Identification of Novel Transcripts involved in the Haematopoietic Differentiation of Human Embryonic Stem Cells

Sun K. Yung

Thesis Submitted for the Degree of

Doctor of Philosophy

Clinical Medical Sciences
Institute of Human Genetics
University of Newcastle upon Tyne

September 2010

ABSTRACT

The holy grail of human embryonic stem cells (HESC) is the generation of a spectrum of differentiated cell types that may be used for clinical and therapeutic use in the treatment and cure of infinite diseases. Over recent years, research into the process of generation of specific cell types from HESC have advanced very rapidly and these advances appear only to be moving with greater speed as scientists sought to understand the specific mechanisms in the generation of these differentiated cells types. This study focused on furthering this understanding through the discovery of novel haematopoietic genes that may contribute to haematopoietic development by screening the entire human genome.

This comparative microarray experiment relied heavily on the isolation and utilisation of haematopoietic-enriched cells acquired through the haematopoietic differentiation of HESC lines H9 and H1. The haematopoietic-enriched populations were generated using an approach that adapted published haematopoietic differentiation protocols, which was then optimised within this experiment for the generation of significantly greater quantities of haematopoietic progenitors compared with the standard foetal bovine serum differentiation approach. These differentiated cells were sorted for a haematopoietic-specific population by fluorescence-activate cell sorting (FACS) for KDR+ CD31+ (markers of early haematopoiesis). Microarray analysis revealed 3,162 transcripts that may play important roles in haematopoiesis, with pathway analysis supporting these findings revealing an overwhelming preference for vasculogenesis, angiogenesis, and erythropoiesis process networks within this gene list. A select number of transcripts were flagged for further investigation, and of these *SOST*, *FLI1*, *EPO*, and *SCL* were of particular interest and hypothesised to enhance *in vitro* haematopoietic differentiation from HESCs.

The effect of SCL on haematopoietic differentiation was assessed using *SCL* over-expressing HESC lines, and it was shown to significantly enhance the efficiency of haematopoietic differentiation. The critical role of *SCL* in erythropoiesis has been well described in literature and within this study, it was discovered that *SCL* over-expression dramatically enhanced erythropoiesis of HESCs by over 25 fold compared to the control. The differentiation of SCL over-expressing HESCs within erythropoietic-optimised media may be the missing link in the efficient generation of therapeutic functional erythrocytes.

ACKNOWLEDGEMENTS

Wow! What a journey! It certainly has been a thoroughly life changing experience for me over the past four years during his study. I thoroughly enjoyed this PhD, although I must admit that at times, it was tough, very tough, and after such low moments in life, I am so happy to be here writing to acknowledge and thank so many people that I am at risk of going over the thesis word limit. I make no apologies regarding the length of this section, because without these people, I would not have been able to achieve what I have in the completion of this thesis, and it therefore gives me great pleasure to take this opportunity to fully acknowledge everyone who have inspired and influenced my life during the course of this project.

To begin, I would like to acknowledge my 'super' supervisor Professor. Linda Lako for literally everything, because if it was not for her I would have never have even had the opportunity to participate in a PhD program. She has always been there for me in guiding the way, and with her amazing knowledge and enthusiasm, she has thoroughly inspired and motivated me at each stage of the way. This past year has been exceptionally hard due to difficulties in my personal life which has unfortunately been brought to the work place, but she has been very understanding and has given me a great deal of support. These past few weeks leading up to the submission deadline has been extremely stressful and I would like thanks Linda for allowing me so much needed time to complete this thesis, and for all her hard work and time that she has also put into this project, which includes reading last minute chapter drafts. With this notice, I would also like to extend my thanks to Mauro Santibanez-Koref, who has been utterly fantastic in delivering very precise chapter corrections for me again, and again, and even over the weekends. Many thanks to Ian Dimmick, our flow cytometry expert who has been solid as a rock and has always been there to help me with my numerous FACS experiments. Thanks to Liming Wang who has been pivitol to the success of this project through her incredible hard work with the processing of the Affymetirx microarray side of the project.

I like to thank my entire family and in particular, my amazing father, who has worked tirelessly to put me through university and it was he who encouraged he to pursue a PhD to complete my studies. He had always been there for me right from the start, and after the sudden death of my mother, he has remained there for me throughout my life. My father has always supported me in everything I have set out to achieve, no matter whether if it was right or wrong, he would be there behind me 100%. Of course I have made mistakes but his support has never been less than maximum, and to reach this far today, I owe a lot of it to him.

I would like to thank literally everyone in my group over the four years of my PhD, which begins with Theresia Walters (my first mentor who took hold of my hand through my first experience of tissue culture), Rebecca Brocklehurst (this superwoman introduced me to the wonders of cloning), Maria Ledran (my master of haematopoiesis who can be quite viscous with a foil), Owen Hughes (great friend and awesome listener), Nicholas Slater (the afternoon/midnight king of kings with a truly unique work ethic), Saj Ahmad (always brings a big smile to my face no matter what), Lyle Armstrong (a very reserved thinking machine with the classic Geordie humour), Xin Zhang (a mummy figure who will always be there to give you sound advice), Chunbo Yang (a stem cell machine), Jin Pei (the cybrid generator), Sai Kolli (the nicest and most generous guy you can possibly meet), Stuart Aktinson (the fastest qPCR machine in the world (aka Sexy Stu), Anna Golebiewska (the smartest / gingerest Polish on the planet), Jim Parris (the American Judo / Boxing master), George Anyfantis (such a hard working fellow, an amazing friend, a music God, quite possibly the greatest but slowest Greek on this planet), Irina Neganova (such a kind-hearted Russian mummy to me), Iliana Paraskevopoulou (the most beautiful [and strongest] Greek woman I have ever met), Ian Dimmick (the king of flow cytometry, there is not a cell population which this man cannot sort, even if it's 0.0001%), Carla Mellough (the most beautiful Scottish/Australian ever), Joe Collin (always cheery and a pleasure to have worked with), Mary Chol (quite possibly the chattiest girl I have ever met), Felipe Vilella (the awesome king of protein, with the highest of class, culture and sophistication), Inma Moreno Gimeno (a beautiful classy lady who certainly has plenty of heart in combination with a dazzling smile), Rita Pilar Cervera Juanes (one of the hardest working stem cell experts I have ever met), Maria Lloret (the smallest, cutiest little Spanish guapa who is always available for a cortado), Dario Melguizo (the most hard working, friendliest, and ambitious colleague I have had the sheer pleasure of working alongside with),

Lastly but not least, I would like to thank someone very important to me, my Miss. Perfect, Katarzyna Tilgner (aka Prof. Tilgner). Kasia was an instant friend when we met over five years ago, and during this time she has always been there for me through all of the highs and of course all of the lows. She was the most influential person in my life during the course of this PhD and in times of trouble, if there were one person who I know I can rely on it would be her. Although she herself may disagree with the immense role she has played in my life, the past few months really showed me how much she means to me. She has helped me in every single way possible during this period of writing hell, and after all of this, I know I will be happy to spend the rest of my life cherishing her, my Miss. Perfect. Moj kochanie, kocham cie.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

Abbreviation	Full Nomenclature
°C	Degrees centigrade
α-MEM	α-minimal essential medium
µg	Microgram
µl	Microlitre
µM	Micromolar
AGM	Aorta-gonads-mesonephros
ALL	Acute lymphoblastic leukaemia
AP	Alkaline phosphatase
ASGM1	Asialo -GM1
bFGF	Basic fibroblast growth factor
BFU-E	Burst forming unit – erythrocyte
bHLH	Basic helix-loop-helix factor
BL-CFC	Blast colony forming cell
BM	Bone marrow
BMP4	Bone morphogenic protein 4
CBF	Core binding factor
CFC	Colony forming cell
CFE	Colony forming efficiency
CFU-E	Colony forming unit – erythrocyte
CFU-G	Colony forming unit – granulocye
CFU-GEMM	CFU – granulocye- erythrocyte-monocyte-macrophage
CFU-GM	Colony forming unit – granulocye-macrophage
CFU-M	Colony forming unit – macrophage
CM	Conditioned medium
dATP	Deoxyadnoside triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DMEM	Dubecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNA	deoxyribonucleic acid
dTTP	Deoxythymidine triphosphate
eNOS	Endothelial NO synthase
EPC	Endothelial progenitor cell
Epo	Erythropoietin
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FBS	Foetal bovine serum
FSC	Forward scatter
GFP	Green fluorescent protein
GMP	Good manufacturing practice

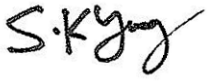
HESC	Human embryonic stem cell
HSC	Haematopoietic stem cell
HUVEC	Umbilical vein endothelial cells
ICM	Inner cell mass
KDR	Kinase Insert Domain Receptor
KO-DMEM	Knockout- Dubecco's modified Eagle's medium
KO-SR	Knockout serum replacement
MEF	Mouse embryonic fibroblast
MESC	Mouse embryonic stem cells
mg	Milligram
M-MLV RT	Moloney murine leukaemia virus reverse transcriptase
mRNA	Mssenger RNA
NK	Natural killer
ng	Nanogram
OCT4	Octamer-4
PBS	Phostphate buffered solution
PI3K	Phosphoinositide 3-kinase
PMT	Photomultiplier tube
P-Sp	Para-aortic splanchnopleura
RNA	Ribonucleic acid
RPM	Rotations per minute
RT-PCR	Reverse transcription polymerase chain reaction
SSC	Side scatter
Stat	Signal transducer and activator of transcription
U	Units
UBC	Umbilical cord blood
VEGF	Vascular endothelial growth factor
vWF	von Willebrand factor
YS	Yolk sac

Abbreviation	**Full and Alternative Cluster of Differentiation nomenclature**
CD31	Platelet endothelial cell adhesion molecule 1 (PECAM-1)
CD43	Leukosialin
CD44	PgP-1
CD56	NCAM-1
CD117	C-Kit
CD144	vascular endothelial cadherin (VE-Cadherin)
CD150	Signalling lymphocyte activation molecule (SLAM)
CD184	Chemokine C-X-C motif receptor 4 (CXCR4)
CD235a	Glycophorin A
CD309	KDR/VEGFR2

DECLARATION

I, Sun Yung, confirm that no part of the material offered has perviously been submitted by me for a degree in this or any other University. Material generated through joint work has been acknowledged and the appropriate publications cited. In all other cases, material from the work of others has been acknowledged, and quotations and paraphrases suitably indicated.

Signature:



Date: 21/09/10

CHAPTER 1 – INTRODUCTION

1.1 Human Embryonic Stem Cells (HESCs)

1.1.1 Origins of Human Embryonic Stem Cells (HESCs)

Human Embryonic Stem Cells (HESCs) are an extraordinary population of primitive cells that have the unique ability to self renew, proliferate indefinitely and differentiate into a vast range of cell types within the body of the organism. Cells with such flexibility to differentiate, develop and mature into a broad variety of cell types are said to be pluripotent [1] and are derived from the inner cells mass at the blastocyst stage of embryonic development (Figure 1) [2, 3]. The inner cell mass has the ability to give rise to the three germ layers termed the ectoderm (external layer), mesoderm (middle layer), endoderm (internal layer) and germ cells (Figure 1). The ectoderm can be split into three parts which includes the external ectoderm (skin, epithelium, enamel, lens, cornea, apical ectodermal ridge, and sensory receptors), the neural crest (pigment cells, ganglia, dorsal root ganglia, Schwann cells, facial cartilage, spiral septum, and ciliary body) and the neural tube (brain, spinal cord, motor neurons, retina, and posterior pituitary)[4]. The mesoderm gives rise to cardiac muscle, skeletal muscle, tubule cells of the kidney, smooth muscle gut, as well as the whole haematopoietic system which is of particular interest in this project. The endoderm is the inner most of the three germ layers and during early development forms the wall of the primitive gut which later develops into the complex gastrointestinal tract and includes pancreatic cells, thyroid cells, larynx, urinary tract, and lung cells. Pera *et al.* (2000) defined a generic criteria for the definition of an embryonic stem (ES) cell and this states that the ES cell must originate from a pluripotent cell population, maintain a normal karyotype of 46 chromosomes, immortal (capacity to propagate indefinitely), and able to spontaneous differentiate producing cells representative of all three embryonic germ layers [5].

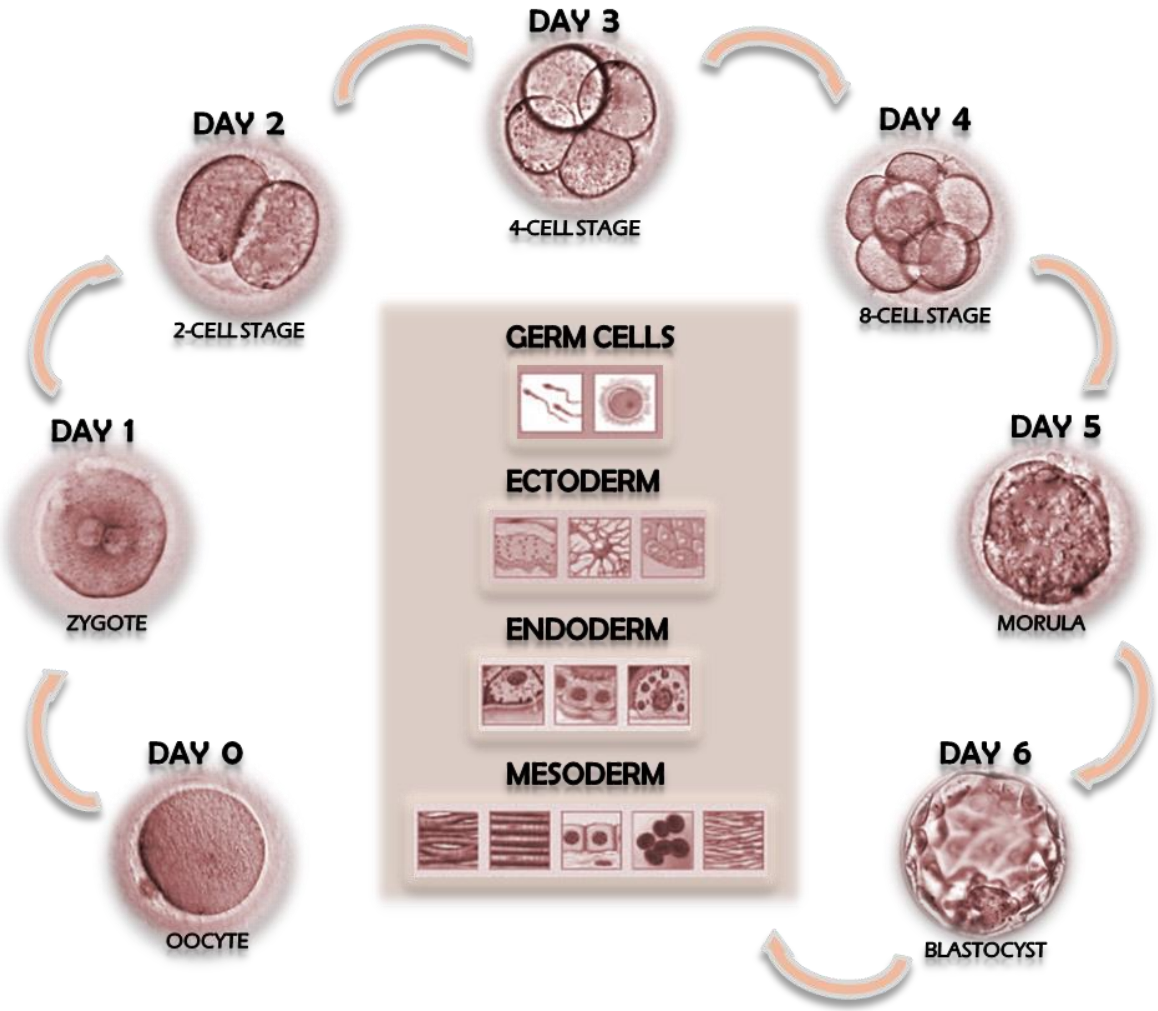


Figure 1: Developmental stages of the pre-implantation blastocyst in human

This diagram shows the multiple stages of development starting from the singled cell oocyte (day 0) to the totipotent morula (day 5) and to the multicellular pluripotent blastocyst (day 6). The inner cell mass from within the blastocyst has the potential to give rise to all three germ layers - ectoderm (including skin cells of epidermis, neuron of brain, pigment cells), mesoderm (including cardiac muscle, skeletal muscle, tubule cell of the kidney, red blood cells, smooth muscle in gut), endoderm (including pancreatic cell, thyroid cells, lung cells). The potential of the cell to differentiate into the range of different cell types decreases after each development stage starting from totipotent to pluripotent to multipotent and finally unipotent. (Reproduced from Terese Winslow (2001) [6])

1.1.2 Uses of Human Embryonic Stem Cells (HESCs)

A cell which has within it the potential to specialise into any cell type in the body carries with it enormous potential value in both medical and commercial terms. By understanding how HESCs transform into the amazing variety of specialised cells, it will provide a detailed insight into normal cell development and may give clues to many serious diseases, which are possibly caused by problems during this process. Therapeutic uses may include transplanting these manipulated HESCs into patients to replace cells in diseased tissues and application for these cells will include the treatment for a myriad of diseases, conditions, and disabilities

including heart disease, diabetes, spinal cord injuries, Parkinson's, Alzheimer's, Crohn's, lupus, burns, blindness, leukaemia and many more [7].

Before HESCs can be used for any therapeutic applications involving human patients, there is still much basic research required to tackle the many technical hurdles that exist between the huge promise of this technology and the realisation of their uses. Through intensive stem cell research this promise is definitely moving closer to the clinic and already many complex events that occur during human development have been identified as a result. A primary goal of this research is to identify and understand the mechanisms of how undifferentiated pluripotent HESCs become specialised, and eventually gain enough knowledge to direct this specialisation process into desired cell types. Upon availability of specialised cell types from HESCs, they may then be use in drugs trials within the pharmaceutical industry to test new medications prior as animal models alone may not be able to accurately reflect the effects on human cells. Thus far there have been many cell lines which are already being used in this way, for example, cancer cell lines are being used to test potential anti-tumour drugs [8] [9], however with the availability of HESCs this will open the door to a much wider range of cell types for drugs testing. Cell based therapies are another important application of HESCs which would offer the possibility of generating an unlimited source of cells and tissues to treat many illness and diseases, for example, skin burns which would mean that fresh skins cells produced by HESCs can be use to replace the damaged tissue of the patient. This type of therapy can expand to treat many other diseases such as heart disease and diabetes whereby healthy heart tissue and pancreatic cells are generated in the lab from HESCs and then transplanted into patients. Studies with mouse have shown that when enriched haematopoietic stem cells (HSCs) were transplanted into the coronary artery, the cells migrated to the cardiac muscle and blood vessels where they differentiated to form cardiomyocytes and endothelial cells and restored function to the heart [10]. Incredibly, the first human clinical trial of embryonic stem cell-base therapy has recently been give the green light by the U.S. Food and Drug Administration (FDA) to inject HESC-derived oligodendrocyte progenitor cells into patients with acute spinal cord injury [11]. This human clinical trial was put on hold by the FDA for a while due to the immense safety concerns, however further animal trials have alleviated these precautions and consistently demonstrated that the HESC-derived oligodendrocyte progenitor cells carried remyelinating and nerve growth stimuating capacities which completely restored neural function in animal model with spinal injury [12].

A common concern for people is that current method of HESC derivation, proliferation and maintenance requires the use of many animal and non-human components, and this is a problem when it comes to using HESC derivatives for therapeutic applications. There are many obvious risks of using xenosupport systems in deriving and subculturing HESCs due to the

possibility of cross-transfer of animal pathogens to the patient and to the wider population. However, as such is the pace of advancement in science, many groups are already having much success with animal-free culture systems, either using feeder-free [13-16] or human feeders [17-20], with a number of good manufacturing practise (GMP) HESC lines already established [21, 22].

1.1.3 Derivation of Human Embryonic Stem Cells (HESCs)

Thomson *et al.,* (1998) [23] was the first group to successfully derive HESCs from the inner cell mass of early human embryos in 1998. The cell lines which were produced were from 14 inner cell masses (ICM), were isolated at the blastocyst stage from five different embryos and resulted in five cell lines named H1, H13, H14, which contain a normal XY karyotype, and H7 and H9, which contain a normal XX karyotype [23]. The isolated ICM from the blastocysts were placed upon inactivated mouse embryonic feeders (MEFs) and the resulting cell population was carefully subcultured by serial passaging from old MEFs to new MEFs [24]. Colonies of these newly derived HESCs show that they contain high ratio of nucleus to cytoplasm, prominent nucleoli, and the colony morphology is similar to those of primate ES cells [23]. All cell lines exhibit the stem cell characteristics of self renewal and unlimited proliferation, and the group has also stated that all five cell lines retained a normal karyotype after six months of continuous undifferentiated proliferation [23]. Similar protocols of HESC derivation has since been used by many different groups around the world in deriving additional lines including derivation on human feeders (human adult fallopian tube epithelial [17], human foetal and adult fibroblasts [20], human placental fibroblasts [25], human foreskin fibroblasts [18]) and feeder free (for example BD Matrigel) [13-16].

1.1.4 Characterisation of Human Embryonic Stem Cells (HESCs)

The HESC has many unique and distinct characteristics as discussed previously which allow it to be distinguished from more mature differentiated cell types of the body. There are a number of tests which scientists are able to measure and identify the fundamental properties of HESCs and the simplest begin with culturing the cells and inspecting them under a

microscope to check that they are healthy and remain undifferentiated. The typical morphology of an undifferentiated stem cell should be one that is small, round and contain a noticeably large nucleus with lots of nucleoli which shows high gene transcription / translation activity. Colonies of HESCs typically grow as a thin and flat monolayer, and spontaneously differentiated cells can easily be identified as thicker, larger multilayer cells forming around the colony at first, and gradually moving inwards until all HESCs have differentiated [23].

As well as the physical morphology and size of the cells which can be used to identify HESC, they also share common surface antigens which are specific to pluripotent undifferentiated cells. There are a number of HESC markers which have been originally identified from human embryonic carcinoma (EC) cells, these include stage-specific embryonic antigen 3 (SSEA-3) [26, 27], stage-specific embryonic antigen 4 (SSEA-4) [26-29], tumour-rejection antigen-1-60 (TRA-1–60) [5], tumour-rejection antigen-1-81 (TRA-1–81) [5], and alkaline phosphatase [13]. There are a further number of HESC markers which are shared with other stem cells (for example haematopoietic stem cells) and these include AC133 [30], C-KIT (CD117)[31], FLT3 (CD135)[32], and CD9 [31]. As well as surface markers, there are a number of key transcription factors which play a important role in maintaining HESC characteristics and analysis of their expression is also used to characterise HESCs [31]. *OCT-4* belongs to the *POU* family of transcriptional regulators and it is the most widely recognised marker of pluripotent HESCs as it is required to for pluripotent cell lineage establishment [33-36], and upon differentiation of HESC, transcriptional levels of *OCT-4* decrease immediately [37]. The homeobox transcription factor *NANOG* is another important pluripotent marker which works with *OCT-4, FOXD3* and *SOX2* within a regulatory network to support/limit each other's expression level to maintain pluripotency within ES cells (Figure 2) [38].

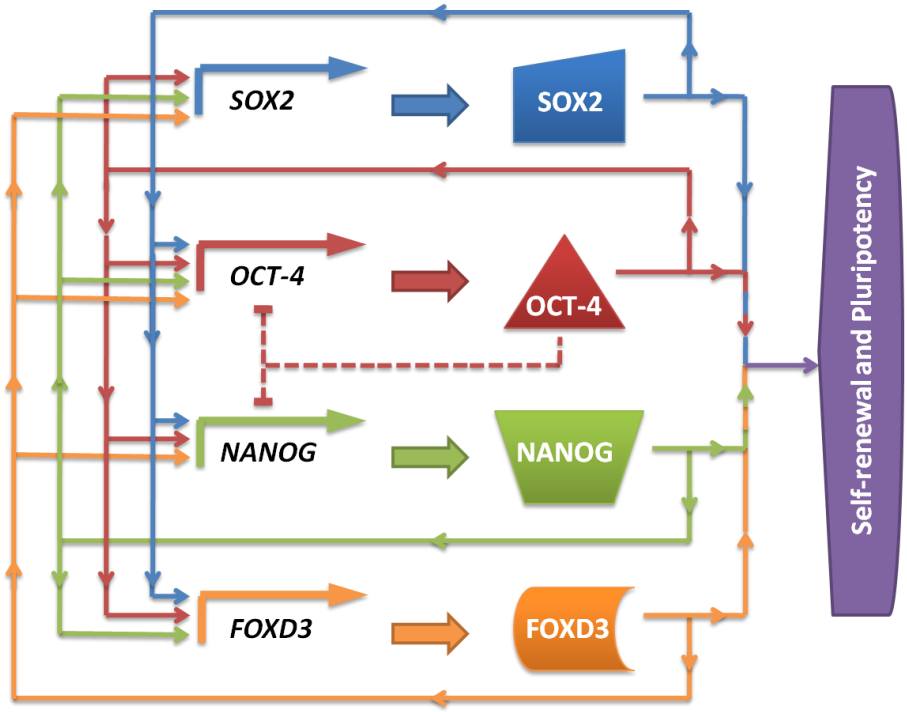


Figure 2: Regulatory network of key transcription factors in the maintenance of ES cell pluripotency

Master pluripotency regulators *OCT4, NANOG, SOX2* and *FOXD3* bind to each other's promoter and enhance / inhibit each other's expression, forming an interconnected auto-regulatory network to maintain ES cell pluripotency and self-renewal [38]. Arrows connected to factors by solid lines indicate positive regulation of a promoter by the factors and broken lines linking from OCT4 indicate negative regulation.

In addition to HESC marker analysis there are a other methods which aid in the characterisation of the undifferentiated, pluripotent state of HESCs and these include expression profiling, cytogenetic analysis, differentiation analysis, and epigenetic analysis [31]. Expression profiling of HESC lines involves the generation of transcriptome profiles using serial analysis of gene expression (SAGE) and comparing these with non-HESCs to produce expression patterns within genes that are unique to HESC cells [39]. Cytogenetic analysis involves the karyotyping of HESC and ensuring that they exhibit a normal complement of chromosomes, as this demonstrates the unique ability of HESC to maintain a normal karyotype through long-term subculturing periods (old H1 and H9 show normal karyotype at passages ranging from 60-100 [16]). The ability of pluripotent HESCs to differentiate into the three embryonic germ layers mesoderm, ectoderm and endoderm is another characteristic test of pluripotency (Figure 1) [31]. The table within Figure 3 demonstrates the pluripotent capacity of the various HESCs through the generation of mesoderm, ectoderm, endoderm, as well as trophoblast and germ cells. It also demonstrates that not all HESC lines are created equally and that some have a superior capacity over others in the generation of particular cell types. *In vivo* undifferentiated, pluripotent HESCs have the ability to form teratomas which contain cells that belong to all three germ layers.

Cell type	Comments	Cell lines	Ref
General Differentiation			
Endoderm, ectoderm and mesoderm derivatives	Analysis of mRNA expression of various markers after growth factor treatment of HESC−derived EBs.	H9.1	[40]
Endoderm, ectoderm and mesoderm derivatives	Analysis of expression of various markers using PCR and immunocytochemistry.	H9	[41]
Endoderm, ectoderm and mesoderm derivatives	Analysis of gene expression during differentiation by microarrays.	H9, H13	[42]
Mesoderm			
Cardiomyocytes	Differentiation into enriched, populations of functional cardiomyocytes.	H1, H7, H9	[43]
Cardiomyocytes	Differentiation into cardiomyocytes.	HES2	[44]
Cardiomyocytes	Characterization of HESC−derived cardiomyocytes.		[45]
Cardiomyocytes	Evaluation of electrophysiological and pharmacological properties	H1, H7, H9, H14	[46]
Cardiomyocytes	Evaluation of electrophysiological and pharmacological properties.	H9.2	[47]
Cardiomyocytes	Demonstration of functional integration after transplantation into guinea pig model.	H1	[48]
Haematopoietic colony−forming cells	Differentiation of multi-potent haematopoietic precursors.	H1, H1.1, H9.2	[49]
Haematopoietic progenitor cells	Differentiation of multi-potent CD34+ / CD45+ haematopoietic precursors.	H1, H9	[50]
Leucocytes	Differentiation into antigen-presenting leucocytes.	H1	[51]
Endothelial cells	Isolation and characterization of HESC−derived endothelial cells.	H9	[52]
Endothelial	Identification of primitive endothelial-like cells that generate endothelial and haematopoietic cells.	H1, H9	[53]
Endothelial	Evaluation of endothelial markers during spontaneous differentiation.	H9.2	[54]
Vasculogenesis	Vasculogenesis using alginate scaffolding.	H9.2, H13	[55]
Vasculogenesis	Vasculogenesis in teratomas.	H9.2, H13, I6	[56]
Ectoderm			
Neurons	RA and NGF enhance neuronal differentiation.	H9	[57]
Neurons and glia	Differentiation into enriched populations of	H1, H7, H9	[58]

	functional neurons and neural progenitors.		
Neurons and glia	Differentiation into neural progenitors, neurons and glia. Transplantation into neonatal rats.	H1, H9, H9.2	[59]
Neurons and glia	Differentiation into neural progenitors, neurons and glia. Transplantation into neonatal rats.	HES1	[60]
Neurons	Differentiation into neural precursors and neurons.	BG01, BG02	[61]
Neurons	Differentiation into dopaminergic neurons. Transplantation into Parkinsonian rats, resulting in functional improvement.	HES-1	[62]
Neural precursors	Differentiation into neural precursors.	HES-2, HES-3	[63]
Neurons	Differentiation into midbrain dopaminergic neurons.	H1, H9, HES-3	[64]
Neurons	Differentiation into dopaminergic neurons.	BG01, BG03	[65]
Neurons	Differentiation into motoneurons.	H1, H9	[66]
Oligodendrocytes	Differentiation into enriched populations of oligodendrocytes and myelination after transplantation.	H7	[67]
Endoderm			
Insulin-producing cells	Differentiation into insulin producing clusters.	H9, H9.2, H13, I6	[68]
Hepatocyte-like cells	Differentiation into enriched populations of hepatocyte-like cells.	H1, H9	[69]
Trophoblast			
Trophoblast	Differentiation into trophoblast cells.	H1, H7, H9, H14	[70]
Trophoblast	Identification of trophoblast cells in EBs.	H1	[71]
Germ cells			
Germ cells	Spontaneous differentiation of germ cells.	HSF-1, HSF-6, H9	[72]

Table 1: Summary of Differentiation of HESCs

This table shows the wide variety of cell types of each of the three embryonic germ layers that are produced from the differentiation of various HESC lines [31]. The two most common HESC lines in use in our laboratory are H1 and H9 which both have the ability to give rise to cells from all three germ layers, but their efficiency to proliferate and give rise to a chosen cell type differs significantly between the two lines. Studies involving the differentiation of haematopoietic and mesodermal cell types have tended to prefer the use of H1 cell line and differentiation of neural and ectodermal cells prefer H9 cell line simply because of their natural tendency to differentiate down a particular lineage when they spontaneously differentiate.

1.1.5 Differentiation potential of Human Embryonic Stem Cells (HESCs)

Pluripotent HESCs require carefully nurturing in a specialised culture environment in order to maintain their pluripotency and remain in an undifferentiated state. They will quite readily lose their pluripotent state and spontaneously differentiate into a cell type representative of the three embryonic germ layers, (mesoderm, endoderm and ectoderm) (Table 1) [31]. Studies from many groups (Table 1) have shown that both HESC lines H1 and H9 have the pluripotent capacity to differentiate into a wide range of cell types of the three embryonic germ layers, however it has only been reported that H9 HESCs were able to differentiate into germ cells [72].

The study of early human development has and remains a difficult area to approach due to the ethical, as well as technical limitations of studying human embryos, and hence many groups have resorted to using animals as developmental models for their answers. Although this will provide good background to early development and form a good foundation for clearer understanding of the process, however direct extrapolation of development in animals to humans may be inappropriate. Recent studies have shown that HESCs can differentiate into cell types that are representative of all three embryonic germ layers within embryonic bodies (EBs). In particular, specific stages of early development may be studied using human EB system, and this includes differentiation down the mesoderm lineage to form haematopoietic colony-forming cells (CFCs) [49, 50, 73, 74]. The formation of cellular spheres called embryonic bodies (EBs) created from HESC colonies provides an excellent system for the differentiation of HESCs and formation of more committed cell lineages *in vitro* [40, 41]. Many studies have utilised this EB system in studying spontaneous and directed differentiation of HESCs as they recapitulate many aspects of early development together with the capacity to produce cell types from all three embryonic germ layers [38-78]. "*Hematopoietic cells develop within embryoid bodies (EBs, in vitro differentiated ES cells) faithfully following in vivo developmental kinetics*" – Chung *et al.,* (2002)[75]. Within carefully controlled culture conditions and environment, HESC differentiation within EBs may be specifically directed towards a particular lineage or cell type [76] and within this project I will also be utilising the EB method in my attempt to persuade the HESCs to develop into haematopoietic progenitors using a combination of growth factors and cytokines.

1.1.6 Ethics of Human Embryonic Stem Cell (HESC) Research

In the search for ways to create ethical human embryonic stem cells, many groups are trying to find ways to obtain these precious cells without destroying human embryos. Within our group, Zhang *et al.* (2006) [77] has developed a technique by which human embryonic stem cells are derived from very young embryos which have arrested during embryonic development. The technique utilises embryos which have arrested (stopped dividing) and in any IVF clinic in the UK these embryos would not meet the standard criteria for implantation, and would typically be discarded. Many of the arrested embryos contained cells which were distorted in shapes and appeared to have damaged chromosomes, but the few healthy cells which are available were nurtured and using various culture conditions gave rise to a new cell line, NCL-9. "*In theory this approach is similar to harvesting organs from dead people – but in the latter case there are clearly defined criteria for proclaiming someone irreversibly brain dead. There are no such rules for an embryo*" – Pearson and Abbott (2006) [78]. The definition describing when an embryo is dead has fired up much debate and will continue on until agreed criteria are established. Currently Landry (2006) [79] is working on trying to establish "ironclad" criteria to define the death of an embryo by looking for specific molecules/markers which are produced solely by terminally arrested embryos.

UK government ministers have recently granted a licence which will enable scientists to create human-animal hybrid embryos within the laboratory to address the shortage of human oocytes available for research [80]. This technique involves the transfer of human DNA into an empty animal oocyte which have had almost all of its genetic material removed producing an embryo which is about 99.9% human and the remainder animal (egg shell and cytoplasmic mitochondria) [81]. However, even this method of HESC generation has been fraught with ethical dilemmas brought on by the public's misconceptions and fears of the possible generation of a mixed human-animal being, with religious groups branding these as 'experiments of Frankenstein proportions' [82]. Despite these outcries, the Human Fertilisation Embryology Authority (HFEA) after a lengthy consideration, granted licenses in 2008 to a number of groups including ours (Dr. Lyle Armstrong). Unfortunately thus far, this method has had limited success with studies showing that human-cow cytoplasmic hydrids (cybrids) are failing to progresss to the blastocyst stage of development (where HESC are derived) and to this date there are not been any publications on the successful derivation of cybrid HESCs.

Very recently, a number of groups have described techniques which would mean that human embryonic-like stem cells may be obtained from somatic adult cells (for examples skin cells) and thus removing the need for eggs and embryos [83]. Takahasiki and Yamanaka (2006)

[83] first described the induction of mouse embryonic/adult fibroblasts back into embryonic-like cells known as mouse induced pluripotent stem cells (mIPSCs) by inducing four transcription factors, *OCT4, SOX2, C-MYC,* and *KLF4*. However these results were met with much scepticism because although these mIPSCs appear to have the ability to proliferate continuously, form colonies and cancerous growths (teratomas), they also lacked many other abilities that are characteristic of embryonic stem cells such as the ability to produce a chimaera (a mix of DNA from both original embryo and mIPSCs throughout the body) when introduced into a developing mouse embryo. A year later, the efficiency in the isolation of these reprogrammed mIPSCs have greatlly improved by including a gene for antibiotic resistance which is activated by the expression of key genes that are characteristic of stem cells [84, 85]. Improved protein markers in the identification and isolation of IPSCs using *Nanog* and *OCT4* resulted in the successful production of chimearic mice, leading researchers to begin applying the same method with human cells [86]. The breakthrough occurred shortly later by Yamanaka and colleagues in 2007 when they successful reprogrammed human somatic cells into IPSCs using human versions of the same reprogramming factors (*OCT4, SOX2, C-MYC,* and *KLF4*) delivered by retroviral vectors [87]. Another group led by Thompson also successfully reprogrammed human somatic cells into IPS cells albeit using a different combination of reprogramming factors (*OCT4, SOX2, NANOG,* and *LIN28*) with a lentiviral approach [88]. Within 3 years of the initial break through by Yamanaka and Thomson, there has since been almost 2000 published IPS-related papers (as indicated by a Pubmed search) as more and more research groups focus their attention into this promising area of research. This surge in IPSC research has rapidly begin to assess the potentials of these IPSCs, and also many of the fundamental problems which have limited IPSC generation, including reprogramming efficiency and also safety in terms of using non-integrating methods in the delivery of the reprogramming factors. From the rapid growing number of publications it seems more and more likely that IPSCs may replace the requirement for HESCs in the future and may ultimately be used in patient-specific therapy, but these all depend on the efficiency of reprogramming, true representation to HESCs, and of course the safety aspect of these cells.

1.2 Haematopoietic System

1.2.1 Introduction to Haematopoietic Stem Cells (HSCs)

Haematopoietic Stem Cells (HSCs) are in great demand in the treatment of many blood disorders which include leukaemia, multiple myeloma, aplastic anaemia, thalassemia, and sickle cell disease. HSCs are commonly used in the treatment of these diseases and are obtained primarily from bone marrow and peripheral blood, but these are in short supply due to limited compatible donors. An alternative source of HSCs may be acquired from umbilical cord blood which offer a more youthful population than those in the marrow and peripheral blood, but problems still occur due to limited quantities obtainable which translate into lower engraftment and hence slower repopulation rates in the marrow, and thus leaving the host more susceptible to infections and other complications. A future source of HSCs may be acquired from the controlled differentiation of pluripotent progenitors such as human embryonic stem cells (HESCs). A distinct advantage of obtaining HSCs derived from HESCs over bone marrow and umbilical cord blood is that the source of HESCs in theory has the ability to expand indefinitely, which would mean that the quantity of HSCs which may be obtained will be limitless. As well as a potentially limitless supply of HSCs, these will also be a more youthful population and as a result have considerably less exposure to oxidative stresses leading to less DNA damage and chromosomal abnormalities, and there will also be a lower risk of pathogenic contaminants and autoimmune diseases. For these reasons the prospect of utilising HESCs in the generation of HSCs has brought enormous attention into this area of research. In particular differentiation studies from HESCs to HSCs have been greatly hindered by the lack of positive markers for the accurate identification of specific cell types at the various stages of development. The accumulation of knowledge through this research may assist in the discovery of novel markers which play an important role in haematopoiesis and contribute to further our understanding into how various processes induce efficient differentiation of HESCs into HSCs, which will pave the way for the future treatment of many blood disorders and make this a reality.

1.2.2 Haematopoietic Hierarchy

The haematopoietic system is firmly based on a hierarchical tissue-specific stem cell system whereby haematopoietic stem cells with their unique capacity to self-renewal have also the potential through asymmetric division to give rise to all haematopoietic cell types within the body (Figure 3). With such a high turnover of cells as is the case of the haematopoietic system, the requirement for there to be a constant source of new cells to replace those that have been lost is clearly of critical importance. The haematopoietic hierarchy begins with a master pluripotent stem cell which differentiates and gives rise to cell types with progressively restricted lineage capacities. The formation of the multipotent haematopoietic stem cell retains its long-term self-renewal capacity but its progeny of multipotent progenitors (MPP; for example CFU-GEMM) although still capable of full lineage potential has however lost their ability to self-renewal. The oligopotent progenitors (for example CFU-Ba) which are derived from MPP are committed lineage-restricted progenitors and therefore can only give rise to mature functional cell types which typically have shorter life spans and higher turnover rates.

With reference to the haematopoietic hierarchical illustration presented in Figure 3, the multipotent haematopoietic stem cell, as previously described have the unique ability to self-renewal and to give rise to all of the cell types within the adult haematopoietic system. The red blood cells (erythrocytes) contribute to approximately 25% of all cells in the human body with 2.4 million being created every second and circulating between 100 to 120 days before being recycled by macrophages [89]. Mature erythrocytes are enucleated and consist largely of haemoglobin which allows it to temporarily bind to oxygen molecules in high concentration regions (for example at the lungs) and releasing it in low concentration regions (for example at tissue) throughout the body [89]. Erythrocytes release adenosine triphosphate (ATP) and nitric oxide when they experience shear stress within constricted blood vessels, which in turn causes the vessel walls to dilate, assisting in the regulation of vessel size in the promotion of blood flow [90]. Studies have recently shown that erythrocytes also play a role in the immune system whereby when pathogens break open erythrocytes, the haemoglobin contained within release unstable free radicals that break down the microbe's cell wall and membrane, and thus killing the pathogen [91]. The full complement of the immune system is composed of five different types of white blood cells (leukocytes) which fall into one of two groups including polymorphonuclear leukocytes (granulocytes) and mononuclear leukocytes (agranulocytes) [92]. The granulocytes are thus termed because their cytoplasm is filled with granules which contain enzymes and antimicrobial chemical that lyse and destroy pathogens, and these are split into three types, neutrophils, basophils, and eosinophils. The agranulocytes consists of

two types, lymphocytes and monocytes, and in contrast to granulocytes, these contain no visible cytoplasmic granules.

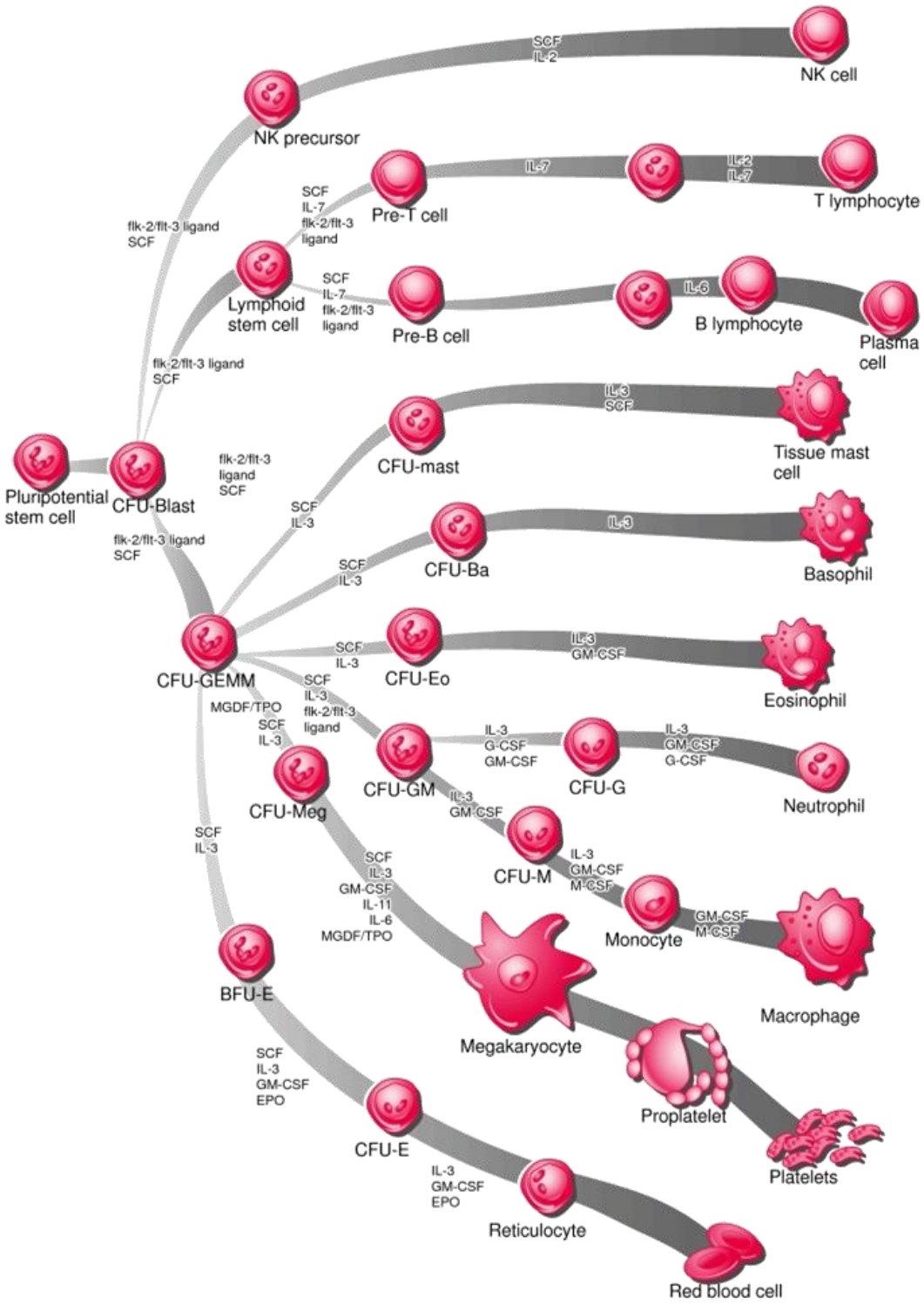


Figure 3: Haematopoietic Hierarchy

The haematopoietic system begins with the pluripotent stem cell which gives rise to the multipotent CFU-Blast cell (haematopoietic stem cell) which in turn has the capacity to form the myeloid precursor - granulocytes, erythroids, monocytes, and megakaryocytes (CFU-GEMM). This cell type in turn has the ability to provide cells of granulocytes and monocytes (CFU-GM), erythrocytes (from burst-forming units (BFU-E) intermediate), megakaryocytes (CFU-Meg), mast cells (CFU-mast), eosinophils (CFU-Eo), and basophils (CFU-Bas). (Reproduced from Holland-Frei Cancer Medicine (2003) [93])

Neutrophil granulocytes are the most abundant white blood cells (accounting for 70% of all white blood cells) in humans with a life span of about 12 hours, and during an infection they will be one of the first to respond to the site of inflammation in the defence against bacteria and fungi through phagocytosis followed by the release of lysosomes used in the digestion of the microbes [92]. Basophil granulocytes are one of the rarest white blood cells (representing 0.01% to 0.3% of all circulating white blood cells) in humans and are primarily responsible for promoting inflammation through the release of histamine and prostaglandins in response to allergic and antigen stimuli in the recruitment of other white blood cells [94]. Eosinophil granulocytes make up between 1% and 6% of white blood cells with a life span of about 12 hours in circulation and up to 12 days within tissue, they are responsible against parasitic infections and similarly to basophils they contain many chemical mediators (such as histamine, eosinophil peroxidise, ribonuclease, deoxyribonueclease, lipase, plasminogen, and major basic protein) which are released during eosinphil activation and are toxic to both parasitic and host cells [92].

Lymphocytes are derived from the lymphoid stem cell (with the exception of natural killer (NK) cells) act to mediate the adaptive immune responses. Natural killer (NK) cells form the bulk of the innate immune system by which they target and eliminate cells through antibody-dependent cellular cytotoxicity whereby small cytoplasmic granules of proteins are released causing cells that are missing ‘self’ markers of major histocompatibility complex (MHC) class 1 to apoptose [92]. The MHC plays a vital role in the immune system and autoimmunity, primarily functioning as a cell surface marker of normal cells (self) or infected cells (non-self), with the MHC changing in infected cells and this change is detected by NK or T-cells which in turn initiates an immune response [95]. The B-cell lymphocytes make up between 10 and 20% of all lymphocytes in humans, and their role is to mediate humoral immunity through the production of antibodies against specific antigens exhibited by invading pathogens, and also form memory cells which recognise the antigens exhibited by the pathogen resulting in a faster immune response in future infections [95]. The T-cell (also known as T-lymphocyte) are responsible for cell mediated immunity, existing in a multitude of forms, these are responsible for a number of roles which include T-helper cells (activation of lymphocytes), cytotoxic T-cells (destruction of infected cells), memory T-cells (recognition of re-exposed antigen), regulatory T-cells (suppression of T-cell mediated immunity), and natural killer T-cells (elimination of infected cells through recognition of glycolipid proteins) [92]. Monocytes perform a similar role to that of the neutrophil through the process of phagocytosis of pathogens (as macrophages), but they differ in that they have a longer life span (brought on by the ability to replenish their lysosomal content whereby neutrophils can not). They also perform a role in antigen presentation whereby microbial fragments are

incorporated into MHC molecules and presented on their cell surface to T-cells which results in an immune response against that specific antigen [96].

Mast cells do not reside in the circulatory system but rather within tissues and they perform a similar role to basophils in that they both contain granules which are rich in chemical mediators (histamine and heparin), and upon detection of injury, these chemicals are released into the blood stream [97]. The release of histamine stimulates the blood vessels to dilate and capillary walls to become more permeable and thus allowing more blood to reach the site of injury. Heparin assists in the action of histamine by preventing the blood from clotting and thus promoting unimpeded blood flow to the site of injury. The formation of blood clots at sites of vessel damage requires the activation of irregularly-shaped anuclear cell fragments called platelets, which clump together upon activation to temporarily reduce blood loss and maintain blood pressure [89].

This brief summary of the components that make up the haematopoietic system provides an idea of the huge diversity of cell types and function which are involved, and how all of these incredibly originate from a single cell type, the multipotent haematopoietic stem cell. The adult haematopoietic stem cell is used routinely in clinical practical in transplantation-base therapies but the scarcity of this cell type (due to limited donors) have meant that the generation of this multipotent progenitor from human embryonic stem cells (and induce pluripotent stem cells) remain the ultimate goal for use in research and clinical applications. The differentiation of haematopoietic stem cells from human embryonic stem cells commences with the generation of the mesoderm germ layer.

1.2.3 Mesoderm Induction

The development of the mesoderm begins when pluripotent human embryonic stem (ES) cells from the inner cell mass (ICM) of the formed blastocyst at about day 5 post fertilisation gives rise to the epiblast and primitive visceral endoderm. The epiblast gives rise to the three germ layers including endoderm, ectoderm, and mesoderm during gastrulation, and the development of these germ layers and their subsequent cell fates are tightly controlled by a series of complicated mechanisms and processes. The process of gastrulation is highly dependent on spatial and temporal regulatory control with the appearance of the primitive streak signifying the beginning and provides one of the earliest structural identifiers of the mesoderm. From as early as the 1930s, morphological studies in the mouse and rat had suggested that mesodermal and definitive endodermal cells to be the main types derived from the primitive streak [98]. Gastrulation studies into mesoderm formation in mouse by Tam and Beddington, (1987) [99] identified cells of the paraxial mesoderm, gut and notochord to be primarily located at the anterior region of the primitive streak, with the posterior region giving rise to extraembryonic mesoderm and the midsection giving rise to the lateral mesoderm. The lateral mesoderm is later split by the coelom (intraembryonic cavity that separates the outer and inner components of the body) into two layers with the mesoderm that lies dorsal to the coelom called the somatic lateral mesoderm and the mesoderm at the ventral region of the coelom called the splanchnic lateral mesoderm (Figure 4) [100]. Studies have suggested that the cells of the splanchnic lateral mesoderm will give rise and populate the haematopoietic and circulatory system (Table 2) as well as those of the heart, smooth muscle, and membranes that surround many of the organs [101, 102].

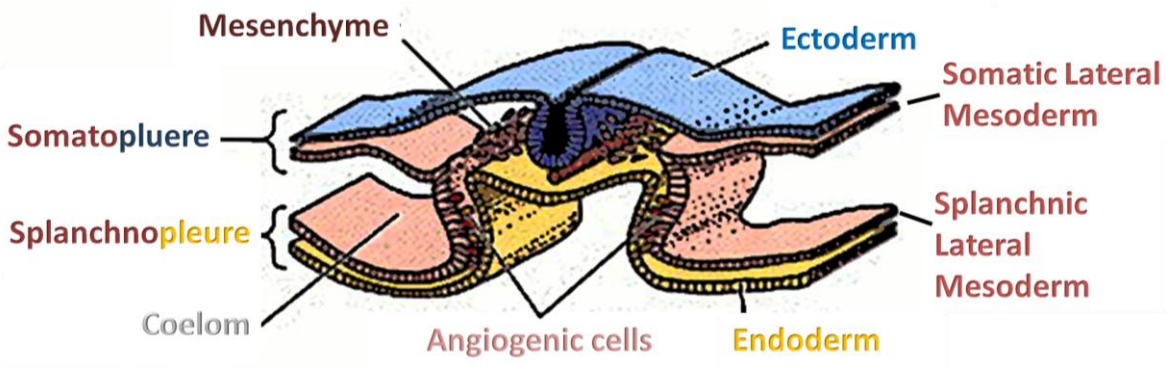


Figure 4: Gastrulation

A transverse section through part of the developing embryo shows the various components of gastrulation including the coelom which separates the lateral mesoderm into somatic and splanchnic mesoderm. Adapted and reproduced from Patten's Foundations of Embryology [103]

Mesoderm Types	Developmental Fates
Paraxial Mesoderm	Axial skeletal muscles Bones Dorsal dermis
Intermediate Mesoderm	Kidneys Gonads Sexual and reproductive ducts Upper vagina and uterus in females
Somatic Lateral Mesoderm	Connective tissues Parietal pleura (lining of the inner surface of chest wall) Parietal pericardium (outer most membrane attaching heart to diaphragm and sternum) Peritoneum (lining of the intra-abdominal organs)
Splanchnic Lateral Mesoderm	Heart Blood vessels Blood cells Smooth muscle Visceral pleura (lining of the lungs) Visceral pericardium (inner most membrane that surrounds the heart) Peritoneum (lining of the intra-abdominal organs) Mesenteries (double layer of membrane which supports the small intestine) Outer layer of the adrenal gland

Table 2: Mesodermal types and their developmental fates

Mesoderm induction requires the extracellular signal regulated kinase 2 (*ERK2*) and together with mitogen-activated protein kinase (*MAPK*) casade, this signal transduction pathway that links extracellular simuli to a transcriptional response, regulates a number of key signal pathways which include cell proliferation, differentiation, and apoptosis [104]. Studies by Yao *et al.* (2003) [105] had shown when exon 2 of *erk2* was deleted by homologous recombination in mouse, resulted in the complete absence of mesodermal cells in the mutant embryos. However when *erk2-/-* ES cells were studied it was observed that a second member of the MAPK casade, ERK1, was able to compensate (albeit at a significantly lower level) for the

absence of ERK2 and rescue mesoderm formation *in vitro*. The same study showed that in the absence of ERK1/2, mutant ES cells were able to proliferate at a normal rate, which contradicted past studies which suggested ERK1/2 to be indispensible for cell proliferation during early development. A separate study by Kren *et al.* (2008) analysed the gene expression signature of what genes are affected by ERK1 and ERK2 knockdown using specific morpholino antisense oligonucleotides which block the translation of the target gene [106]. It was revealed that ERK1 is involved in dorsal-ventral patterning and cell migration whereas ERK2, which shares 75% sequence identity at the amino acid level and shares the same substrate specificity as ERK1, is involved in mesoderm and endoderm induction, maintenance, differentiation and patterning.

As well as ERK2, other studies have implicated bone morphogenetic protein 4 (BMP4) [107, 108] together with fibroblast growth factor (FGF) [109, 110] and transforming growth factor beta (TGF-β) / Nodal / Activin [111] signalling activities as playing important roles in mesoderm development. Zhou et al. (2007) [112] demonstrated that the MAPK/ERK cascade is activated by BMP4 (Figure 5) in human umbilical vein endothelial cells (HUVECs), and in the presence of BMP4, collagen gel-embedded HUVECs sprouted capillaries which were significantly longer and faster than negative controls. The same study showed that ERK1/2 was essential for capillary sprouting and without the MAPK/ERK signalling cascade, considerably fewer and shorter capillaries formed in the presence of BMP4 alone. BMP4 is part of the superfamily of TGF-β proteins which studies have shown to be involved in a wide range of biological processes including bone formation, limb development, fracture repair, muscle development, and dorsal-ventral axial formation during embryogenesis [113, 114]. Winnier *et al.* (1995) [115] demonstrated the importance of BMP4 in mouse mesoderm development by targeted disruption studies of *Bmp4* and its receptors (*BmprIa* and *BmprII*) which resulted in embryonic lethality between 6.5 and 9.5 days post cotium (d.p.c.) as a result of the lack of mesoderm differentiation. A number of studies [112, 116] have suggests that BMPs may also signal through Smad independent pathways such as MAPK/ERK (Figure 5). BMP4 in parallel induces the small mother against decapentaplegic (Smad) cascade of proteins which mediates their action through intracellular signalling, forming complexes which enter the nucleus and regulates transcription of a number of genes (Figure 5) [116].

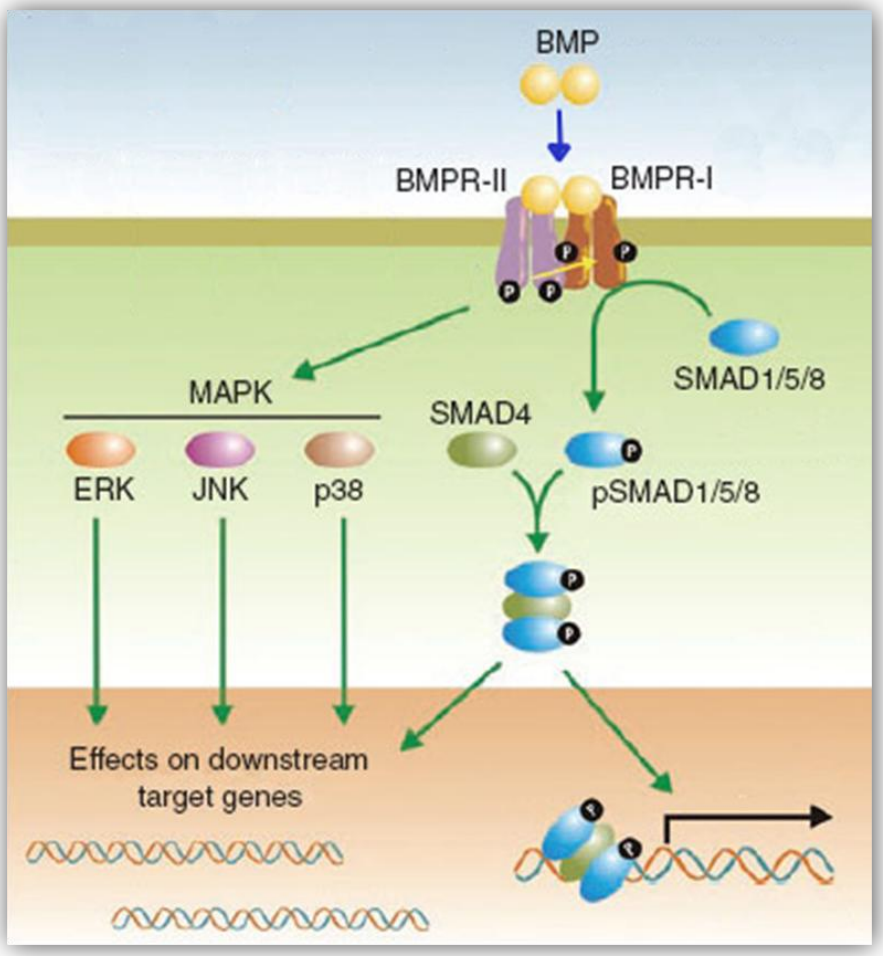


Figure 5: BMP and Smad signalling pathway

The BMP pathway is activated by through the binding of the BMP ligand to the constitutively active kinase domains of type II receptor (BMPR-II) which phosphorylates BMPR-1. Receptor regulated Smad signalling proteins (Smad1/5/8) are activated through phosphorylation by BMPR-I. Activated Smad proteins associate with common-partner Smad4 to form a heteromeric complex which translocates to the nucleaus and regulates the transcription of a number of genes. Reproduced and adapted from Anderson, G.J. and Darshan, D. (2008) [116]

As illustrated in Figure 5, BMPs binds and activates specific cell surface receptors (type I and type II serine / theronine kinase) which in turn phosphorylates and activates receptor-regulated Smad1 / Smad5 / Smad8 proteins. Hetero-oligomeric complexes are formed between the activated receptor-regulated Smads and common-partner Smad4, which translocates into the nucleus whereby it regulates the transcription of a wide range of genes. This complex is however subjected to regulation by a negative feedback loop through the actions of inhibitory Smads (Smad6 / Smad7) which functions by disrupting complex formation or interfering with the phosphorylation of the receptor-regulated Smads [112]. Targeted homozygous disruption of *Smad5* by Chang *et al.* (1999) [117] resulted in embryonic lethality between 9.5 and 11.5 d.p.c., due to abnormal vasculature and defects in the developing aminon, gut, and heart, which suggests that Smad5 plays a key role in angiogenesis [117]. The

Smad1 knockout mouse also resulted in embryonic lethality around 9 d.p.c. caused by chorio-allantois defects resulting in the lack of a placental connection and a definite embryonic circulation [118]. Loss-of-function studies of Smad1 and Smad5 in zebrafish reveal distinct roles during embryonic haematopoiesis and it appears that Smad1 plays an important part in the production of mature embryonic macrophages, whereas Smad5 is involved in primitive erythrocyte generation, and both are required for definitive haematopoiesis [119]. This suggests that BMP signalling pathway is not only important for mesoderm production but the resultant downstream effectors also means it is important for both primitive and definitive haematopoiesis during embryonic development.

The mesodermal stage of development may be identified by the gene expression of number of markers which include *Brachyury* (*BRY*), *Wingless-type MMTV integration site family member 3a* (*WNT3a*), and *Mix1 homeobox-like 1* (*MIXL1*). *Brachyury* is a T-Box DNA-binding transcription factor which asserts its regulation by recognising and binding to the highly conserved 'T-box' sequence TCACACCT, and since its discovery in 1927 in mouse, its vital role in the formation of posterior mesoderm and axial development in vertebrates has been well studied [120]. Beddington *et al.* (1992) [121] found that mice with defective *Brachyury* die *in utero* as a result of notochord formation defects, a reduction in allantois, and absent somites (masses of mesoderm). Studies by Halpern *et al.* (1993) [122] observed that a mutation in the *Brachyury* homology of zebrafish resulted in death shortly after hatching, further analysis reveals that these mutant zebrafish lacked notochords, tails and somites. Homologs of the *Brachyury* gene has been identified in a wide range of vertebrate species including mouse, zebrafish, xenopus, chicken, and human, and in all cases its expression was observed in posterolateral mesoderm and notochord which suggests that its role in development is highly conserved [120]. *Brachyury* is a direct target of *Wingless-type MMTV integration site family member 3a* (*WNT3a*) which studies has shown to be required for paraxial mesoderm development, and its absence during mouse development results in the down-regulation of *Brachyury* [123]. *Wnt3a* is part of the *Wnt* family of 19 genes (human) which activates (ON-state) the canonical Wnt / β-catenin signalling pathway (Figure 6) through binding of cell surface Wnt specific receptors, composed of frizzled (FZ) in a complex with low-density lipoprotein receptor-related protein 5 (LRP5) or low-density lipoprotein receptor-related protein 6 (LRP6). These receptors recruit members of the APC / Axin / GSK-3β destruction complex, and β-catenin is therefore not degraded and is allowed to accumulate before entering the nucleus, whereby it forms a complex with lymphoid enhancer (LEF) / T cell factor (TCF) leading to the regulation of a number of body axis specification and morphogenic signalling genes.

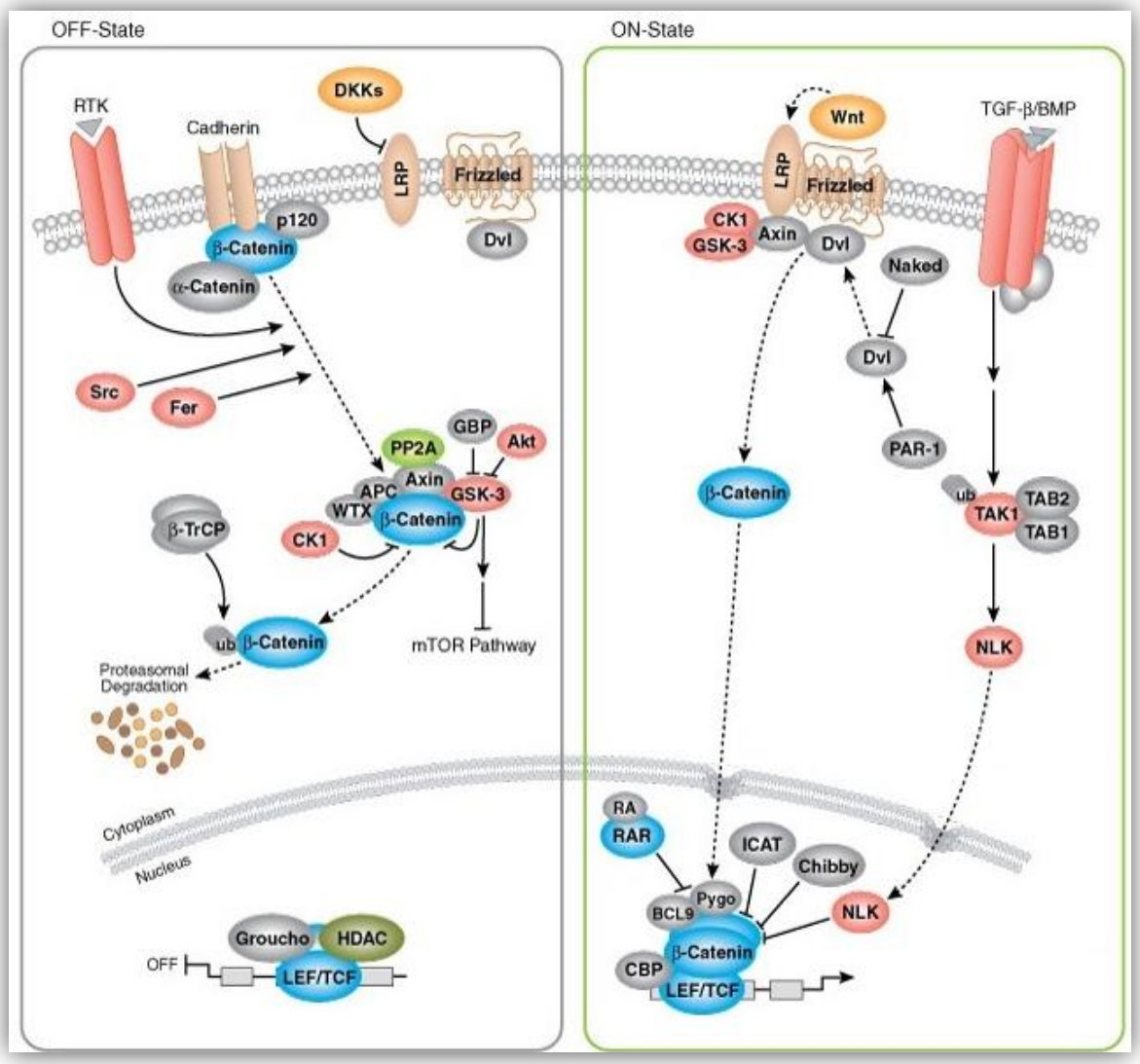


Figure 6: Wnt / Beta-catenin Pathway

The canonical Wnt / β-Catenin pathway is crucial in the regulation of a number of developmental processes including an important role in the control of osteoblast differentiation. In the absence of Wnt (OFF-State), β-Catenin is degraded by the APC/Axin/GSK-3β-complex, but in the presence of Wnt (ON-State), this degradation complex is disrupted by the actions of LRP (binding to inhibit Axin) and Frizzled (binding to activate Dvl). β-Catenin accumulates and enters the nucleus, activating lymphoid enhancer factor (LEF) / T cell factor (TCF) which leads to the activation of a number of developmental genes. This image was adapted and modified from Cell Signalling Technology [124].

The expression of the *mix1 homeobox-like 1 (MIXL1)* gene is another marker of mesoderm formation and studies have shown that *MIXL1* homolog (*Mix.1*) in mouse is required for axial mesoderm morphogenesis and patterning of the embryo [125]. *Mix.1* is a member of the Pax class of homeodomain proteins which function as homodimers or heterodimers in the regulation of dorsal-ventral axial patterning [126]. *Mix.1* over-expression does not induce mesoderm formation but patterns mesoderm from a dorsal to ventral fate in the embryo which is characterised by loss of head structures and increased blood formation. Mead *et al.* (1996) [126] showed that both *Mix.1* and *BMP4* are expressed during gastrulation and both produce ventral axial patterning, but in *Mix.1-/- mutants*, BMP4 signalling alone was

unable to produce a ventral phenotype. The direct injection of Mix.1 partially rescued this phenotype, which suggests that *Mix.1* is the key constituent in the establishment of dorsal-ventral axis, with BMP4 playing an induction role upstream. As previously described, BMP4 is part of the superfamily of TGF-β family of proteins and it was suggested by Mead *et al.* (1996) [126] that BMP4 may regulate the transcriptional activation of *Mix.1* but the mechanisms were not investigated. Studies by Hart *et al.* (2005) [127] revealed that the *Mix.1* promoter was activated by TGF-β type 1 receptor together with the winged-helix / forkhead co-activator FoxH1 which recruits Smad complexes and regulates transcription upon translocation to the nucleus. Further investigation showed that from the TGF-β / Activin signalling pathway (Figure 7), Smad2 and Smad3 in particular acted as positive effectors in the activation of the *Mix.1* promoter. This suggests that the presence of TGF-β / Activin / Nodal ligand may be required in order for the TGF-β / Activin pathway to be initiated. As well as a role in axial mesoderm morphogenesis and patterning, research by Guo *et al.* (2002) [128] has revealed that a homologue of the Mix.1 gene in Xenopus (*MIXL*) could induce expression of the α-globin gene and thus suggests that *MIXL* may also play an essential role in embryonic, foetal, and adult haematopoiesis.

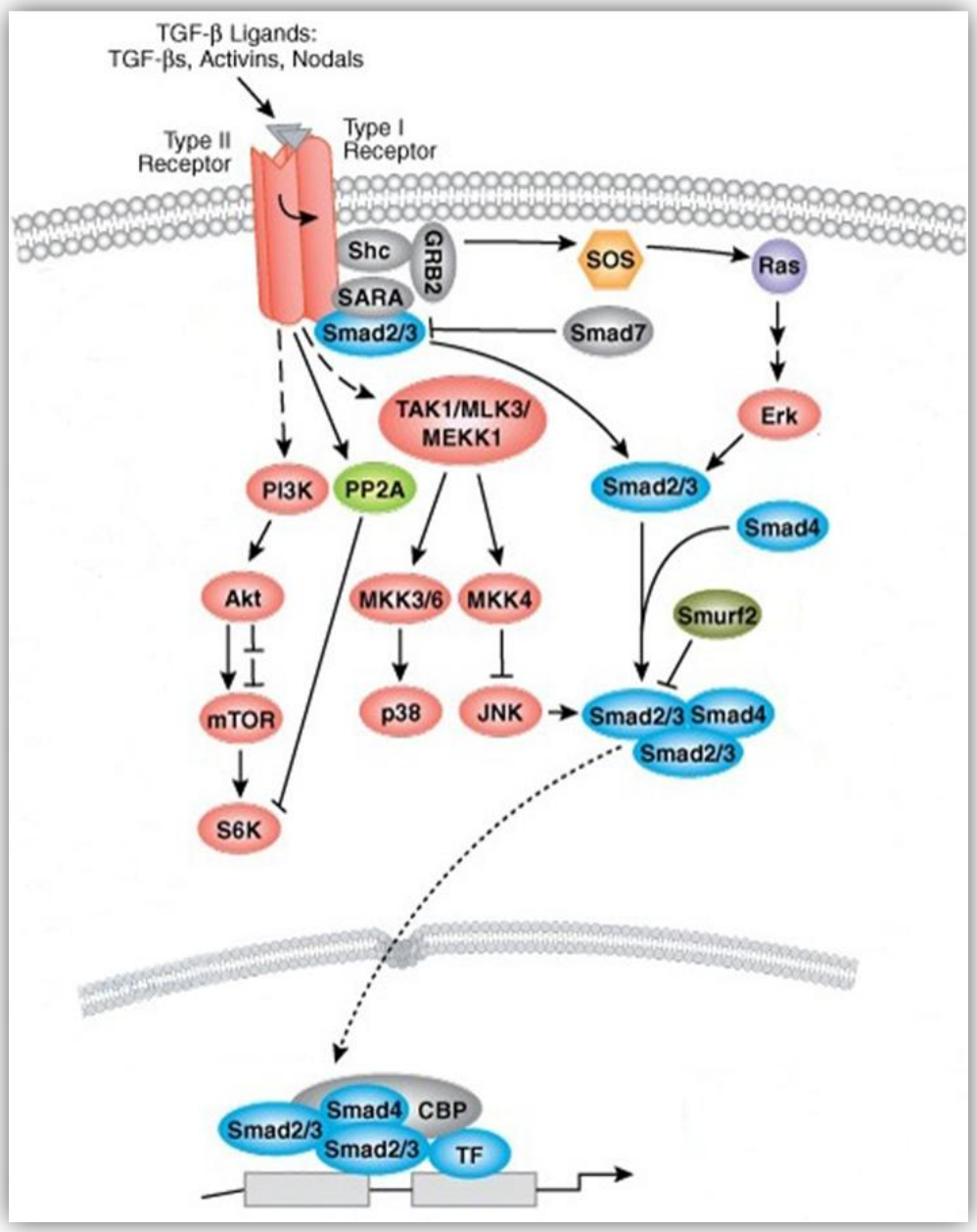


Figure 7: TGF-β / Activin Pathway

From proliferation to differentiation, the TGF-β / Activin pathway plays a crucial regulatory role in a wide range of biological systems. This simplified overview of the TGF-β / Activin pathway shows the activation of Smad2 and Smad3 by an ALK-4 receptor. The common-partner Smad4 binds to Smad2 and Smad3 forming a complex which is then translocated into the nucleus, where it regulates the transcription of a number of genes. This image was adapted and modified from Cell Signalling Technology [124].

1.2.4 The Haematopoietic Niche

Studies in chick and mouse have shown that haematopoiesis begins from the clusters of cells within the yolk sac where lateral and posterior mesodermal cells have migrated to and these were termed the 'blood islands'. The simultaneous emergence of both haematopoietic and endothelial cells from the yolk sac has led to the hypothesis that these may have derived from a common progenitor [129]. The term 'haemangioblast' was used to identify these precursors which appear to cluster together during migration in mouse, where the cells surrounding the peripheral of these cluster appear flatter and over time differentiates to endothelials, whereas the more central cells give rise to a limited number of haematopoietic cell types which include primitive nucleated erythroblasts. The emergence of haematopoietic cells occurs in mammals in two overlapping waves of development which are known as primitive (short-term and limited haematopoietic cell types) and definitive (long-term and gives rise to all haematopoietic cell types) haematopoiesis. The concept of the yolk sac being the origin of the onset of primitive haematopoiesis was widely accepted, but mechanisms of how sites for definitive haematopoietic were colonised have been intensely debated over for the last century. It was suggested that during development, haematopoietic stem cell production migrates sequentially from the yolk sac to multiple niches which include the foetal liver, the spleen, the thymus, and the bone marrow. In support of this theory, Moore and Owen (1965) [130] demonstrated the existence of vascular migration utilising chromosomal labelled cells, and implied that haematopoietic stem cells did not appear *de novo* as in the same way as with the yolk sac, but rather originated and migrated from the yolk sac to colonise the other organ rudiments. This principle of migration from the yolk sac was supported by Houssaint (1981) [131] who demonstrated using mouse liver rudiments taken during early development (before 10 days of gestation) and grafted in chick / quail do not give rise to haematopoietic cells, but liver taken at a later stage of development and after circulatory system had established, resulted in haematopoiesis. This study shows the migration of haemangioblast cells in the colonisation of the foetal liver, but does not directly demonstrate that these haematopoietic stem cell progenitors which populated the foetal liver originated from the yolk sac, and may actually have migrated from an alternative haematopoietic niche (Figure 8).

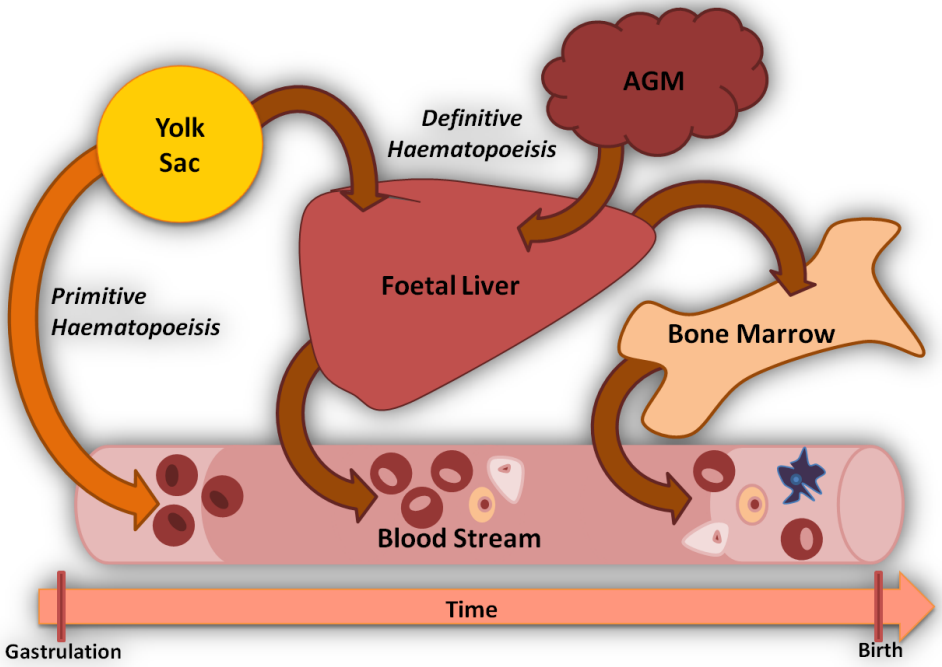


Figure 8: Model of Haematopoietic Niches during Development

The onset of haematopoiesis occurs in two distinct waves with the yolk sac contributing to primitive haematopoiesis and cells migrate to repopulate the foetal liver together with haematopoietic progenitors from the AGM which gives rise to definitive haematopoiesis. The bone marrow becomes the major niche shortly before birth and remains the primary source of haematopoietic stem cells throughout life. Adapted from McGrath et al. (2005) [102]

Dieterlen-Lievre *et al.* (1975) [132] discovered the presence of an alternative niche of haemangioblasts from chick-quail chimera studies, involving the grafting of quail embryos on to chick yolk sacs at various developmental stages and it was detected that the definitive haematopoietic cells types (granulocyte, erythrocyte, lymphocyte) were of quail origin (from the embryo proper). It showed strong evidence that the definitive haematopoietic cells did not originate from the yolk sac and had arisen from an intra-embryonic source. But a possible flaw in this study may be a lack of consideration into the effects of the removal of the quail embryo from its natural environment and its relocation into a bioassay system with differing factors and conditions which in its own way may affect cell fate determination of the organism. A study by de Brujin et al. (2000) [133] shows the presence of the definitive haematopoietic stem cells developing from with the major arterial region of the mouse embryo by dissection of the mouse at 11 d.p.c. This intra-embryonic source of definitive haematopoietic progenitors was later shown to be the paraaortic splanchnopleura (P-Sp) which develops into the aorta, gonads, and mesonephros (collectively known as the AGM) in mouse [134] and human [135]. Cell tracking studies by Kinder *et al.* (1999) [136] showed that mesoderm cells from the anterior region of the primitive streak enter the embryo in formation of the paraaortic splanchnopleura (P-Sp), demonstrating a difference in origin between extra-embryonic cells of

the yolk and the intra-embryonic cells of the P-Sp / AGM. The AGM has been shown to harbour multipotent haematopoietic progenitors and also pluripotent long term renewal haematopoietic stem cells, which are capable of giving rise to all cell types of the adult haematopoietic system [133]. Experiments with chick-quail [132], chick-chick [137], and amphibian [138] embryo grafts also revealed the presence of this intra-embryonic region which gives rise to definitive haematopoietic precursors. Although the AGM is a source for haemangioblasts, its role is however not to provide a niche for the expansion of these cells but rather to seed the foetal liver which will then later seed the bone marrow in the establishment of long-term haematopoiesis [102].

However, the story does not end here as recent studies have shed new light on the possibility that the yolk sac contributes towards adult haematopoiesis [139]. Samokhvalova *et al.* (2007) [139] undertook long-term *in vivo* tracing of yolk sac cells and observed the colonisation of the umbilical cord, the AGM region and ultimately the liver. However, Palis *et al.* (1999) [140] noticed that the *C-myb* gene which is essential for definite haematopoiesis, was detected (albeit at low levels) in the mouse yolk sac just prior to the emergence of definitive erythroid progenitors (colony forming unit – erythroid). The appearance of definitive erythroid progenitors from the yolk sac suggests that definitive haematopoietic progenitors arise from the yolk sac and then migrates through the bloodstream to seed and populate in other rudiments such as the foetal liver / bone marrow. These studies have shown that under the right conditions, the yolk sac can give rise to cells that possess the essential properties of haematopoietic stem cells. It seems logical that there is an overlap between the primitive and definitive haematopoiesis within the yolk sac, and it has been suggested that yolk sac progenitors might acquire functional characteristics of haematopoietic stem cells in the AGM region, however the strength and potency of the yolk sac's participation in the establishment of the adult haematopoietic system remains to be determined [139].

In conclusion, the combination of these studies have nevertheless illustrated that the site of haematopoietic development is not fixed, and during the process of development haematopoiesis occurs in a number of sequential niches in two distinct waves which begin from their short term stay at the blood islands of the yolks sac (primitive) to their long term residence in the bone marrow (definitive).

1.2.5 The Haemangioblast

1.2.5.1 The Mouse Haemangioblast

The concept that a common mesodermal stem cell was responsible for the generation of haematopoietic and endothelial cells was hypothesised almost a century ago by Sabin *et al.* (1920) [141]. It was noted that both haematopoietic and endothelial cells share expression of a number of different genes and this close developmental association within the blood islands has reinforced the hypothesis of a common precursor [142, 143]. Choi *et al.* (1998) [144] described a transient cell population derived from differentiating mouse embryonic stem cells (MESCs) which in response to vascular endothelial growth factor (VEGF) gives rise to blast colonies. These blast colonies in liquid culture together with the appropriate cytokines (such as VEGF) are able to give rise to both haematopoietic and endothelial cell types. These observations provide strong evidence that these blast colonies may represent the *in vitro* equivalent of the long hypothesised haemangioblast. Progressive lineage analysis by Nishikawa *et al.* (1998) [145] defined intermediate stages of differentiation from MESCs to blood cells using a combination of 6 monoclonal antibodies (Table 3), and identified cells which are FLK1+ VE-cadherin+ CD45- as at the diverging point of haematopoietic and endothelial lineages.

Developmental Stage	Cell Surface Markers
Mouse Embryonic Stem Cells	E-cadherin+ Flk1- PDGFRalpha-
Proximal Lateral Mesoderm	E-cadherin- Flk1+ VE-cadherin-
Progenitor with Haemoangiogenic Potential	Flk1+ VE-cadherin+ CD45-
Haematopoietic Progenitor	CD45+ c-Kit+
Mature Blood Cells	c-Kit- CD45+ or Ter119+

Table 3: Developmental stages from MESC to Blood Cells

Identification of cell types at the various development stage using cell surface antigens as reported by Nishikawa *et al.* (1998) [145].

1.2.5.2 Markers of the Mouse Haemangioblast

FLK1 is a kinase insert domain receptor (a type III receptor tyrosine kinase) which encodes one of three receptors of vascular endothelial growth factor (VEGF), and functions to mediate endothelial proliferation, survival, migration, tubular morphogenesis, and sprouting [146]. As with many genes with multiple synonyms, it can be rather confusing especially when attempting to follow multiple genes across interspecies studies, for example, *FLK1* is used in mouse studies, whereas studies in chick / quail refers to this gene as *VEGFR2*, and studies in human utilises *KDR.* However, herein the synonym, *KDR* will be used to describe this gene in all species for consistency. *KDR* is the receptor for the ligand VEGF-A (but also binds VEGF-C and VEGF-D) which is a member of the vascular endothelial growth factor (VEGF) family of homodimeric glycoproteins consisting of VEGF-A, VEGF-B, VEGF-C, VEGF-D, and PLGF (placenta growth factor) [147]. These VEGF family of ligands target and bind to cell surface receptors VEGFR1, KDR (VEGFR2), and VEGFR3 which studies have implicated roles in the activation of a number of signalling pathways (Figure 9) including protein kinase c pathway (transcription regulation, cell proliferation, immune response), nitric oxide synthase pathway (vasopermeability), AKT pathway (cell growth, cell survival, cell cycle progression, differentiation, transcription, translation, glucose metabolism), Paxillin (cytoskeletal rearrangement and cell migration), and MAPK/ERK pathway (cell migration, proliferation, and division) [148, 149].

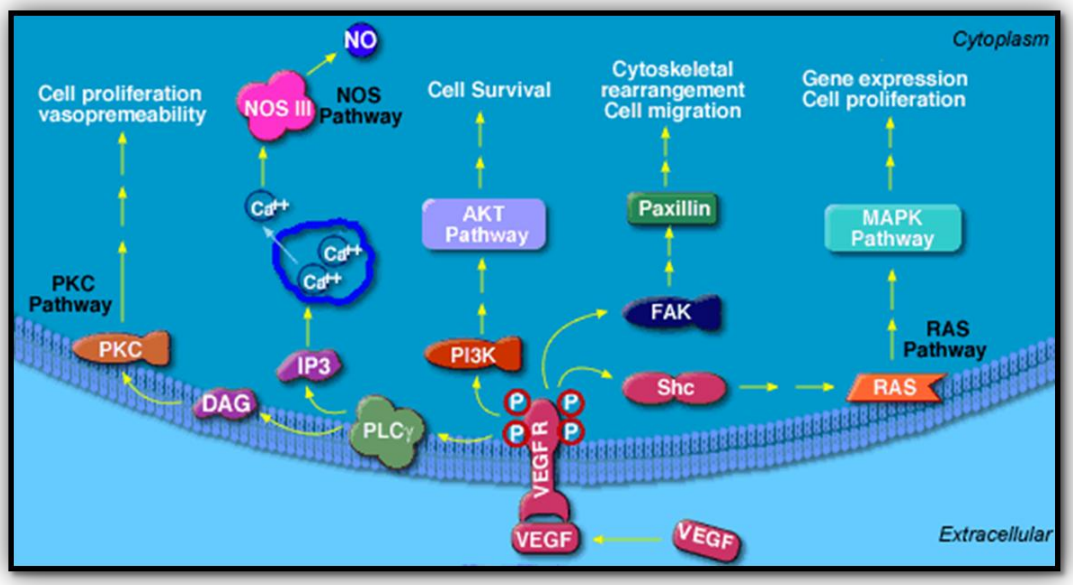


Figure 9: VEGF/KDR Signalling Pathway

Upon induction of VEGF by hypoxia, VEGFR is activated which in turn regulates a number of pathway which include the protein kinase c pathway (transcription regulation, cell proliferation, immune response), nitric oxide synthase pathway (vasopermeability), AKT pathway (cell growth, cell survival, cell cycle progression, differentiation, transcription, translation, glucose metabolism), Paxillin (cytoskeletal rearrangement and cell migration), and MAPK/ERK pathway (cell migration, proliferation, and division). This pathway was adapted and modified from Biocarta Pathways [150].

KDR gene disruption studies by homologous recombination results in the death of the mouse embryo between 8.5 to 9.5 d.p.c. due to the absence of yolk sac blood islands and organised blood vessels [151]. Shalaby *et al.* (1997) [152] revealed that *KDR* may play a role in the movement of cells from the primitive streak to the yolk sac, as *KDR* -/- cells are shown to be unable to migrate to the correct location for the formation of blood islands. During development, Quinn *et al.* (1993) demonstrated by *in situ* hybridisation that the level of *KDR* was restricted to the vascular endothelium and umbical cord stroma, and suggests a role for KDR in vascular development. These findings strongly suggest that as the cell surface receptor KDR is crucial for the development of both haematopoietic and endothelial cells, this in turn suggests that the downstream pathway is regulated by the receptor ligand VEGF. As a result, VEGF has been included in the differentiation media used within our study as an essential component in the specification of haematopoietic development.

The surface antigen VE-cadherin (also known as CD144 and CDH5) is a calcium-dependent cell to cell adhesion glycoprotein from the cadherin superfamily and is comprised of five extracellular repeats, a transmembrane region, and a highly conserved cytoplasmic tail [153]. Studies has suggest that VE-cadherin plays an important role in endothelial cell biology enabling cells to adhere in a homophilic manner and maintaining cohesion of the intercellular junctions [153]. Targeted inactivation of the *VE-cadherin* gene by Carmeliet *et al.* (1999) [154] resulted in lethality at 9.5 days of mouse gestation, and showed that *VE-cadherin* was required not for the assembly but for the remodelling and maturation of endothelial cells. *VE-cadherin* was also found to control endothelial survival as down-regulated *VE-cadherin* resulted in endothelial apoptosis and affected KDR signalling [154].

Cluster of differentiation 45 (CD45) is encoded by the *protein tyrosine phosphatase receptor type C (PTPRC)* gene, which is a type I transmembrane protein that is specifically present on all differentiated haematopoietic cells (except erythrocytes and plasma cells) [155]. CD45 is a member of the protein tyrosine phasphatase (PTP) family which are composed of signalling molecules that regulate a range of cellular processes including mitotic cycle, cell growth, and differentiation.

1.2.5.3 The Human Haemangioblast

The mouse embryoid (EB) system was adapted and used in the differentiation of human embryonic stem cells (HESCs), a few groups have successfully generated haematopoietic and endothelial cells in a temporal pattern that recapitulates early haematopoietic development *in vivo* [156]. Utilising comparative reverse transcriptase-mediated polymerase chain reaction, Keller *et al.* (1993) [156] demonstrated that gene markers for primitive endoderm and mesoderm were present in the developing EB and that this *in vitro* system accurately recapitulates haematopoietic development *in vivo*. The heterogeneity of the EB which contains a multitude of cell types of the various stages of development make identification very challenging, and has thus far heavily relied on the gene expression signatures and surface antigen markers. Kennedy *et al.* (2007) [157] identified CD34, CD31, KDR and VE-Cadherin positive cells from within human EBs which also display both haematopoietic and vascular potential. It was shown that these markers identified a transient population that appears in BMP4 induced human EBs between 72 hours (3 days) and 96 hours (4 days) of differentiation which give rise to primitive erythroid cells, macrophages, and endothelial cells. These findings by Kennedy *et al.* (2007) [157] pinpointed the earliest stage of haematopoiesis through the identification of the bipotent progenitor in human development, the haemangioblast.

1.2.5.4 Markers of the Human Haemangioblast

The haemangioblast markers specified by Nishikawa *et al.* (1998) [145] which included FLK1+, VE-cadherin+, and CD45-, are very similar to those identified by Kennedy *et al.* (2007) [157] for the human haemangioblast albeit with the addition of CD31 and CD34.

The platelet / endothelial cell adhesion molecule (PECAM1; is also known as cluster of differentiation 31 (CD31)) is a 130-kD member of the immunoglobulin superfamily that is normally expressed on the cell surface of endothelial cells, platelets, macrophages, granulocytes, T-lymphocytes, natural killer cells, megakaryocytes, osteoclasts, and neutrophils [158]. Studies have shown that the cell adhesion receptor CD31 is involved in diverse roles in vascular biology including angiogenesis, platelet function, thrombosis, haemostasis, immunity by leukocyte regulation, and inflammatory response [159]. A study by Furuta *et al.* (2006) [160] involving specific staging and dissecting of mouse embryos through development identified CD31+ cells (at 7.5 d.p.c). with the ability to give rise to haematopoietic and endothelial cells, whereas CD31- cells could only give rise to endothelial cells, which suggests an important role for CD31 in haematopoiesis.

The cluster of differentiation 34 (CD34) protein is a member of the transmembrane sialomucin proteins and similarly to CD31, it is also expressed on early haematopoietic and endothelial cells. Unfortunately little is known as its exact role or function, although it was been used for many years to distinguish early progenitor cells from quiescent cells, and also as a crucial marker during clinical haematopoietic stem cell transplants [161]. As well as being expressed on haematopoietic stem cells, CD34 is also expressed on mesenchymal stem cells (non-haematopoietic cells), and vascular endothelia in all organs with especially high expression on small vessels [142]. As mentioned previously, the CD34 protein is only expressed on haematopoietic / endothelial precursors and not on differentiated haematopoietic lineages although it does remain on endothelial cells in the adult [162]. In conjunction with CD31, these dual markers of expression of CD31 and CD34 will contribute to the identification of the earliest stages of haematopoietic / endothelial development.

1.2.5.5 Challenging the Haemangioblast Concept

Although there appears to be ample evidence of the existence of bipotential progenitor, some studies have challenged the haemangioblast concept, with Ueno *et al.* (2006) [163] claiming that in the mouse model *in vivo* the majority of haematopoietic cells are generated from KDR- cells, whilst endothelial cells are principally from KDR+ cells. Lineage tracing studies by Ueno *et al.* (2006) [163] using MESCs marked with fluorogenic protein injected into mouse blastocysts showed that during development, haematopoietic and endothelial cells produced were of different clonal origins of these chimeric mice. It was shown that the yolk sac blood islands were derived from multiple clonal precursors and that the majority of haematopoietic cells are not derived from KDR+ precursors. However this study suffers a number of flaws that were not suitably accounted for, which include failing to select for stably transfected clones resulting in the possible inclusion of non-transfected MESCs, and the exact composition and quantity of transfected MESCs in the chimeric embryo was not discussed. Hence in contrast to this, a more recent and more efficient lineage tracing study by Lugu *et al.* (2009) [164] permanently marked KDR+ cells using the Cre/loxP system (in the generation of *KDR+/Cre,Rosa26R-EYEP* mice) and observed their progenies during development. Fluorescence-activated cell sorting analysis revealed that at 9.5 d.p.c. 97% of erythroid cells (using Ter119 as a marker) and 97% of macrophages (Mac1 as a marker) were positive for EYFP expression. It was also observed that all CD45+ (marker of haematopoietic cells) were also positive for EYFP and the CD45+ population represented 2.7% of the yolk sac cells. This study concludes in favour with the general opinion within the current literature that most, if not all,

haematopoietic cells, primitive and definitive, (and endothelials) in mice are the progeny of KDR+ cells.

Other studies has suggested that the potential of the haemangioblast may not be solely restricted to just haematopoietic and endothelial cells. Yamashita *et al.* (2000) [165] and Ema *et al.* (2003) [166] demonstrated that KDR-expressing cells obtained from EB differentiation in mouse can also develop into vascular smooth muscle (Figure 10). These findings do not disprove the existence of the haemangioblast but rather implies that the haemangioblast population can not identified by KDR alone. It was suggested that KDR-expressing cells represent a mesodermal progenitor which can give rise to either haemangioblasts (haematopoietic and endothelial) or to angioblast (endothelial and smooth muscle) and this appears to be governed by *Tal1/SCL* expression [166].

The *T-cell acute lymphocytic leukaemia 1* (*Tal1*) gene is a basis-helix-loop-helix (bHLH) transcription factor, which gene ablation studies have revealed is involved in key roles in primitive haematopoiesis, vasculogenesis, angiogenesis, definitive erythroid and megakaryocytic development [167]. *Tal1* is also know as *stem cell leukaemia* (*SCL;* hereafter *SCL* will be used), and this gene was first implicated in human T cell acute lymphoblastic leukaemia (T-ALL) caused by its over-expression due to chromosomal rearrangements [168]. SCL is a member of the basic-helix-loop-helix (bHLH) family of proteins which dimerises with bHLH proteins E47 / E12 through the HLH domain and regulates gene expression by binding to E-box motif (CANNTG) on the DNA. Targeted disruption of the *SCL* gene in mouse have resulted in the complete obliteration of yolk sac haematopoiesis leading to death at 9.5 d.p.c. suggesting an importance in primitive haematopoiesis [169, 170]. This finding is further confirmed up by studies with SCL-deficient EBs from MESCs which resulted in the absence of blast colony formation, and reiterated the importance of SCL for the progression to haematopoietic commitment [171]. Ema *et al.* (2003) [166] describes the importance of SCL in the regulation of cell fate and have shown that increased *SCL* expression in KDR-expressed cells results in the inhibition of smooth muscle production. The hierarchical model of the differential expression of *KDR* and *SCL* in Figure 10 shows that in the absence of SCL, KDR-expressed mesodermal progenitor cells will give rise to angioblasts (KDR+SCL-) and may differentiate to smooth muscle (KDR-SCL-).

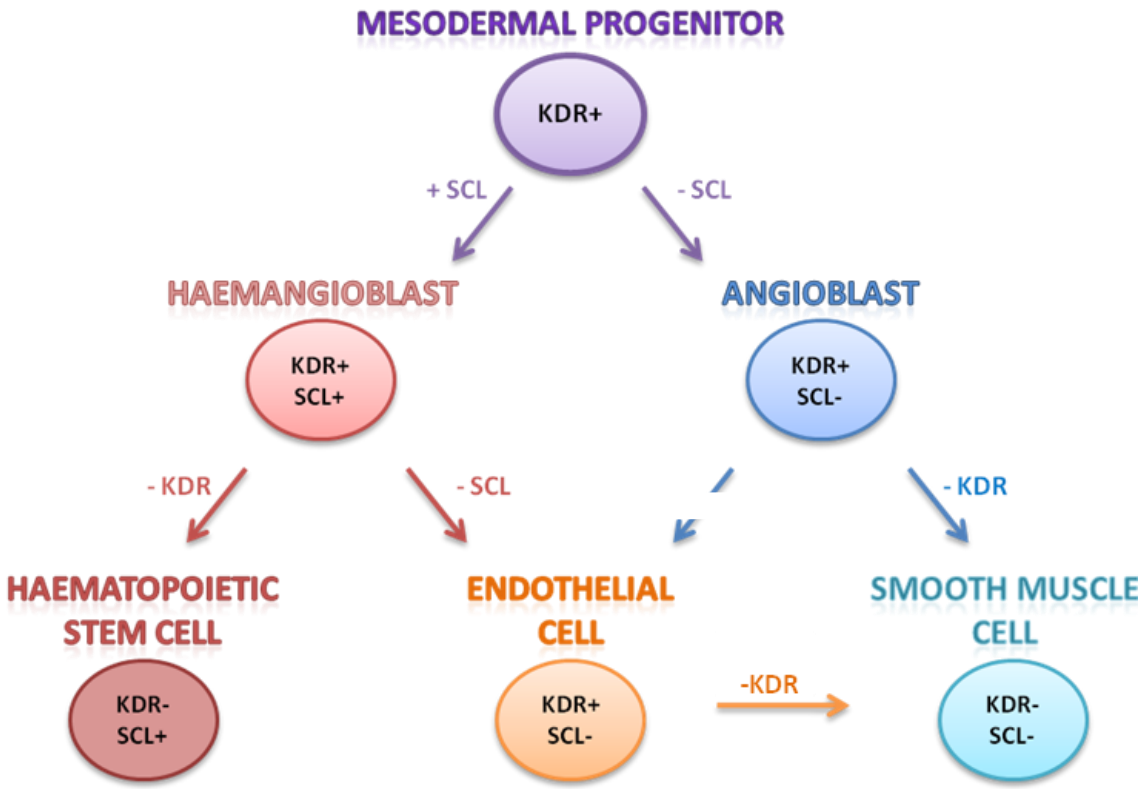


Figure 10: Model of KDR and SCL expression in early development

This hierarchical model of the simplifies the general roles of KDR and SCL during haematopoietic and vascular development shows that only KDR+ and SCL+ cells may give rise to haematopoietic stem cells, but shows that the production of endothelial cells may be from KDR+SCL+ or KDR+SCL- cells. In the absence of SCL, KDR+ cells of the angioblast population as well as endothelial-like cells may give rise to smooth muscle cells. This figure was reproduced and adapted from Ema *et al.* (2003) [166].

1.2.5.6 The Adult Haemangioblast

Although the haemangioblast model at embryonic haematopoietic development appears to be quite well established, the concept of a similar bipotential precursor in the adult is relatively new [172]. Studies have suggested that there may be a common precursor involved in blood and blood vessel production, and this was termed the adult haemangioblast. However this population does not appear to exist in large numbers (1-5% in the bone marrow), with varying contributions to the arrangement of blood vessels [173], and hence questions the important of the role of this adult haemangioblast population in normal physiology and pathophysiology [174].

1.2.6 Mesoderm to Haemangioblast

Despite the intense research into haematopoietic development, the underlying molecular mechanisms of how the haemangioblast is formed from mesodermal cells remains largely unknown. However, though gene knockout and over-expression studies, a number of key haemangioblast development genes have been identified to be important during this stage which includes *KDR* and *SCL*. As previously described, genes *KDR* and *SCL* appears to play a crucial role during mesoderm to haemangioblast transition, however the precise mechanisms by how these key regulators function during this stage of development and how they interact with other genes within a network and / or signalling pathway remains largely unknown. A tracking study by Fehling *et al.* (2003) [143] of mesoderm development (by the addition of a GFP reporter construct to the *Brachyury* (*Bry*) promoter) identified the emergence of three distinct populations which represents the stages of mesoderm differentiation from pre-mesoderm (Bry-KDR- cells), to pre-haemangioblast mesoderm (Bry+KDR- cells), and to haemangioblast (Bry+KDR+ cells). In recent decades, angiogenesis has been extensively studied due to the critical role it plays in the development of cancer and tumour progression. The formation of new capillaries and vessels to the tumour greatly facilitates growth through the delivery of oxygen and nutrients (and removal of waste products) which rapidly switches the tumour from a slow state of cellular multiplication to a state of rapid uncontrolled proliferation [175]. Therefore, these studies in angiogenesis have provided some insight on the effects of the VEGF / KDR pathway (Figure 11) of which plays an integral role in capillary formation from pre-existing blood vessels during early development.

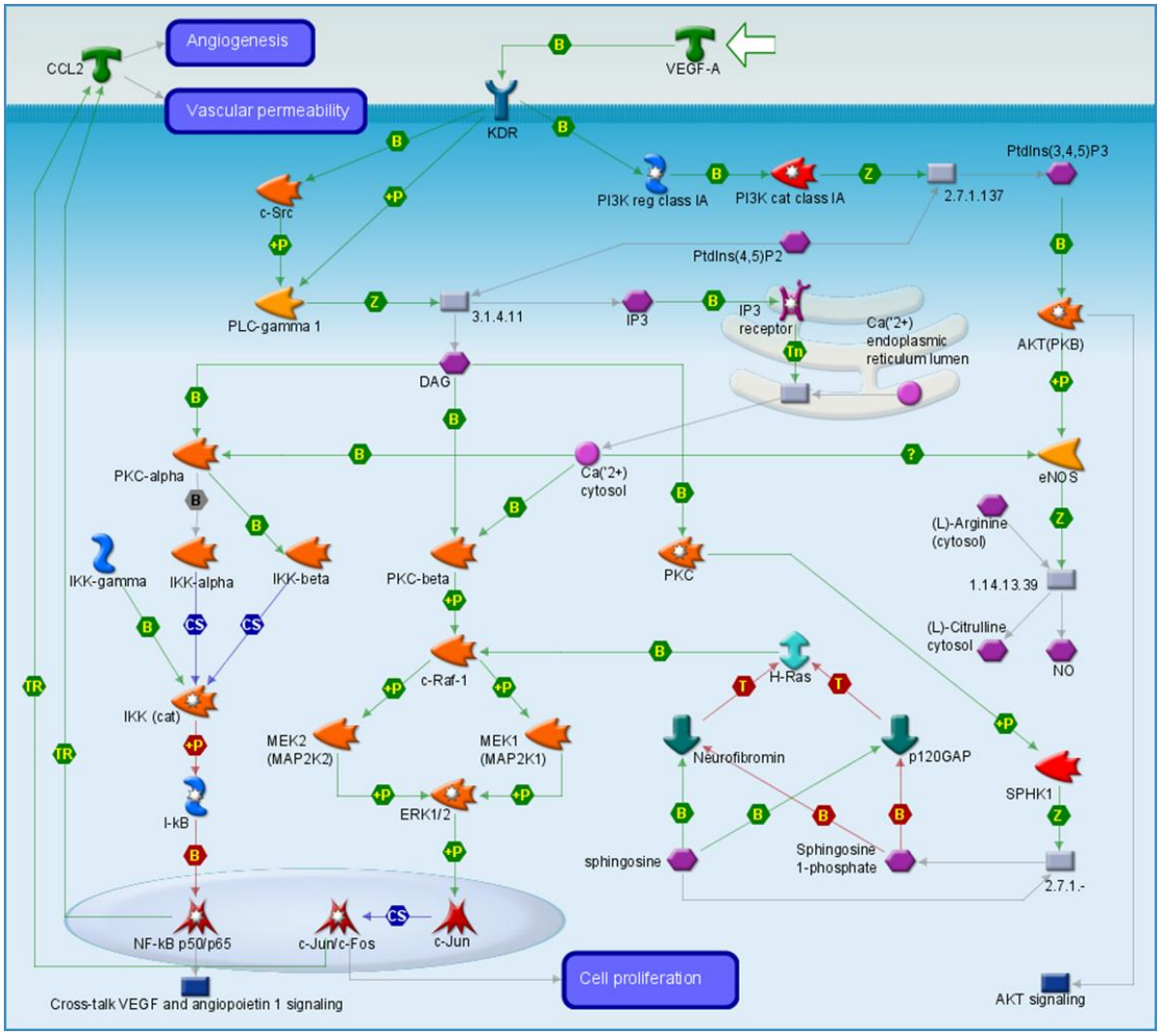


Figure 11: VEGF signalling though KDR (VEGFR2)

KDR is activated by the ligand VEGF-A, which in turn activates a number of signalling pathways including phosphoinsositide-3-kinase regulatory subunit (PI3K reg class IA) leading on to the activation of V-akt murine thymoma viral oncogene homolog 1 (AKT) which plays a role in cellular survival through the inhibition of apoptotic pathways. The transcriptional activation of nuclear factor of kappa light polypeptide gene enhancer in B-cells and V-reticuloendotheliosis viral oncogene homolog A (NF-kB p50/p65) and jun oncogene (c-Jun) / v-fos FBJ murine osteosarcoma viral oncogene homolog (c-Fos) results in the activation of chemokine ligand 2 (CCL2) which promotes angiogenesis and vascular permeability, and c-Jun / c-Fos activation also results in DNA synthesis and cell proliferation. (referenced and adapted from Metacore Pathways [176])

The presence of the VEGF-A ligand together with KDR expression is required for angiogenesis, as it seems that the presence of KDR during the haemangioblast stage appears to be readying the bipotential cells for vascular development. Studies to date have not suggested a specific role for KDR during the transition from mesoderm to haemangioblast, but rather acts as a marker which is only later required for the proceeding stages of development.

A role for *SCL* in embryonic development was highlighted by Gering *et al.* (1998) [177] where by it was observed that *SCL* over-expression in the zebrafish embryo resulted in the marked increase of haematopoietic and endothelial lineages albeit at the expense of somatic and pronephric duct tissue. This suggests that *SCL* functions in the specification of the mesoderm to form the haemangioblast, however a later study by D'Souza *et al.* (2005) [178] showed that the majority of haemangioblasts do not express *SCL*, and implicated that *SCL* expression is not essential for the formation of the haemangioblast but rather for its differentiation to haematopoietic and endothelial lineages. More recent studies have used chIP-Seq (coupling chromatin immunoprecipitation with deep sequencing) technology to generate a SCL-binding network of target genes in providing valuable insight into the genomic biological function and role of SCL through development [179]. The SCL target genes (228 binding events) were highly enriched for haematopoietic transcription factors and involved in signal transduction, and of these a number of genes (*Runx2, Hhex, Hfe2, Mafk, Cebpe, Cbfa2t3h, Gfi1b, Runx1, Klf2, Erg, Fli1, Lyl1, Gata2, Zfpm1,* and *Tox2*) have been validated through *in vivo* experiments (Figure 12). This list contains a number of genes (*Gata2, Fli-1, Runx1*) from which literature has already implicated to be targets of SCL regulation and are expressed in haematopoietic or endothelial cells [179].

Numerous studies [143, 180-185] have suggested that the haemangioblast is marked by the expression of a number of genes which include *KDR, CD31,* and *CD34* which suggest that these may play a role in the formation for this progenitor or in preparation for proceeding stages of haematopoietic / endothelial differentiation. The roles and interactions of *CD31* and *CD34* during haemangioblast formation has not been well described in literature but their role during later stages of development especially in the establishment of contact and maintenance of tight, adherens, and gap junctions between endothelials cells is of critical importance during vasculogenesis and angiogenesis [186, 187].

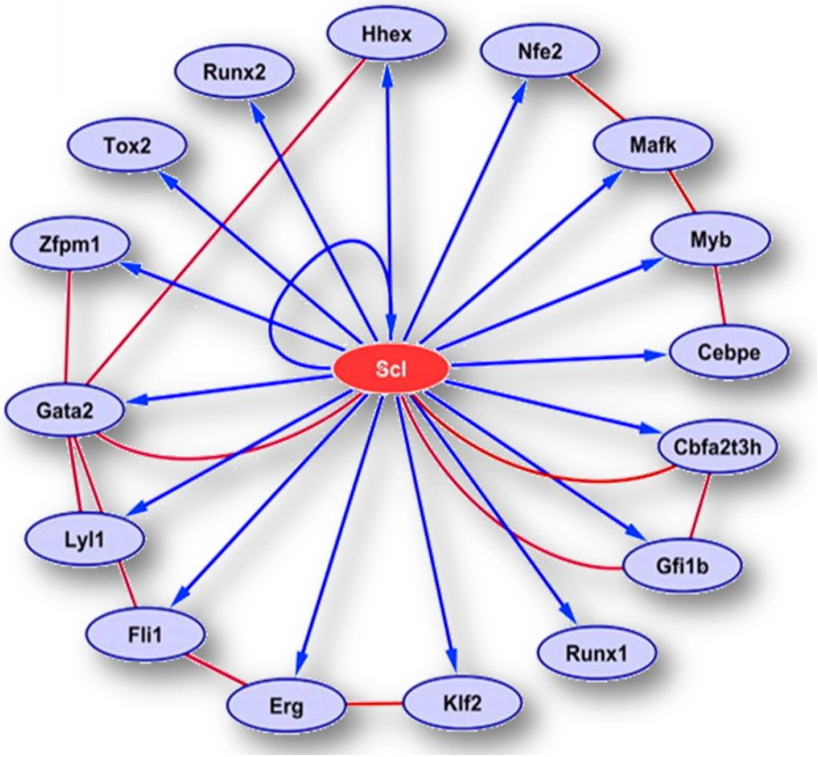


Figure 12: SCL Regulatory Network

A regulatory network produced by Wilson et al. (2009) [179] using ChIP-Seq technology shows 16 genes which were found to be targets of SCL interaction and validated through *in vivo* studies. Blue arrows shows interactions with SCL that were validated through transgenic mice studies and red arrows showing interaction that were obtained from published literature. (Adapted from Wilson et al. (2009) [179])

1.2.7 The Haemogenic Endothelium

The haemogenic endothelium was first hypothesised when it was noticed that within the AGM region, newly generated haematopoietic stem cells were found to be closely associated with those forming the walls of the blood vessel of the aorta [188]. As previously described, the origin of definitive haematopoietic stem cells were from the embryonic aorta-gonad-mesonephros (AGM) region. Pardanaud *et al.* (1996) [189] identified within chicks, cells of the ventral aortic endothelium have dual haematopoietic and endothelial potential. However, this phenomenon was restricted only to cells that constitute the floor of the aorta, with it being reported that the aortic endothelium is comprised of a mosaic origin of cells, with the roof and sides contributed by the somites and the floor by the splanchnopleure. Through quail / chick transplant experiments, it has been shown that endothelial cells originate from two different mesodermal lineages, with the first (somites) only giving rise to endothelial cells, and the second (splanchnopleure) giving rise to both endothelial and haematipoietic stem cells [189]. The differences in potential between these two endothelial progenitors may be connected to a study by Furuta *et al.* (2006) [160] who showed two endothelial progenitor populations, which

one was CD31+ and gives rise to endothelial and haematopoietic cells, and the other CD31- and only gives rise to endothelial cells. From these results, it appears that the cell surface marker CD31 may be used in the identification of haemangioblasts from other cell types including angioblasts (endothelial-restricted progenitors). It also questions the widely accepted concept that all endothelial cells are derived from the bipotent haemangioblast together with haematopoietic cells, and suggests that angioblasts may have been derived independently from the mesoderm prior to the appearance of haemangioblasts.

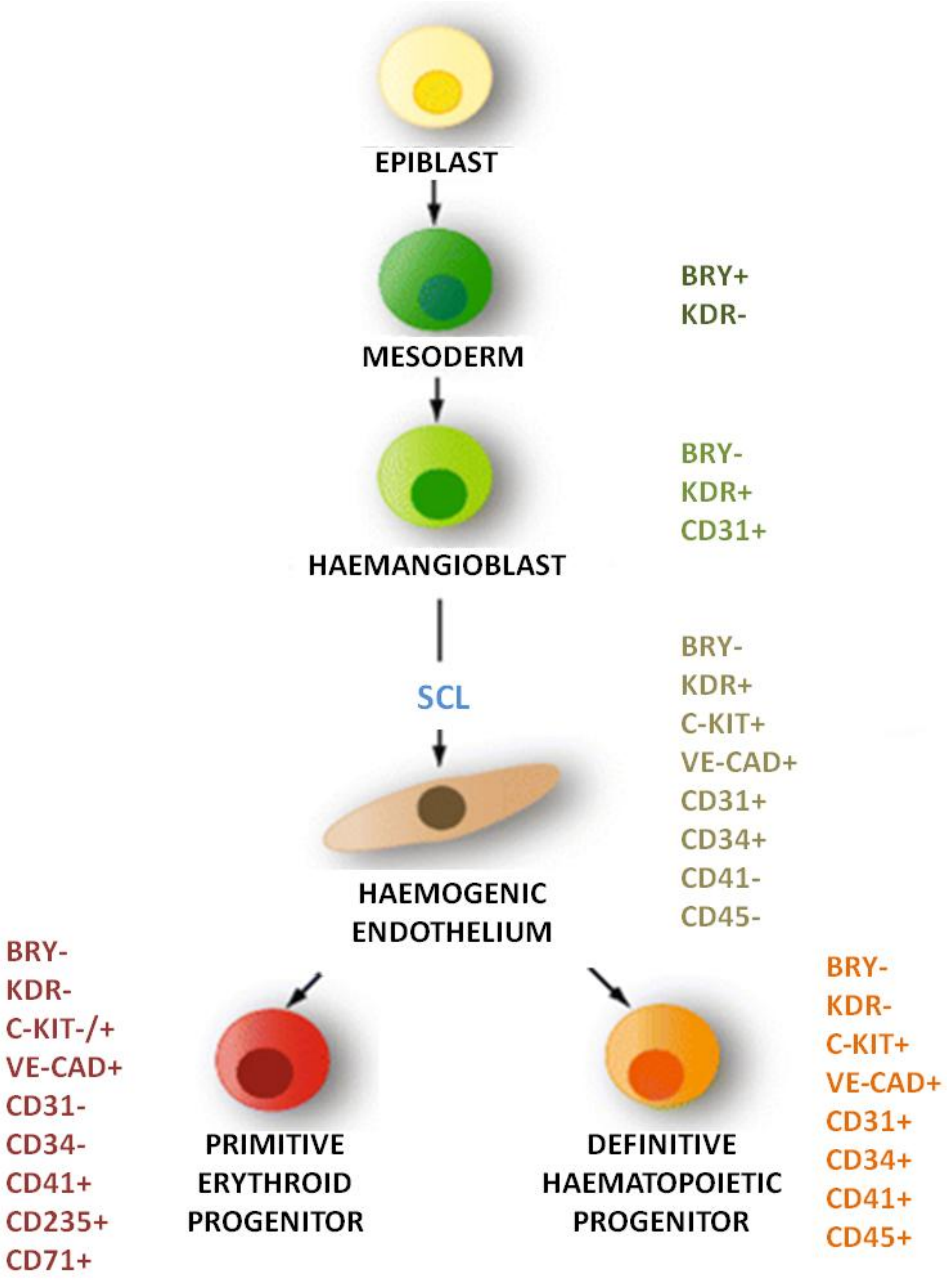


Figure 13: Model of Haemangioblast Formation and Progeny

The mesoderm is marked by the presence of BRY and the formation of the haemangioblast is marked by expression of KDR and CD31. In the presence of the haematopoietic transcriptional regulator, SCL, the haemogenic endothelium is formed and is marked by a number of antigens including KDR, CD117, CD144, CD31, and CD34. Hemogenic endothelium appears to be regulated by a number of transcription regulators and signalling pathways (Notch1, Runx1, and Wnt) in giving rise to cells that make up primitive and definitive haematopoiesis. (Reproduced and adapted from Lancrin *et al.* (2010) [190]).

Although the hypothesis that the haemangioblast does not directly give rise to haematopoietic cells but rather to an endothelial intermediate, the haemogenic endothelium, which was proposed by Pardanaud *et al.* (1996) [189]. However, It was only recent that this hypothesis was proven *in vitro* by Jaffredo *et al.* (2000) [191] who traced the progeny of the aortic haemangioblast in avian embryos and demonstrated the existence of such a cell type. Eilken *et al.* (2009) [192] used a novel cell tracking method whereby single mesodermal cells of the mouse was continuously observed over long-term, and it was showed that embryonic endothelials can be haemogenic. These results suggest that the generation of haematopoietic cells occurs through the formation of a haemogenic endothelial intermediate cell type from the haemangioblast (Figure 13). Lancrin *et al.* (2009) [193] also supported these findings and utilising time-lapse photography they identified the emergence of non-adherent round cells budding off from a tight adherent endothelial sheet, and these non-adherent cells were positive for the CD41 antigen which specifies haematopoietic commitment. It appears from these studies, that the haemogenic endothelium concept is valid and this is derived only from splanchnopleural cells, but ectopic grafts studies by Pardanaud *et al.* (1996) [189] demonstrates that the niche too plays a very important role in the haematopoietic potential of these cells. It was shown that these grafts only produced haematopoietic clusters when they were located to the aortic floor, which is tightly associated with the endoderm, and suggests that signals from the endoderm may be required for haematopoietic differentiation.

Pardanaud and Dieterlen-Lievre (1999) [194] showed through chick / quail grafting studies that the transient contact of mesodermal cells with those of the endoderm promoted production of haemogenic endothelial cells even those from the somatopleural (which would normally only give rise to endothelial cells). It was also demonstrated that this feat could be mimicked by the addition of a cocktail of cytokines during mesoderm differentiation, which include vascular endothelial growth factor (VEGF), basic fibroblast growth factor (bFGF), and transforming growth factor beta 1 (TGFβ1). However, in contrast, contact with the ectoderm or the addition of epidermal growth factor (EGF) and transforming growth factor alpha (TGFα) inhibited the haemangiopoietic potential of the splanchnopleural mesoderm. In the search for haemangioblast induction factors of the endoderm during development, Belaoussoff *et al.* (1998) [195] digested sections of primitive mouse endoderm and ectoderm, and discovered that the Indian hedgehog (Ihh) factor alone was able to generate haematopoietic and endothelial cells in their assays. It was later shown by Zhang *et al.* (2001) [196] that the active expression of Ihh was actually not essential as Ihh-/- mutant embryos do still exhibit haematopoietic and endothelial lineages albeit with vascular remodelling defects. Ihh signalling appears to regulate *SCL* expression through interactions with the *SCL* 3' enhancer

region during haemangioblast specification. SCL has already been described as being important for the specification of the haematopoietic stem cell and the differentiation into subsequent primitive and definitive lineages [166, 169-171].

1.2.8 Haemogenic Endothelium to Haematopoietic Cells

1.2.8.1 Primitive Haematopoiesis

The concept that the yolk sac is the sole niche and source of primitive haematopoiesis has been well established over a century ago beginning with the discovery of the 'blood islands' which are mesodermal cell aggregates that give rise to angioblasts (vascular endothelial cells) and haematopoietic cells [197]. These haemangioblasts from the blood islands of the yolk sac will give rise to blast colony types that are restricted to primitive haematopoietic potentials, which predominantly include erythrocytes as shown in Figure 14. Large primitive nucleated erythroblasts are one of the first blood cells that are generated from the blood islands, which mature in circulation into primitive erythrocytes and represent the predominant population during the yolk sac stage of haematopoiesis [198]. The mechanisms behind early primitive erythropoiesis have not been very well understood but there appears to be tightly regulation and control due to the transient nature of this stage of development. A number of factors have been implicated in the regulation of yolk sac haematopoiesis that include VEGF ligand / KDR receptor and TGFβ1, as previously described play an important role in the establishment of haematopoietic and vascular lineages [152, 199]. A recent study by Cheng *et al.* (2008) [200] has implicated *Notch* and *Wnt* as important regulators of early primitive erythropoiesis. The Notch signalling pathway is composed of four different Notch receptors (NOTCH1, NOTCH2, NOTCH3, and NOTCH4), located on the cell surface of cells and functions through cell-to-cell contact as the ligands are also transmembrane proteins. Following activation, Notch is cleaved intracellularly and the liberated intracellular part of the protein migrates into the nucleus where it associates with a number of DNA-binding sites in the regulation of gene expression [201]. *Notch* knockout studies by Hadland *et al.* (2004) [202] showed that Notch plays an inhibitory role during primitive erythropoiesis, as *Notch*-/- ES cells and mutant embryos resulted in enhance primitive erythroid potential. Wnt signalling however appears to have the opposite effect to the Notch pathway, as demonstrated by Nostro *et al.* (2008) [203] when the Wnt pathway was blocked, primitive erythroid development was

significantly reduced in mouse ES differentiation. Cheng *et al.* (2008) [200] investigated the relationship between Notch and Wnt in the specification of early haematopoietic development, and concluded that activation of the Wnt canonical signalling pathway in conjunction with the inhibition of the Notch signalling pathway by Numb was required for primitive erythroid specification.

The erythroid transcription factor, GATA-binding 1 (GATA-1) is essential for primitive haematopoietic as studies have shown that plays an important role in erythroid development by regulating the switch from foetal to adult haemoglobin, and is also responsible for the initiation of megakaryoblastic proliferation [204]. The generation of megakaryocytes and erythroids are tightly associated during differentiation and it has been shown in the mouse model that these are derived from a bipotent megakaryocyte-erythroid progenitor (MEP) [205]. It has been reported that GATA-1 interacts with the Ets domain of *friend leukaemia virus integration 1* (*FLI1*) in the synergistic expression of megakaryocyte-specific genes [206]. Guo *et al.* (2009) [207] found that the v-myc myelocytomatosis viral oncogene homolog (c-Myc) protein controls the fate of MEP progenitors, resulting in a dramatic increase of megakaryocytes when over-expressed and an apoptotic loss of erythroblasts was also observed in *c-Myc* knockout mice. Studies of c-Myc have revealed critical roles in multiple biological processes including embryonic development, angiogenesis, tissue regeneration, and embryonic and adult haematopoiesis [207]. A large number of genes appear to be involved in the specification of primitive haematopoietic and in particular primitive erythropoiesis, but the stage at which all of this occurs during the haemangioblast development is still largely unknown, and many molecular effectors remain to be discovered in the creation of pathways, which may provide a clearer understanding of the underlying mechanisms.

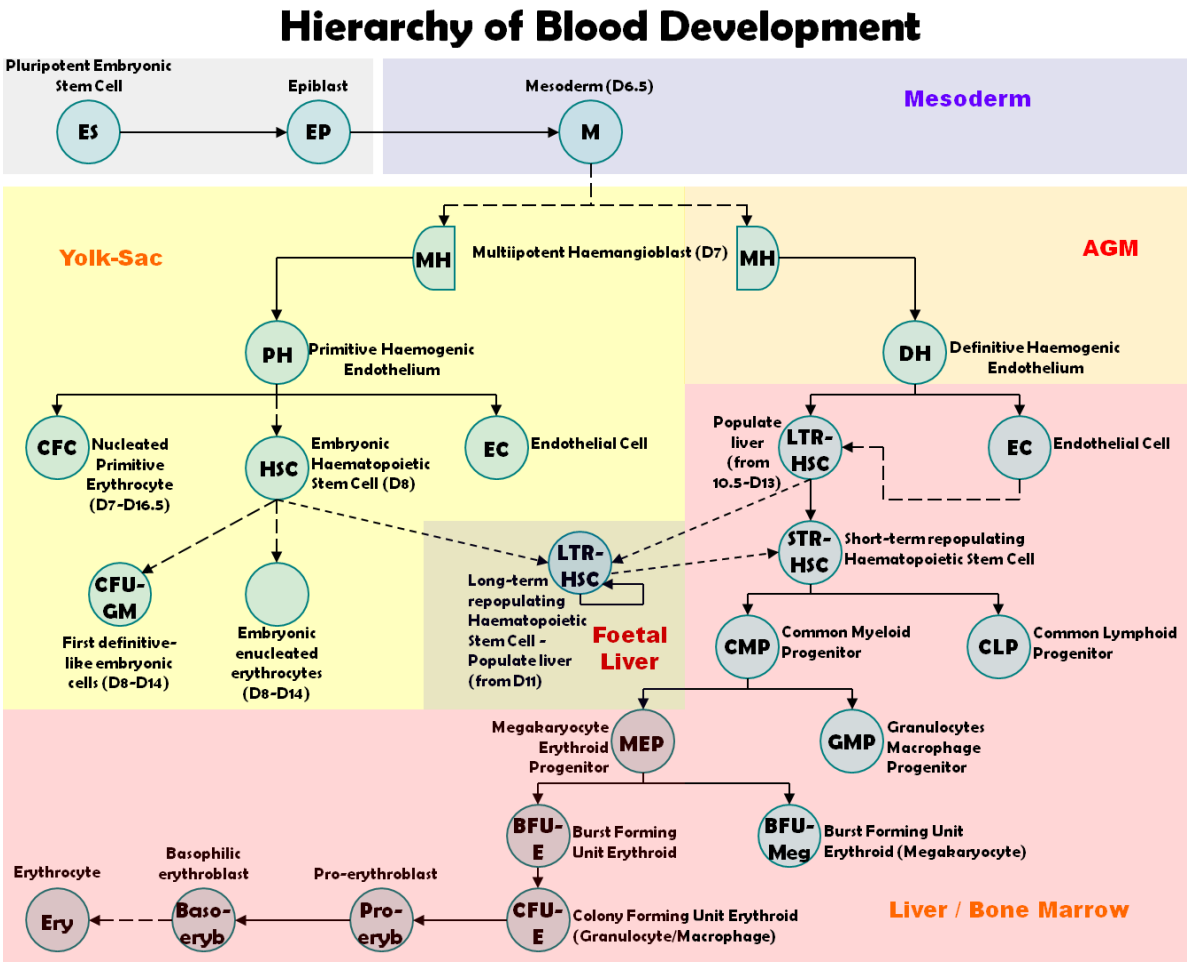


Figure 14: Hierarchy of blood development

This diagram shows a simplified developmental hierarchy of the haematopoietic system whereby development starts from pluripotent hESC stage and specialising through to more definitive cell types. (Adapted from Lai et al., (2006) [208])

1.2.8.2 Definitive Haematopoiesis

During definitive haematopoiesis, haematopoietic stems cells give rise to not only erythrocytes and megakaryocytes, but to the full range of haematopoietic cell types including lymphocytes and myelocytes [93]. The switch from primitive haematopoiesis to definitive haematopoiesis occurs in overlapping waves identified by the dynamic change of haemoglobin composition and erythrocyte morphology within the circulating blood. Lux *et al.* (2008) [209] reported that all primitive and definitive haematopoietic progenitor cells in circulation before day 10 d.p.c. in the mouse embryo are derived from the yolk sac. With reference to Figure 14, it has been previously discussed that although there appears to be overwhelming evidence that HSCs responsible for seeding the liver / bone marrow for definitive haematopoiesis

originate from the AGM region, studies have shown that the yolk sac does have the ability to produce definitive haematopoietic cells which may contribute to seeding the definitive niches.

The generation of haematopoietic cells from these definitive niches rely on the expression of a number of genes, of which *RUNX1* and *SCL* are the two key transcriptional regulators. Studies have implicated the importance of a transcription factor called runt-related transcription factor 1 (*RUNX1*) in haematopoietic development linking it with various types of leukaemia [172, 210, 211]. The runt-related transcription factor 1 (*RUNX1*) is also known as AML-1 and CBFA2, due to its discovery and link with translocations in human leukaemia [212]. *RUNX1* is a member of a family of proteins called core binding factors (CFB), and these transcription factors consist of heterodimeric subunits CBFα (*RUNX1* Runt domain) and CBFβ which are involved in DNA binding and non-DNA binding respectively [172, 213]. As mentioned previously, *RUNX1* appears to have an effect on haemangioblast development and commitment to the primitive and definitive haematopoietic lineages, and *in vitro RUNX1* knockout studies results in the generation of 20 times fewer blast colonies [172]. It was shown that this group of blast colonies was restricted to a primitive haematopoietic fate, suggesting that *RUNX1* does not play a vital role in primitive haematopoietic development, but it is indispensible for the generation of definitive haematopoiesis. Surprisingly, Notch has been described in a contrasting role to that within primitive haematopoiesis, with Burns et al. (2005) [214] reporting that *RUNX1* gene expression is induced by Notch, and this pathway is important in the specification of HSC fate in definitive haematopoiesis. In *Drosophila* studies, it was noticed by Mandal et al. (2004) [215] that *Notch* expression is required for the specification of dorsal mesoderm towards haematopoietic lineages, and in the absence of *Notch*, the mesoderm appears to be endothelial-like, whilst in the over-expression of *Notch*, the mesoderm fate becomes almost entirely haematopoietic.

1.3 Aims and Hypotheses

The principle aim of this project is the discovery of novel transcripts which have not to this date been discovered to play an essential part in haematopoietic development. This ambitious aim was made possible through the advent of microarray technology which allows the gene expression of the entire human genome at various stages of in vitro haematopoietic differentiation to be investigated. A crucial starting point is the optimisation of the differentiation protocol for the efficient generation of haematopoietic progenitor cells from human embryonic stem cells (HESCs). The microarrays within this project will consist of these haematopoietic progenitors isolated at various stages of haematopoietic differentiation for transcriptome comparisons and analyses. Due to the exploratory nature of this approach in the discovery of novel haematopoetic transcripts, specific transcript predictions may not be possible, but one would expect a number of well-studied haematopoietic transcripts such as SOX17, SOX7, HOXA9, and LMO2 to be present within the list of haematopoietic important transcripts. At least one novel transcript will be selected from this list and its role investigated through the over-expression of the transcript within H9 and H1 HESCs, with its effects on in vitro haematopoietic differentiation analysed and compared to the corresponding negative controls.

The principal aims of this project were as follows:

1. To establish optimal methods for the differentiation of human embryonic stem cells to haemangioblasts.

2. To produce a time course of microarray gene transcript changes from the RNA of haemangioblast-enriched sorted cells.

3. To analyse the results of microarray gene transcript changes and identify novel transcripts that may play an important role in haematopoietic differentiation.

4. To investigate the effects of factors from Aim 3. of the microarray analysis on the haematopoietic differentiation of human embryonic stem cells.

CHAPTER 2 - DIFFERENTIATION

2.1 Introduction

The incredible pluripotent nature of human embryonic stem cells (HESCs) enables them to differentiate into a multitude of cell types and in conjunction with their unique self renewal capacities, these immortal pluripotent cells appear to be the perfect model for the study of early human development. HESCs have the potential to differentiate into cell types representative of all three germ layers including those of the mesoderm, which in turn has the ability to gives rise to all cell types of the haematopoietic system [216, 217]. In this chapter, protocols of HESC sub-culture and differentiation will be explored in tandem with a number of essential methods required for the maintenance of HESC, including feeder layer derivation, sub-culture, and inactivation. Strategies for differentiation will also be investigated beginning with comparisons between foetal bovine serum-induced differentiation and cytokine-directed haematopoietic differentiation. In addition, a number of differentiation optimisations will be implemented in an attempt to reduce variables and to promote haematopoietic differentiation efficiency. Unfortunately, despite the wealth of research accumulated over the past century in haematopoietic development, directed control of HESC differentiation into haematopoietic stem cells (HSCs) to this day remains a challenge with many underlining processes and mechanisms governing early embryonic haematopoiesis still unclear. The resulting *in vitro* efficiencies of haematopoietic differentiation thus remains very low, however it is hoped that the data acquired from this project may contribute to furthering the understanding of embryonic haematopoietic development and as a result, enhance the efficiency of haematopoietic differentiation from HESCs.

2.2 Methods

2.2.1 Mouse Embryonic Fibroblast (MEF)

2.2.1.1 Introduction to Mouse Embryonic Fibroblast (MEF)

Undifferentiated HESCs are notoriously difficult to maintain in culture and require a feeder layer (mouse embryonic fibroblasts) to provide an *in vitro* stem cell niche. Mouse embryonic fibroblasts (MEFs) play a dual role in the culture of human embryonic stem cells (HESCs), providing both an extracellular matrix (ECM) on which the stem cells attach onto and essential growth factors which support growth and the maintenance of pluripotency [218]. Proteome analyses by Lim and Bodnar (2002) [219] of conditioned media from MEFs revealed the presence of numerous growth factors that include insulin-like growth factor binding protein 4 (IGFBP-4), transforming growth factor beta 1 (TGFβ1), fibroblast growth factor 2 (FGF2; albeit at very low levels), Activin A, and bone morphogenetic protein 4 (BMP4). Studies have shown that activin / nodal signalling through Smad2/3 and FGF signalling play a cooperative role in the maintenance of HESC pluripotency [220, 221]. It has been shown that the most critical growth factor for pluripotency is Activin A of which MEFs produce an abundance of, but also the presence of FGF2 is also required for long-term maintenance of pluripotency [221]. MEFs do not however produce enough of FGF2 and therefore an additional FGF2 supplement must be added to the HESC media. Human feeders conversely produce substantially more FGF2 but secrete lower levels of Activin A, and through the analysis for pluripotency markers, it was discovered that the efficiency of the maintenance of undifferentiated HESCs using human feeders were not as great as MEFS [221].

2.2.1.2 Mouse Embryonic Fibroblast (MEF) Isolation

Swiss MF1 pregnant mice containing 12.5 – 13.5 day old embryos were humanely sacrificed by experienced prior to collection, by qualified staff from our certified animal house facility. The mice were dissected and the uterine horns were removed and placed in a Petri dish containing Phosphate Buffered Saline (PBS) supplemented with 10% Foetal Bovine Serum (FBS). The limbs, heads, tails and all visible organs removed from the Swiss MF1 mouse embryos and washed in fresh PBS supplemented with 10% FBS. Fine scissors were used to mince the remaining torsos before they were further dissociated by repeated pipetting and then were incubated in 5 ml of 0.25% Trypsin (Invitrogen) containing 0.53 mM EDTA for 5 minutes at 37°C. Inactivation of trypsin was achieved by the addition of twice the volume of

FBS after dissociation. The cells were centrifuged at 1000 rpm for 3 minutes, the supernatant was removed and the cells were transferred into MEF media (Table 4), and then were distributed into T75 (Iwaki) flasks (distribution density of about one embryo per flask). Flasks were incubated at 37°C with 5% CO2 and monitored on a daily basis.

MOUSE EMBRYONIC FIBROBLAST (MEF) MEDIA

Components	Quantity	Supplier
Dulbecco's Modified Eagle's Medium Without Sodium Pyruvate	450 ml – 90%	Invitrogen
Heat Treated Foetal Bovine Serum (FBS)	50 ml – 10%	Sigma
Non-Essential Amino Acids (MEM)	5 ml – 1%	Invitrogen
L-Glutamine/Penicillin-Streptomycin Solution (1 mM)	5 ml – 1%	Invitrogen

Table 4: Mouse Embryonic Fibroblast (MEF) Media Composition

2.2.1.3 Mouse Embryonic Fibroblast (MEF) Culture

Each MEF T75 flask was inspected daily and checked for cell confluency and morphology. The MEF media was changed every 2 days and when the MEF cells within the flask reached approximately 90% confluent, the flask was ready for passaging in a ratio of 1 to 3. The passaging process was required to provide the MEF cells space within the flask to continually expand. The MEF cells were washed with Phosphate Buffered Saline (PBS; 1x; Invitrogen) and incubated with 5 ml of 0.05% trypsin EDTA 4Na (Invitrogen) for 5 minutes at 37°C. The trypsin was inactivated by the addition of 10 ml of MEF media (Table 4) and cell suspensions were collected by centrifugation at 800 x g for 5 minutes. The supernatant was removed and cells were resuspended in 36 ml of MEF media and transferred equally into three new T75 flasks. Flasks were incubated at 37°C with 5% CO2 and checked on a daily basis.

2.2.1.4 Mouse Embryonic Fibroblast (MEF) Inactivation by Mitomycin C

Prior to plating and use with human embryonic stem cells (HESC), mouse embryonic fibroblasts (MEF) were treated with Mitomycin C (Sigma) which is a potent DNA crosslinker that causes DNA replication arrest in cells and thus prevents cell division [222]. Plating HESC onto inactivated MEF allows for plate to plate consistency in MEF quantity which aids in reducing overall HESC variation between plates, and as inactivated MEF are non-dividing they will also utilise less nutrients from the media but still produce the essential factors for maintenance of HESC pluripotency.

Mitomycin C was added to the culture media within the flask of MEF at a final concentration of 10 µg / ml and incubated for 2 hours at 37°C with 5% CO_2. During Mitomycin C incubation, the 6-well and 4-well plates were gelatinised by the addition of 1.5 ml of filter-sterilised 0.1% Gelatin solution (Table 5) into each well of a 6-well plate and 0.5 ml into each well of a 4-well plate. These gelatinised plates were incubated at 37°C for 2 hours, and the Gelatin solution was only removed immediately prior to plating of inactivated MEF. After Mitomycin C incubation, the media was removed and placed into a Mitomycin C waste container for specialised disposal afterwards. The cells were carefully washed three times with excess PBS and the discarded wash was also placed into a Mitomycin C waste container. The cells were detached from the tissue culture flasks by incubation with 5 ml of 0.05% trypsin with EDTA (Invitrogen) at 37°C for 5 minutes. The trypsin was inactivated by the addition of 10 ml of MEF media, and the cell suspension was collected by centrifugation at 800 x g for 5 minutes. The supernatant was discarded and the cell pellet was resuspended in 10 ml of fresh MEF media. 20 µl of MEF cells suspension were combined with 20 µl of 0.4% Trypan Blue solution (Sigma) which stains dead cells and helps to provide a measure of the number of viable MEF cells within the sample. 20 µl of the Trypan blue stained cells suspension were placed onto a haemocytometer (Fisher Scientific) and numbers of viable cells were recorded under a Nikon TS100 inverted microscope at 10 x magnification. The MEF cells were plated at a concentration of 1.6 x 10^4 cells / cm^2 on pre-gelatinised 4-well or 6-well plates (Iwaki), and at a concentration of 4.0 x 10^4 cells / cm^2 on pre-gelatinised flasks for conditioned media.

0.1% GELATIN SOLUTION

Components	Quantity	Supplier
Gelatin from Porcine skin	0.5 g	Sigma
Distilled Water	500 ml	-

Table 5: 0.1% Gelatin Composition

2.2.1.5 Mouse Embryonic Fibroblast (MEF) Storage – Cryopreservation

Confluent T75 flask of MEF cells was washed with Phosphate Buffered Saline (PBS) and incubated with 5 ml of 0.05% trypsin with EDTA for 5 minutes at 37°C. The cell suspension was collected by centrifugation at 800 x g for 5 minutes, and the supernatant was discarded and replaced with 200 µl of MEF freezing down media (Table 6) per original T75 flask of MEF cells. 200 µl of cell suspension were transferred to each cryopreservation cryovials (Nunc) and placed into an isopropanol (VWR) freezing container which was then stored in a -80°C freezer.

MOUSE EMBRYONIC FIBROBLAST (MEF) FREEZING MEDIA

Components	Quantity	Supplier
Dulbecco's Modified Eagle's Medium Without Sodium Pyruvate	6 ml	Invitrogen
Dimethyl sulfoxide (DMSO)	2 ml	Sigma
Heat Inactivated Foetal Bovine Serum (FBS)	2 ml	Sigma

Table 6: Mouse Embryonic Fibroblast (MEF) Freezing Media Composition

2.2.2 Human Embryonic Stem Cell (HESC)

2.2.2.1 Human Embryonic Stem Cell (HESC) Characterisation

Human embryonic stems lines 'H9' and 'H1' used within this project were obtained from WiCell [223] which characterised these HESCs over a period of 20 passages to ensure that these are healthy and display the classic stem cell phenotype. Characterisation testing included a number of assays for genetic stability (Karyotype by G-band, comparative genome hybridisation array), expression of hESC markers (analysis of SSEA3, SSEA4, TRA-1-60, TRA1-81, and OCT-4 pluripotency marker by flow cytometry), gene expression profile (microarray analysis), differentiation capacity, and morphology analysis [224].

2.2.2.2 Human Embryonic Stem Cell (HESC) Culture

Human embryonic stem cells (HESC) were grown in colonies within 4-well plates (Nunc) or 6- well plates (Iwaki) supported by MEF cells in HESC media (Table 7). A plate which contained HESC colonies that cover up to 90% of each well was ready to be passaged in a ratio of 1 to 3. Prior to passaging of the HESC plate, differentiated areas were carefully scraped away under an IVF hooded dissection microscope using a 29 gauge (29G) needle. The differentiated areas of cells within each colony appeared much thicker and whiter when compared with non-differentiated HESC areas, and under an inverted light microscope (Zeiss Aviovert 200M with x10 magnification objective lens) these cells appear larger, with smaller nuclei, and more variable in shape. After cleaning, the media was aspirated and replaced with 1 ml of Collagenase type IV (1 mg / ml; Sigma; Table 8) per well of a 6-well plate and incubated for 15 minutes at 37°C. The colonies were inspected and at this stage the edge of the colonies should be clearly defined and appear slightly raised. However if these were not seen then the colonies were incubated in Collagenase type IV for an additional 5 minutes, with the colonies re-checked for raised edges up to a maximum total incubation of 30 minutes. After incubation the Collagenase type IV solution was carefully removed and replaced with 1 ml of HESC media per well of a 6-well plate, and incubated for another 15 minutes at 37°C. The majority of colonies at this point should be on the verge of detaching from the plate surface, and the gentle pipetting motion of a 1 ml pipette was used to detach the colonies completely and break them up into smaller pieces. These fragmented pieces of HESC were transferred to a freshly prepared inactivated MEF plate containing HESC media and incubated at 37°C in a humidified atmosphere with 5% CO_2. The HESC colonies were monitored daily with any

differentiated area of cells removed and media changed every 2 days. Typically 3 - 6 days (dependent on cell line) post-passage, the colonies would reach 90% confluent within the wells and then passaged into fresh MEF plates.

HUMAN EMBRYONIC STEM CELL (HESC) MEDIA

Components	Quantity	Supplier
Knockout-Dulbecco's Modified Eagle's Medium (KO-DMEM)	120ml – 80%	Invitrogen
Knockout-Serum Replacement (KO-SR)	30ml – 20%	Invitrogen
Non-Essential Amino Acids (MEM)	1ml – 1%	Invitrogen
L-Glutamine/Penicillin-Streptomycin Solution	1ml – 1%	Invitrogen
Basic Fibroblast Growth Factor (BFGF)	8ng/ml	Invitrogen

Table 7: Human Embryonic Stem Cell Media Composition

COLLAGENASE (1 X)

Components	Quantity	Supplier
Collagenase Type IV from Clostridium histolyticum	50 mg	Invitrogen
Knockout-Dulbecco's Modified Eagle's Medium (KO-DMEM)	50 ml	Invitrogen

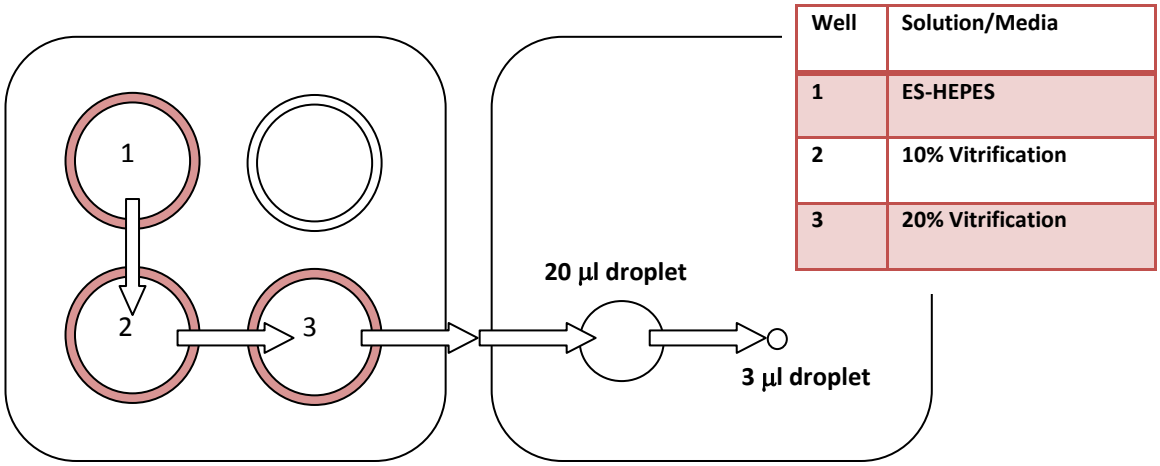
Table 8: Collagenase Composition

2.2.2.3 Human Embryonic Stem Cell Storage - Open Straw Vitrification Freezing

Open straw vitrification is a very simple and rapid method for freezing and storing human embryonic stem cells in liquid nitrogen. This method allows high cooling and warming speeds (over 20,000°C per minute) which aids to circumvent chilling injury and decrease toxic and osmotic damage to the cells [225].

0.5 ml of ES-HEPES solution (Table 9) was pipetted into well number 1 (Figure 15) of a 4-well plate, 0.5 ml of 10% vitrification solution (Table 10) was pipetted into well number 2 (Figure 15), and 0.5 ml of 20% vitrification solution (Table 12) was pipetted into well number 3 (Figure 15). A 20 µl drop of 20% vitrification solution was placed on the lid of the plate (Figure 15) and both lid and plate were warmed to 37°C. HESC colony fragments that have been passaged using Human Embryonic Stem Cell Culture protocol were carefully placed into well number 1 of the vitrification plate and incubated for 1 minute at 37°C. 10 colony fragments were transferred from well number 1 to well number 2 of the vitrification plate and incubated

for 1 minute at 37°C. The colony pieces were transferred from well number 2 to well number 3 of the vitrification plate and incubated for 25 seconds at 37°C. The colony fragments were transferred in the smallest possible volume to the 20 µl drop of 20% vitrification solution placed on the lid. The colony fragments were aspirated in a 3 µl volume from the 20 µl drop and deposited as a separate small, peaked droplet on the lid (Figure 15). A vitrification straw was immediately inserted (narrow end) into the 3 µl drop at a 30° angle to the plane of the dish, and the fragments were drawn into the straw by capillary action. The straw was plunged into liquid nitrogen and stored in a labelled tube within a nitrogen storage canister.



Well	Solution/Media
1	ES-HEPES
2	10% Vitrification
3	20% Vitrification

Figure 15: Open Straw Vitrification Plate – Freezing

ES-HEPES SOLUTION

Components	Quantity	Supplier
Knock Out-Dulbecco's Modified Eagle's Medium (KO-DMEM)	15.6ml	Invitrogen
Knockout-Serum Replacement (KO-SR)	4ml	Invitrogen
HEPES (1M)	0.4ml	Invitrogen

Table 9: ES-HEPES Solution Composition

10% VITRIFICATION SOLUTION

Components	Quantity	Supplier
ES-HEPES Solution	2ml	-
Ethylene Glycol	0.25ml	Sigma
DMSO	0.25ml	Sigma

Table 10: 10% Vitrification Solution Composition

1M SUCROSE STOCK

Components	Quantity	Supplier
Sucrose	3.42g	VWR
ES-HEPES Solution	14ml	-
Knockout-Serum Replacement (KO-SR)	2ml	Invitrogen

Table 11: 1M Sucrose Stock Composition

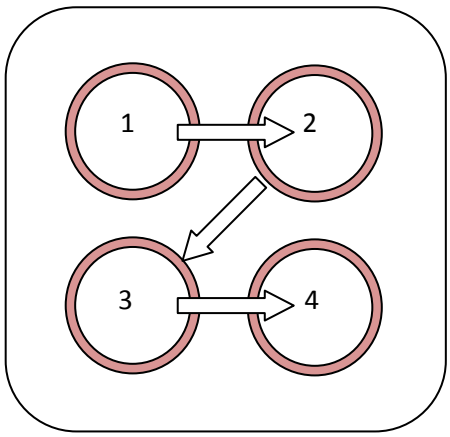
20% VITRIFICATION SOLUTION

Components	Quantity	Supplier
ES-HEPES Solution	0.75ml	-
1M Sucrose Stock	0.75ml	-
Ethylene Glycol	0.5ml	Sigma
DMSO	0.5ml	Sigma

Table 12: 20% Vitrification Solution Composition

2.2.2.4 Human Embryonic Stem Cell Storage - Open Straw Vitrification Thawing

0.5 ml of 0.2M Sucrose solution (Table 14) was pipetted into well number 1 (Figure 16) of a 4-well plate, 0.5 ml of 0.1M Sucrose solution (Table 13) was pipetted into well number 2, and 0.5 ml of ES-HEPES solution (Table 9) was pipetted into wells number 3 and 4. A vitrification straw which contained cells in vitrification solution was removed from nitrogen storage and the narrow end was submerged into 0.2M Sucrose in well number 1 (Figure 16) at 30° angle to the plane of the plate. As soon as the liquid within the straw melts, a 10 µl pipette was placed at the top end of the straw and the liquid was gently pipetted out. The colony fragments were incubated in 0.2M sucrose solution in well number 1 (Figure 16) for 1 minute at 37°C and then they were transferred to 0.1M sucrose solution in well number 2 and incubated for 5 minutes at 37°C. The colony fragments were transferred to ES-HEPES solution in well number 3 and incubated for 5 minutes at 37°C, and then the fragments were moved to ES-HEPES solution in well number 4 and incubated for a final 5 minutes at 37°C. The fragments were collected and placed into a single well of a 4-well plate which contained inactivated mouse embryonic fibroblasts (MEF) in human embryonic stem cell (HESC) media.



Well	Solution/Media
1	0.2M Sucrose
2	0.1M Sucrose
3	ES-HEPES
4	ES-HEPES

Figure 16: Open Straw Vitrification Plate - Thawing

0.1M SUCROSE SOLUTION

Components	Quantity	Supplier
ES-HEPES Solution	4.5ml	-
1M Sucrose Stock	0.5ml	-

Table 13: 0.1M Sucrose Solution Composition

0.2M SUCROSE SOLUTION

Components	Quantity	Supplier
ES-HEPES Solution	4ml	-
1M Sucrose Stock	1ml	-

Table 14: 0.2M Sucrose Solution Composition

2.2.2.5 Human Embryonic Stem Cell Storage - Cyrovial Cryopreservation Freezing

The advantage of using cryovial cryopreservation in the storage of HESC is that an entire confluent 6-well plate can be stored into just one cryovial, whereas with the open straw vitrification method only a maximum of 10 colony fragments may be stored per straw. Cryovial cryopreservation slow freezes the HESC (-1°C per minute) and therefore this method will result in lower cell viability compared with open straw vitrification, but as the starting quantity (when freezing) is so much larger, the resultant loss of cells (when thawing) is acceptable and colonies appeared within one week after re-plating into a 6-well plate. A study by Watanabe *et al.* (2007) [226] however has shown that the inclusion of Rho-associated Kinase (ROCK) inhibitor

Y-27632 to the freezing media significantly enhances survival by as much as 27 fold protecting the cells from apoptosis. This significant enhancement of survival in combination with advantages that include ease of transport and storage means that this method of freezing within cryovials is now the preferred HESC freezing method in our group.

HESC colony fragments that have been passaged using Human Embryonic Stem Cell Culture protocol were transferred into a falcon tube and the cells were pelleted by centrifugation at 200 x g for 5 minutes. The supernatant was aspirated and replaced with 200 µl of HESC media per 6-well plate of HESC colonies and combined with 200 µl of cryopreservation media. 200 µl of HESC cryopreservation mixture were transferred into each cryovial tube and the cryovials were placed into an isopropanol freezing container and stored in -80°C for 24 hours. The cryovials were later transferred to liquid nitrogen for long-term storage.

CRYOPRESERVATION MEDIA (2x)

Components	Quantity	Supplier
Heat Treated Foetal Bovine Serum (FBS)	6ml	Sigma
DMSO	2ml	Sigma
HESC Media	2ml	-
Rho-associated kinase (ROCK) inhibitor, Y-27632	10µM	Sigma

Table 15: Cryopreservation Media Composition

2.2.2.6 Human Embryonic Stem Cell Storage - Cryovial Cryopreservation Thawing

When required, frozen HESC cryovials were removed from liquid nitrogen storage and thawed quickly within a water bath (Fisher Scientific) set at 37°C for 30 seconds. The HESC cryopreservation mixture was transferred to 10 ml of HESC media (Table 7) and was centrifuged at 200 x g for 5 minutes. The supernatant was removed and the pellet was gently resuspended in 6 ml of HESC media. The HESC suspension was divided equally between 3 wells of a freshly prepared 6-well MEF plate. The thawed HESC plate was incubated at 37°C in a humidified atmosphere with 5% CO_2, and the plate was monitored daily and media changed every two days. Typically 7-8 days post-thawing, the colonies were of suitable size to be passaged onto fresh MEF plates.

2.2.3 Differentiation of Human Embryonic Stem Cell (HESC)

2.2.3.1 Introduction to Differentiation

Human embryonic stem cells (HESCs) will readily differentiate if the signalling pathways responsible for the maintenance of pluripotency are disrupted, and the sole withdrawal of FGF2 from the media is enough to cause spontaneous differentiation [227]. Although this simple adjustment of the media will result in the obliteration of the self-renewal process and HESCs lose their pluripotency, in practice however, the presence of foetal bovine serum (FBS) is utilised in the media in combination with FGF2 withdraw, enhances differentiation [3].

Colonies of HESCs are typically cultured as monolayer on a feeder layer (or feeder-free with conditioned media), however during differentiation these colonies must be separated from the pluripotency enhancing feeders and re-plated as monolayer colonies onto feeder-free substrate (for example, BD Matrigel), differentiation enhancing cell lines as co-cultures (for examples stromal cells), or as embryoid bodies (EBs). Differentiation studies have shown that the formation of a three-dimensional cell aggregate from HESC called an embryoid body (EB) can accurately recapitulate many aspects of *in vivo* cell differentiation during early mammalian embryogenesis [228]. The EBs in combination with standard differentiation media (Table 16) will undergo differentiation, and it is the foetal bovine serum (FBS) which is the crucial component that encourages the EBs to give rise to cells representative of those of the three germ cell layers (mesoderm, endoderm, and ectoderm) [228]. The same standard differentiation media may however be used for monolayer and particularly stromal co-culture differentiation of HESC, of which studies have shown that the haematopoietic niche provided by the mouse stromal cells significantly promotes differentiation of the HESCs towards a haematopoietic lineage [229]. However, the inherent batch-to-batch variability of stromal cells may significantly affect results, and more importantly, if the ultimate goal is for the production of haematopoietic cells derived from HESCs for human clinical uses, then the differentiation process must be completely animal-free. EB differentiation therefore is preferred and although the FBS used is still obtained from bovine and therefore suffers from batch-to-batch variability, companies have since attempted to recreate the basic components of FBS in producing xeno-free and serum-free defined media with no batch-to-batch variation [230].

2.2.3.2 Embryoid Body (EB) Differentiation - Standard Differentiation

For differentiation, one well of Embryoid Bodies (EBs) consisted of HESC colonies from three confluent wells of a six-well plate. Prior to EB formation, differentiated cells (identified by thick heterogeneous population of irregular cells) present within the HESC plates were removed using a fine 29 gauge (29G) needle. The media was aspirated and replaced with 500 µl of Collagenase type IV (1 mg / ml; Sigma), and the HESC plate was incubated for 10 minutes at 37°C. Following incubation, the Collagenase solution was carefully removed and replaced with HESC media, and incubated for another 10 minutes at 37°C. The gentle pipetting motion of a 1 ml pipette was used to detach the colonies and break them up into smaller pieces, and these were transferred to an ultra low attachment culture plate (Corning) containing standard EB differentiation media (Table 16). The plate was incubated at 37°C with 5% CO_2. Every 3 days during differentiation, the EBs were transferred to a 50 ml falcon tube and allowed to settle to the bottom of the tube for 30 minutes. 80% of the media was removed and replaced with fresh EB media into a new ultra low attachment culture plate (Corning).

STANDARD DIFFERENTATION MEDIA

Components	Quantity	Supplier
Knock Out-Dulbecco's Modified Eagle's Medium (KO-DMEM)	120ml – 80%	Invitrogen
Hyclone Heat Treated Foetal Bovine Serum (FBS)	30ml – 20%	Thermoscientific
Non-Essential Amino Acids (MEM) (0.1mM)	1.5ml – 1%	Invitrogen
L-Glutamine/Penicillin-Streptomycin Solution (1mM)	1.5ml – 1%	Sigma

Table 16: Standard Differentiation Media Composition

2.2.3.3 Embryoid Body (EB) Differentiation - Optimised Differentiation

Embryoid bodies (EB) produced from HESC will faithfully differentiate into all three germ layers when cultured in media containing FBS, however as previously discussed, the batch-to-batch variability of FBS and that it is an animal-derived product means that results generated may not be consistent and / or translatable for clinical uses. The use of xeno-free and serum-free media therefore appears to be a more appropriate starting point, and holds a number of advantages over the reliance of FBS for differentiation. Although FBS-based media does induce HESCs to differentiate, the differentiation is however undirected and uncontrolled, resulting the in the heterogeneous production of cells representative of all three germ layers. Studies have shown that the use of synthetically defined serum-free media in combination with

cytokines may persuade the EBs to differentiate down towards a specific germ layer rather than another [157]. The optimised differentiation method described within this section was conceived by Kennedy *et al.* (2007) [157] and was adapted in this experiment to encourage the HESCs to differentiate towards the mesodermal lineage, and thus in theory, results in a greater proportion of haematopoietic progenitors compared with the standard differentiation method with FBS.

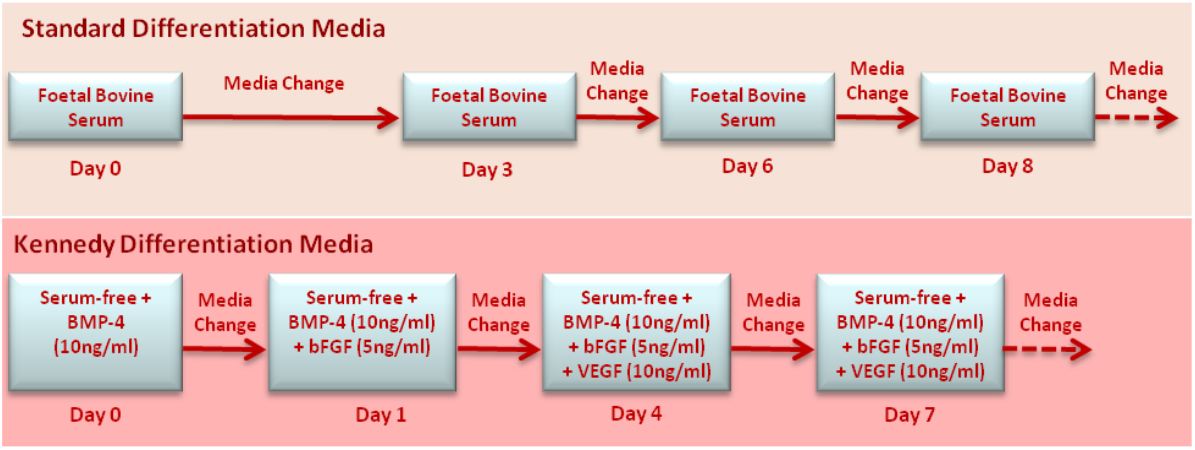


Figure 17: Flow diagram Comparison between Standard and Optimised media

The key component of the standard differentiation media is foetal bovine serum (FBS) which initiates spontaneous differentiation of HESCs down all three germ layers. The serum-free optimised media however utilises BMP4, bFGF, and VEGF cytokines in attempt of inducing differentiation towards mesoderm and haematopoietic lineages.

The process of preparing the EBs from HESC colonies were performed using the same protocol as the standard differentiation. However instead of standard differentiation media (Table 16), optimised differentiation media 1 (Table 17) was used to resuspend and maintain the EBs over the differentiation period. The primary difference between the optimised differentiation media and the standard differentiation media was that the former contain serum-free media (StemPro-34) and the latter contain FBS. StemPro-34 media (Invitrogen) was developed to contain no animal derived components and consist of materials of human and synthetic origin only which promotes reproducibility due to the elimination of batch to batch variation, and being xeno-free means that this differentiation may be applicable for future therapeutic use. StemPro-34 media was specially formulated to support the development of human haematopoietic stem cells and studies have suggested that when compared alongside FBS-based media, it produces over 3 fold more haematopoietic cells over 10 days differentiation period [231]. The differentiation plate was incubated at 37°C with 5% CO_2 for 24 hours, and then the EBs were transferred to a 50 ml falcon tube and were allowed to settle to the bottom of the tube for 15 minutes. The upper 3/4 of the media within the tube which contain single cells was aspirated (taking great care to not aspirate the EBs settled in the

bottom of the tube) and replaced with optimised differentiation media 2 (Table 18), and the plate now containing chiefly whole EBs were incubated at 37°C with 5% CO_2 for 72 hours. Following this incubation, the EBs were transferred to a 50ml falcon tube and were again allowed to settle to the bottom of the tube for 15 minutes. Once again the upper 3/4 of the media within the tube was removed and replaced with optimised differentiation media 3 (Table 19), and the plate was incubated at 37°C with 5% CO_2. The media was changed every 72 hours with optimised differentiation media 3 using the same method until the end of the differentiation period.

OPTIMISED DIFFERENTATION MEDIA 1

Components	Quantity	Supplier
StemPro-34 Serum Free Media	120ml – 80%	Invitrogen
Penicillin-Streptomycin Solution (1mM)	1.5ml – 1%	Sigma
Bone Morphogenic Protein 4 (BMP-4)	10ng/ml	Peprotech
Glutamax (2mM)	1.5ml – 1%	Invitrogen
Monothioglycerol (MTG)	4 x 10-4M	Sigma
Ascorbic acid	50µg/ml	Sigma

Table 17: Optimised Differentiation Media 1 Composition

OPTIMISED DIFFERENTATION MEDIA 2

Components	Quantity	Supplier
StemPro-34 Serum Free Media	120ml – 80%	Invitrogen
Penicillin-Streptomycin Solution (1mM)	1.5ml – 1%	Sigma
Bone Morphogenic Protein 4 (BMP-4)	10ng/ml	Peprotech
Glutamax (2mM)	1.5ml – 1%	Invitrogen
Monothioglycerol (MTG)	4 x 10-4M	Sigma
Ascorbic acid	50µg/ml	Sigma
Basic Fibroblast Growth Factor (bFGF)	5ng/ml	Invitrogen

Table 18: Optimised Differentiation Media 2 Composition

OPTIMISED DIFFERENTATION MEDIA 3

Components	Quantity	Supplier
StemPro-34 Serum Free Media	120ml – 80%	Invitrogen
Penicillin-Streptomycin Solution (1mM)	1.5ml – 1%	Sigma
Bone Morphogenic Protein 4 (BMP-4)	10ng/ml	Peprotech
Glutamax (2mM)	1.5ml – 1%	Invitrogen
Monothioglycerol (MTG)	4 x 10-4M	Sigma
Ascorbic acid	50µg/ml	Sigma
Basic Fibroblast Growth Factor (bFGF)	5ng/ml	Invitrogen
Vascular Endothelial Growth Factor (VEGF) 165	10ng/ml	Peprotech

Table 19: Optimised Differentiation Media 3 Composition

2.2.4 Identification of Haematopoietic Progenitors

2.2.4.1 Flow Cytometry Analysis

Flow cytometry is a powerful tool which allows the identification and interrogation of individual cells based on size (Side Scatter - SSC), granularity (Forward Scatter - FSC) and cell surface antigen markers using antibody conjugated fluorochromes.

Since this method relies on the identification and interrogation of individual cells, a single cell suspension was required. Haematopoietic differentiated embryoid bodies (EB) (created using the Embryonic Body (EB) Differentiation Protocol as previously described) were collected in a 50 ml falcon tube and centrifuged at 600 x g for 3 minutes. The supernatant was removed and Accutase (Chemicon; 1X) was added (approximately 1 ml of Accutase / 1 well of EB of a 6-well plate and incubated at 37°C for 30 minutes. To facilitate the dissociation during the incubation, the EBs were gently pipetted for 1 minute at 10 minutes intervals for a maximum of 30 minutes in total. The Accutase was inactivated by the addition of 1 ml of standard differentiation media (Table 16) / ml of Accutase, and centrifuged at 600 x g for 3 minutes. The supernatant was removed and replaced with 30 ml of PBS supplemented with 10% FBS. The number of cells within the cell suspension was counted and recorded using a Vi-Cell XR cell viability analyzer (Beckman Coulter) so that the correct volume of antibody conjugated fluorochromes (Table 20) can be added. The required antibody conjugated fluorochromes were added to the cell suspension at the manufacturers' specified concentrations, and the

mixture was incubated in the dark at 4°C for 1 hour to allow the antibodies sufficient time to match up and bind to the corresponding antigens on the cell surfaces. The incubated cells were washed with PBS to remove excess antibodies and resuspended in 500 µl of PBS with 5% FBS, and 10 µl of 4',6-diamidino-2-phenylindole (DAPI) (Sigma) were added to the sample and incubated in the dark at room temperature for 30 minutes. DAPI is a fluorescent stain that binds strongly to DNA and is widely used as a viability marker to identify live cells (unstained) from dead cells (stained) [232]. The cells were washed with PBS to remove excess stain and resuspended in 500 µl of PBS with 5% FBS and ready to be put through a flow cytometer for analysis (BD FACS LSRII) or for sorting (BD FACS Aria).

FLUOROCHROME CONJUGATED ANTIBODIES

Antibody	Conjugated Fluorochrome	Volume per 100,000 cells (µl)	Emission (nm)	Excitation (nm)	Supplier
CD34	Phycoerythrin-Cy7 (PE-Cy7)	5.0	800	488	BD Pharmingen
CD31	Allophycocyanin (APC)	0.5	680	595-605	eBioscience
KDR	Phycoerythrin (PE)	10.0	421	402	R&D Systems
CD144	Fluorescein (FITC)	10.0	530	488	Serotec
CD117	Phycoerythrin (PE)	5.0	570	488	BD Biosciences
CD45	Allophycocyanin –CY7 (APC-CY7)	5.0	660	595-605	BD Pharmingen

Table 20: Fluorochrome conjugated antibodies and their respective emission and excitation wavelengths

2.2.5 Haematopoietic Colony Forming Capacity

2.2.5.1 Colony Forming Assay - Colony Forming Units (CFU)

Colony assay is an important cytological technique for measuring the functional capacity of stem cells. There are many types of colony assays available depending on the type of stem cell in the study, and thus within this project the methylcellulose colony forming cell (CFC) assay was used for the study of haematopoietic stem cells. In this assay, the ability of the differentiated haematopoietic stem cell progenitor to proliferate and form colonies in response to cytokine stimulation within a semi-solid media was assessed. Every haematopoietic colony obtained through this assay was derived from a single progenitor cell, and therefore the results of this assay provides an excellent assessment of the original cell population including the type of progenitor present and the frequency, and thus the ability of this population to give rise to haematopoietic cells.

The semi-solid methylcellulose based media, MethoCult GF H4434 (Stemcell Technology) was thawed slowly from -20°C within a fridge at 4°C overnight and not at 37°C as rapid thawing will result in small 'lumps' which will not dissolve at 37°C and affect the efficiency of the assay [233]. Colony assays were set up with 3.0 x 10^4 cells in 1.5 ml of MethoCult GF H4434 which was composed of Methylcellulose in Iscove's MDM, Foetal Bovine Serum, Bovine Serum Albumin, 2-Mercaptoethanol, together with a cocktail of cytokines which include recombinant human stem cell factor (rh SCF), recombinant human granulocyte-macrophage colony-stimulating factor (rh GM-CSF), recombinant human interleukin 3 (rh IL-3), and recombinant human erythropoietin (rh EPO). The inclusion of the cytokine rh SCF promotes the growth of mast cells, myeloid, and lymphoid progenitors, rh GM-CSF promotes the growth of granulocytic and monocytic progenitors, rh IL-3 enhances early myeloid progenitors of all lineages, and rh EPO promotes the formation of erthyroid progenitors. This mixture was equally distributed between two 35 mm suspension culture dishes (Stem Cell technologies), and these were placed into a large 120 mm Petri dish together with a 35 mm dish that contained sterile water to prevent the MethoCult media from drying out (Figure 18). These dishes were incubated at 37°C in a humidified atmosphere with 5% CO_2 for 14 days, and then evaluated and observed under an inverted light microscope (Zeiss 200M). Using a gridded scoring dish, colony types and numbers were carefully enumerated and recorded together with their respective morphologies (Figure 19).

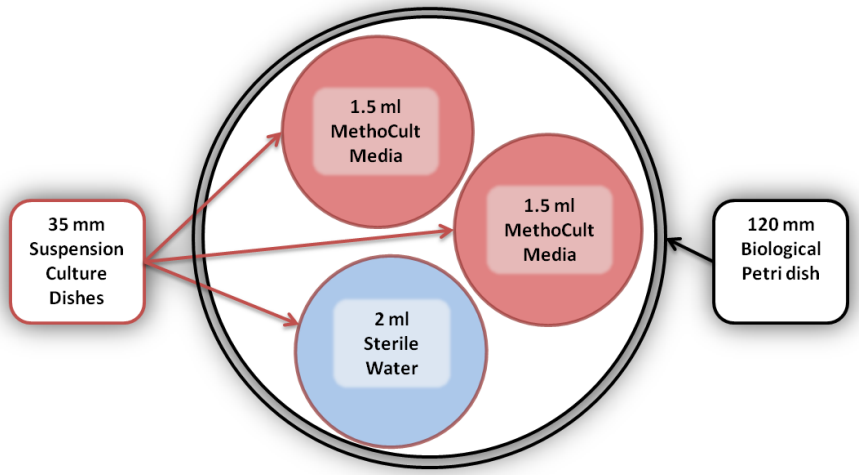


Figure 18: Preparation of Haematopoietic Colony Assay Plates

Haematopoietic colonies were set up in duplicate with 30,000 cells seeded into 1.5 ml of MethoCult media in each 35mm culture dish. These dishes were placed into a larger 120 mm biological Petri dish and an additional 35 mm dish that contained sterile water was included to maintain a humidified environment for the 14 days assay.

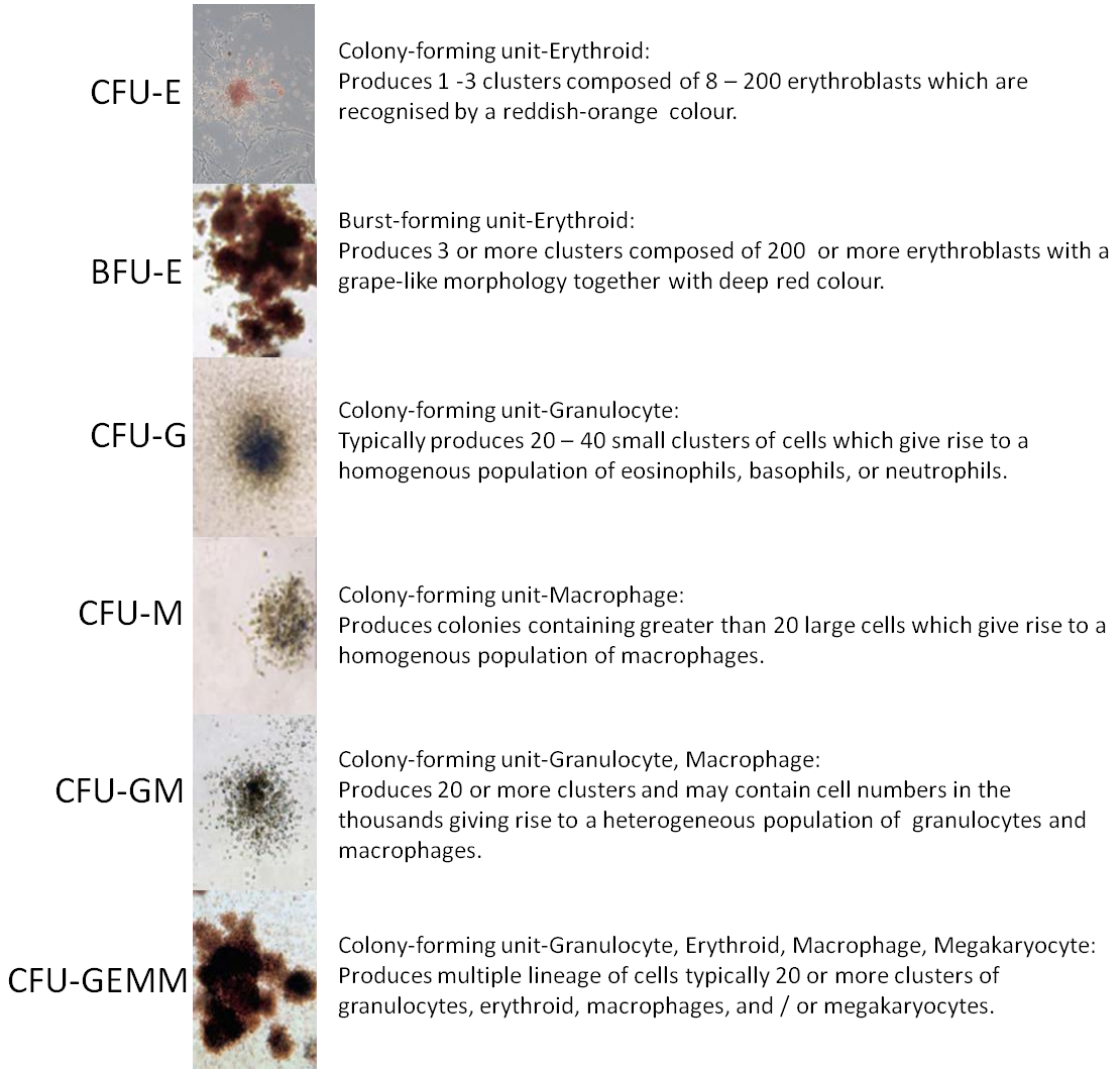


Figure 19: Morphological Identification of Colony Forming Units

The colonies produced from the haematopoietic colony assay may be identified and classified by their morphology into one of the colony forming unit types: CFU-E, BFU-E, CFU-G, CFU-M, CFU-GM, and CFU-GEMM. Adapted from MethoCult User Instructions (Stemcell Technologies).

2.2.5.2 Colony Forming Assay – Cytospin

For histochemistry analysis of the cells produced from the colony assay, the cells were required to be evenly distributed and fixed onto slides using the following cytospin method before the cells can be stained and observed under a microscope. Cells isolated from the colony forming cell (CFC) assay were placed into PBS supplemented with 10% FBS, washed three times and counted using a Casy cell counter TT (Sedna Scientific). The number of cells were recorded and the concentration was diluted to 5.0 x 10^4 cells / ml with PBS supplemented with 10% FBS, and if the recorded concentration was lower than 5.0 x 10^4 cells / ml the cell suspension was centrifuged at 800 x g for 3 minutes and resuspended into the appropriate volume for the required concentration. A clean slide and cytology funnel was placed into an available slot in a cytospin centrifuge (Thermo Shandon) with the cardboard filter facing the centre of the cytospin. 100 μl of PBS supplemented with 10% FBS (this serves to wet the filter and allow more cells to reach the end of the slide) was pipetted into the well of the cytology funnel and centrifuged at 800 x g for 3 minutes. 100 μl of the cell sample (5.0 x 10^3 cells) were immediately aliquot into the well of the cytology funnel and centrifuged at 800 x g for 3 minutes. The cytology funnel was carefully removed from the slide in an upwards vertical motion to prevent smearing and the slide was air dried for 2 hours at room temperature. For fixation of the cells onto the slide, the slide was immersed in 100% methanol for 10 seconds and air dried overnight at room temperature.

2.2.5.3 Colony Forming Assay - Wright-Giemsa Staining

Cells that have been fixed onto the slide the previous night using the cyotspin protocol were stained using Accustain Wright-Giemsa Stain (Sigma-Aldrich) in the following protocol. 50 ml of the Accustain Wright-Giemsa Stain was placed into a Coplin jar and 50 ml of distilled water were placed into second Coplin jar. The slide was placed into the Accustain Wright-Giemsa Stain for 30 seconds and then removed from the stain and placed directly into distilled water for 5 minutes. The slides were removed and air dried overnight at room temperature. The dried stained slides were observed under an inversed light microscope (Zeiss axiovert 200M) and the various haematopoietic cells types were identified, counted and images were recorded using Axiovision software.

2.2.6 Quantitative Gene Expression Analysis

2.2.6.1 PCR Synthesis of Transcripts - Primer Design

Primers were designed using the online primer design tool, OligoPerfect Designer [234] (Invitrogen). The annealing temperature was set in the range of 55 – 75°C (ideally 60°C), the optimal size for each primer should be between 17 - 24 bases in length (ideally 20), and the primer sequence should contain a G + C base composition in the region of 45 – 65%. Information on the genes of interest including full nucleotide sequences (introns and exons) were obtained from the Entrez Gene database within the National Center for Biotechnology Information (NCBI) website [235]. Within the Entrez gene report, contained a URL link to Ensembl [236] which revealed any reported transcript splice variants that may exist and upon selection of the required transcript, the full nucleotide sequence was shown and fragmented so that exons and introns sequences were clearly marked. This nucleotide sequence was imported into OligoPerfect and used in the primer selection process. The product size (amplificon) and the region of amplification within the imported sequence was dependent on whether the primers were for standard Polymerase Chain Reaction (PCR) or quantitative Polymerase Chain Reaction (qPCR). The primer design within the project for both standard Polymerase Chain Reaction (PCR) and quantitative Polymerase Chain Reaction (qPCR) followed the same rules as previous described but differed primarily on the size of the product, with the standard PCR product size being between 300 – 500 bp and for qPCR, a shorter 50 – 150 bp. The primary reason for the difference in products sizes for the two PCRs is that with large amplicons, the qPCR quantitative results tend to be more inconsistent due to poor amplification efficiencies [237]. However with standard PCR results the presence and size of the product is of greater importance than quantity, and thus larger amplificons in general will offer clearer results (as small qPCR amplificons may be masked by primer dimmers when observed on a agarose gel).

2.2.6.2 Ribo-nucleic Acid (RNA) Extraction

Ribo-nucleic acid (RNA) was insolated from cells using TRIzol reagent which immediately lyses the cells upon contact, dissolves cell components, and crucially maintains and protects the integrity of the RNA [39]. The cell suspension was transferred to a 1.5 ml microfuge tube and centrifuged at 1500 x g for 3 minutes to pellet the cells. The supernatant was discarded and 1 ml of TRIzol reagent was add to the cell pellet and this mixture was homogenised by incubation at room temperature for 5 minutes to facilitate the complete dissociation of nucleoprotein complexes. 10 µg of glycogen (Invitrogen) were added to the homogenised sample as a carrier before 200 µl of chloroform (VWR) were added for phase separation. The microfuge tube was shaken vigorously by hand for 15 seconds and incubated at room temperature for 15 minutes, before centrifuging at 12,000 x g for 15 minutes at 4°C. The upper clear aqueous phase of the sample was transferred to a clean 1.5 ml microfuge tube, and the RNA was precipitated by the addition of 500 µl of isopropyl alcohol (VWR). The sample was incubated at room temperature for 10 minutes and centrifuged at 12,000 g for 10 minutes at 4°C. The supernatant was removed and the RNA pellet was washed with 1 ml of 75% ethanol which was mixed by vortexing, and followed by centrifugation at 7,500 x g for 5 minutes at 4°C. The supernatant was discarded and the tube was carefully inverted and placed on a paper towel for 5 minutes, and then the RNA pellet was air dried at room temperature until all of the ethanol had evaporated off. The RNA pellet was dissolved in 10 µl of nuclease-free water and incubated at 60°C for 10 minutes to increase solubility of the pellet. Once dissolved the RNA was checked for its quality and quantity using a Nanodrop Spectrophotometer (Thermo Scientific).

The Nanodrop Spectrophotometer was used to qualify and quantify the amount and concentration of extracted RNA for each sample. The sensors of the spectrophotometer were cleaned with distilled water to remove any dried sample that might have been present before using. 2 µl of distilled water was pipetted onto the bottom sensor to initialise and calibrate the spectrophotometer. The sensors were wiped clean and another 2 µl of distilled water was placed onto the bottom sensor to set the baseline to zero (blanking). The sensors were wiped clean and 2 µl of the extracted RNA sample was pipetted onto the bottom sensor and the samples was measured. The overall RNA quantity was calculated by the recorded concentration reading (ng per µl), and as the upper limit for sample concentration on the Nanodrop Spectrophotometer is 3000 µg / µl, it was important to sufficiently dilute the samples as any concentration higher than this would result in inaccurate quality ratios. The overall RNA quality was given by the ratio of the measure of absorbance at the wavelength of

260 (A_{260}; RNA) and at the wavelength of 280 (A_{280}; protein). In general the identification of a very pure RNA sample would result in an A_{260} / A_{280} ratio of 2.0, but studies have shown that absorption is highly pH dependent [238], and thus dependent on the buffer used, for example, if nuclease-free water (pH 6-7) was used then the optimum A_{260} / A_{280} ratio would be 1.85, and if TE (pH 8) was used the optimum ratio would be 2.14.

2.3.6.3 Reverse Transcription (RT)

The extracted RNA was reverse transcribed into cDNA for quantitative polymerase chain reaction (qPCR) analysis. However before the RNA was reverse transcribed, the sample was treated with TURBO DNase (Ambion) to clear genomic DNA contamination by the cleavage of double stranded DNA which can contaminate the qPCR and result in false positive signals. 5 µl of TURBO DNase Buffer was added to 50 µl of RNA (<10 µg) within a 500 µl microtube, and thoroughly mixed by pipetting. 1 µl of TURBO DNase enzyme (2 units) was added to the 55 µl mixture, and incubated at 37°C for 30 minutes (during which up to 2 µg of genomic DNA was be removed in the reaction [239]). The TURBO DNase enzyme was inactivated by the addition of 5 µl of DNase Inactivation Reagent into the RNA mixture, incubated at 22 °C for 5 minutes, and centrifuged at 10,000 x g for 2 minutes (to pellet the DNase Inactivation Reagent). The supernatant (containing the RNA) was transferred into a fresh 500 µl microtube, and a maximum volume of (1 pg to 5 µl)

2.2.6.4 Quantitative Polymerase Chain Reaction (qPCR)

Real-time quantitative polymerase chain reaction (qPCR) was used to accurately assess the active / inactive state of any gene of interest within the genome by the detection of cDNA (produced from the RT reaction) using SYBR green. SYBR green is one of the most sensitive stains for the detection of double stranded DNA by binding with high affinity (but it also binds to single-stranded DNA albeit with lower affinity [240]), and when the DNA-SYBR green complex is excited at a wavelength of 488 nm (blue light) it emits green light at 522 nm. Due to the small volumes of components used within each reaction (total volume of 10 µl), qPCR is very sensitive to pipetting errors and therefore each reaction was performed in triplicate. The components of the qPCR master mix reaction (Table 21); primers used are specified in Appendix - Table 42) were combined in a 1.5 ml microcentrifuge tube, and 3 µl of cDNA was

added to the same tube. 30 µl of this qPCR-cDNA mixture was pipetted into three wells (10 µl per well) of a 384-well PCR plate (Applied Biosystems), and the plate was centrifuged down at 12,000 x g for 5 minutes to remove any trapped air bubbles within the occupied wells. The plate was placed inside an Applied Biosystems 7900HT Fast Real-Time PCR system, and the machine was programmed to run the experimental protocol. This protocol consisted of an initiation incubation stage at 95˚C for 30 seconds, followed by 39 amplification cycles that consisted of 95˚C for 30 seconds, followed by annealing of primers (at primer specific temperatures for 30 seconds), elongation at 72ºC for 20 seconds, with single fluorescence data collection at the specified melting point of the product (determined from the melting curve). A melting curve (60-95ºC with a temperature transition rate of 0.1ºC per second and continuous fluorescence data collection) was also produced which was used to check the quality of the primers (ensure a single product was produced), as well as to provide the optimal temperature at which fluorescence data should be acquired. The fluorescence of the stably expressed reference genes (for example, *GAPDH, SDHA, RPL13A, TBP*), was also measure and used for the normalisation of the target genes

QUANTITATIVE POLYMERASE CHAIN REACTION (QPCR) MASTER MIX

Components	Quantity	Supplier
SYBR Green JumpStart Taq ReadyMix	15 µl	Sigma
Forward Primer	1.5 µl	Sigma
Reverse Primer	1.5 µl	Sigma
Nuclease-free Water	9 µl	Sigma

Table 21: Quantitative Polymerase Chain Reaction (qPCR) Master Mix Composition

2.2.7 Statistical Analysis

All flow cytometry and quantitative polymerase chain reaction (qPCR) results were analysed for statistical significance with the threshold level of probability (p-value) defined at 0.05 (5%) or 0.01 (1%) for significant results. Statistical significance is used as a measure of how confidently one might reject the null hypothesis (no relationship between measured samples), with the null hypothesis being true if it is viewed that the observed data have at least a 95% probability of occurring purely due by chance [241]. The Student's t-test was used in the calculation of statistical significance between two sets of qPCR samples based on a null hypothesis, with the calculation of p-values performed within Microsoft Office Excel 2007. Statistical analysis between microarray samples (for example, Day 0 against Day 4) was performed within GeneSpring using one-way ANOVA (ANalysis Of VAriance) test (asymptotic p-value computation and Benjamini-Hochberg multiple testing correction).

2.3 Results

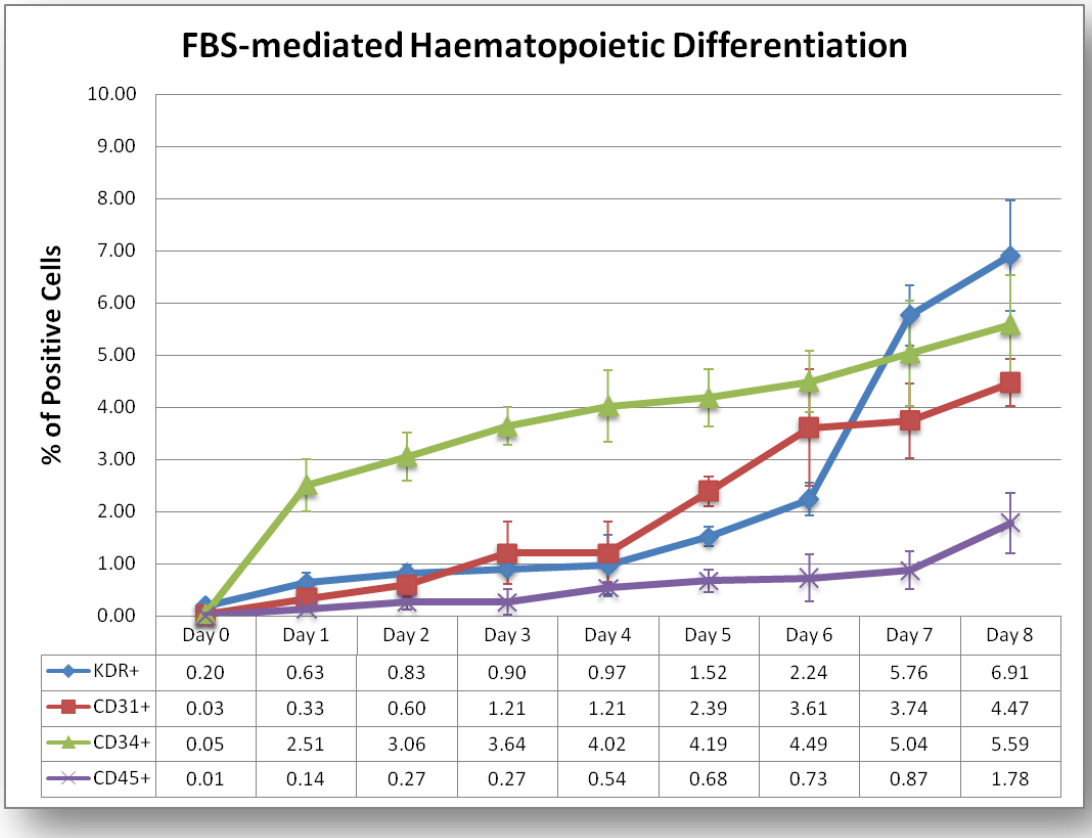
The experiments described in this section were performed in order to assess the differentiation potential and efficiency of human embryonic stem cells in the formation of the earliest haematopoietic progenitor. As previously described in Chapter 1, the first haematopoietic progenitor which emerges during development is the haemangioblast. In essence, the primary aim of this PhD study is to analyse and compare the gene expression profile of this haemangioblast population to a population of undifferentiated, pluripotent stem cells in a bid to discover which genes are required to be over-expressed and which need to be repressed in order to achieve this haemangioblast state. Numerous studies have already suggested that this primitive haematopoietic progenitor may be identified by a number of key cell surface antigens, and as previously explained these are CD31, KDR, CD34, CD117, and CD144 [143, 180, 182-185, 242-246].

The primary objective within this chapter was to attempt to generate the haemangioblast population by means of differentiating HESCs. The theory was that this generated haemangioblast population may be identified and isolated by flow cytometry using a combination of all of the specified haemangioblastic cell surface markers (CD31, KDR, CD34, CD117, and CD144). However, as the preliminary results will reveal, the efficiency of haemangioblast differentiation is very poor and despite numerous efforts to increase the efficiency, the required number of cells for the microarray study remained unachievable. A microarray analysis requires an extraordinarily large number of cells, estimating in the region of hundreds of thousands, and therefore the only viable solution was to decrease the number of haemangioblast markers, which in turn increases the number of positive cells, but unfortunately this as a result reduces the purity of the haemangioblast population. The markers KDR and CD31 in combination were used to identify and isolate an enriched haemangioblast population which will be analysed by microarray at varying stages of HESC EB differentiation in order to assess and examine transcripts expressed during early haematopoietic development.

2.3.1 FBS mediated Haematopoietic-Induction

2.3.1.1 Detection of Haematopoietic Progenitors by Flow Cytometry

Studies have shown that cell surface markers may be used in the identification of developmental cell types, and that CD31, CD34, and KDR in combination are expressed on early haemangioblast cells with both haematopoietic and endothelial potentials [157, 185, 193, 213, 242, 243, 247]. The results from the fluorescence-activated cell sorting (FACS) analysis as shown in Figure 20 indicates the presence of a number of haematopoietic antigens on the membrane of cells from differentiated embryoid bodies (EBs) which include KDR, CD31, and CD34. The emergence of these markers of early haematopoiesis alone suggests that pluripotent HESC line H9 have the capacity to differentiation to the mesoderm lineage in the presence of foetal bovine serum (FBS).

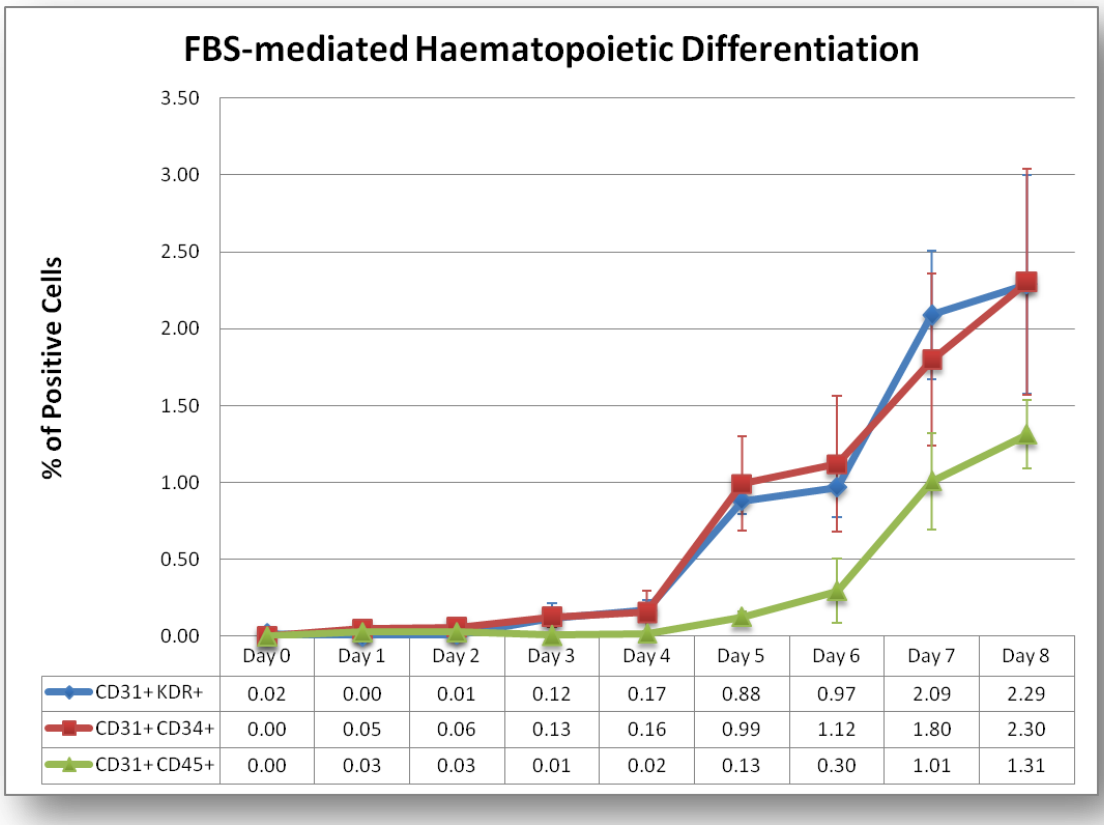


	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
KDR+	0.20	0.63	0.83	0.90	0.97	1.52	2.24	5.76	6.91
CD31+	0.03	0.33	0.60	1.21	1.21	2.39	3.61	3.74	4.47
CD34+	0.05	2.51	3.06	3.64	4.02	4.19	4.49	5.04	5.59
CD45+	0.01	0.14	0.27	0.27	0.54	0.68	0.73	0.87	1.78

Figure 20: Kinetics of single haematopoietic markers during FBS-mediated differentiation

The percentages of KDR, CD31, CD34, and CD45 positive EB cells were measured by flow cytometry from day 0 to day 8 of FBS-mediated differentiation. Isotype controls were used to represent the negative population, and error bars were calculated using the standard errors of the mean with a sample size of 3.

The flow cytometry results within Figure 20 show the kinetics of a number of haemangioblast markers with the overall trend showing a gradual increase of antigen-positive cells throughout the eight days of differentiation. Upon closer inspection, differing rates of antigen presentation occurred between the haematopoietic markers, with CD34+ cells appearing on about 2.51% of the cells by day 1, and both CD31+ and KDR+ cells only present on 0.33% and 0.63% respectively of the total differentiated EB population at the same stage. While antigens of CD31 appear on increasing numbers of cells steadily through differentiation and peaking at day 8 covering about 5.59% of the cell population, it is not until day 4 when both CD31 and KDR antigens increase to about 1% of cells. From here on, KDR and CD31 increase by on average 130% and 200% respectively from day 4 to day 6 of differentiation, and whereas CD34+ cell numbers only increased by about 10% during the same period. From day 6 to day 8, the rate of KDR expression increased a further 200% and was present on about 7% of all differentiating cells by day 8, with CD31 expression on about 4.5% of all cells. The antigen expression pattern of CD45 (a pan-haematopoietic marker) was included to detect late / mature haematopoietic cells which was observed at very low levels through differentiation, however this may be due to non-specific binding of the antibody rather than real expression.



	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
CD31+KDR+	0.02	0.00	0.01	0.12	0.17	0.88	0.97	2.09	2.29
CD31+CD34+	0.00	0.05	0.06	0.13	0.16	0.99	1.12	1.80	2.30
CD31+CD45+	0.00	0.03	0.03	0.01	0.02	0.13	0.30	1.01	1.31

Figure 21: Kinetics of double-positive haematopoietic markers during FBS-mediated differentiation

The percentages of double positive CD31+KDR+, CD31+CD34+, and CD34+CD45+ EB cells were measured by flow cytometry from day 0 to day 8 of FBS-mediated differentiation. Isotype controls were used to represent the negative population, and error bars were calculated using the standard errors of the mean with a sample size of 3.

The analysis of double positive markers as shown in Figure 21 shows that the number of positive cells is far lower than that of individual single markers as one might expect. Days 4, 6, and 8 have been singled out in Figure 22 during the 8 days differentiation analysis as key points of early haematopoietic development and closer examination of Figure 21 appears to support this hypothesis. CD31+KDR+ and CD31+CD34+ cells each account for on average 0.17% of the total cells at day 4 but this rapidly increases by day 6 by 470% and 830% respectively (Figure 22). This large and sudden change may suggest the beginnings of haematopoietic specification from day 4 and therefore it would be interesting to examine gene expression at this stage of differentiation. The gradient of CD31+KDR+ and CD31+CD34+ double positive populations appear to plateau from day 5 to day 6, before increasing rapid from day 6 to day 8, again suggesting another interesting point of analysis in order to explain this dip in marker expression between day 4 and day 8. From day 6, CD31+CD45+ appears to increase with greater urgency at a similar rate to both CD31+KDR+ and CD31+CD34+ populations, suggesting an establishment of early haematopoiesis.

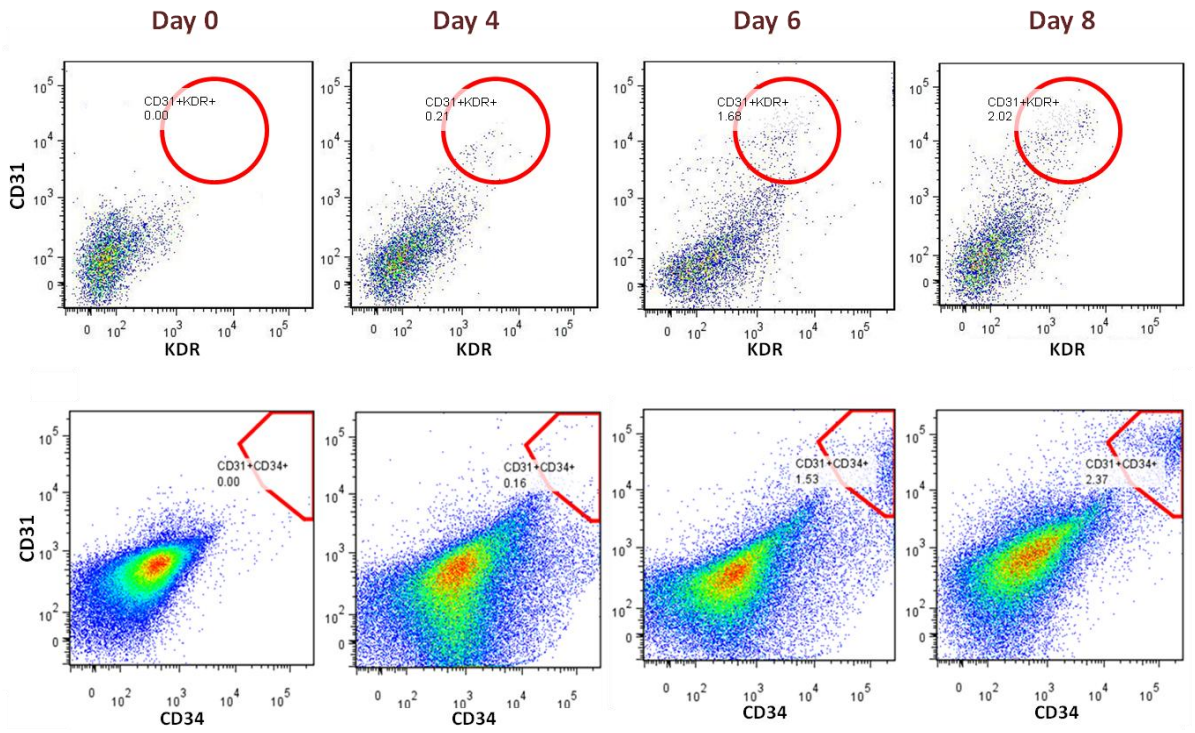


Figure 22: Kinetics of KDR+CD31+ and CD31+CD34+ population during FBS-mediated differentiation

The percentages of double positive CD31+KDR+ and CD31+CD34+ EB cells were measured by flow cytometry at day 0, 4, 6, and 8 of FBS-mediated differentiation. Isotype controls were used to represent the negative population, and 200,000 events were recorded.

Population	Events	%Total
All Events -	200,000	100.0
CD34	12,280	6.14
CD31	9,560	4.78
CD45	4,140	2.07
KDR	15,020	7.51
NOT (CD45)	195,860	97.93
NOT (CD34)	187,720	93.86
NOT (CD31)	190,440	95.22
NOT (KDR)	184,980	92.49
CD31 AND CD34	4,740	2.37
CD31 AND KDR	4,700	2.35
CD31 AND CD45	2,580	1.29
CD34 AND CD45	5,620	2.81
NOT (CD31) AND NOT (CD34)	195,260	97.63
NOT (CD31) AND NOT (KDR)	195,300	97.65
NOT (CD31) AND NOT (CD45)	197,420	98.71
NOT (CD34) AND NOT (CD45)	194,380	97.19
CD31 AND CD34 AND KDR AND NOT (CD45)	**380**	**0.19**

Figure 23: Tabulated results from H9 Day 8 of differentiation with FBS-mediated media

This table of flow cytometry results show a sample of the possible combinations of the 4 markers (CD31, CD34, KDR and CD45) and the corresponding number and percentage of cells. A total of 200,000 differentiated day 8 cells were analysed and recorded by flow cytometry.

As previously discussed, the early haematopoietic progenitor known as the haemangioblast may be identified during *in vivo* development by a combination of cell surface antigens (CD31, KDR, and CD34) and through flow cytometry analysis, these markers may be used to sort for these cells *in vitro* [157, 185, 193, 213, 242, 243, 247]. A microarray analysis requires a minimum of 100,000 sorted cells per time point, and at the time point of day 8 (the stage whereby these markers are at the highest expression during differentiation) these markers are expressed on just 0.19% of cells. Therefore, in order to sort for the minimum number of CD31, CD34, KDR (and not CD45) cells, over 53 million cells will be required to begin differentiation with (assuming zero loss of cells through media changes, cell dissociation and 100% survival). In practise, the predicted number of cells required would easily be double this calculated figure, and in addition when attempting to sort for this population at earlier days (day 4 and day 6) where the percentage of haemangioblast markers are even lower than at day 8 will mean that the number of cells required to start differentiation will be astronomical. Therefore, a different approach must be taken in order to sort enough of this cell population, and an alternative differentiation approaching utilising a serum-free base media in combination with haematopoiesis-inducing cytokines may be the solution [157].

2.3.2 Cytokine-Induced Haematopoietic Differentiation

In order to improve haematopoietic differentiation efficiencies, studies have sought to using serum-free media in combination with haematopoiesis-inducing cytokines which were added at specific stages to control and direct differentiation [108, 157].

2.3.2.1 Detection of Haematopoietic Progenitors by Flow Cytometry

The flow cytometry plots of the kinetics of CD31+KDR+ and CD31+CD34+ cells (Figure 24) of cytokine-optimised differentiation show distinct, discrete double positive populations. The kinetics of the CD31+KDR+ population increased from 0.17% on day 0, to as much as 6.26% by day 8, and CD31+CD34+ appears to be following the same trend, achieving similar numbers of double positive cells by day 8.

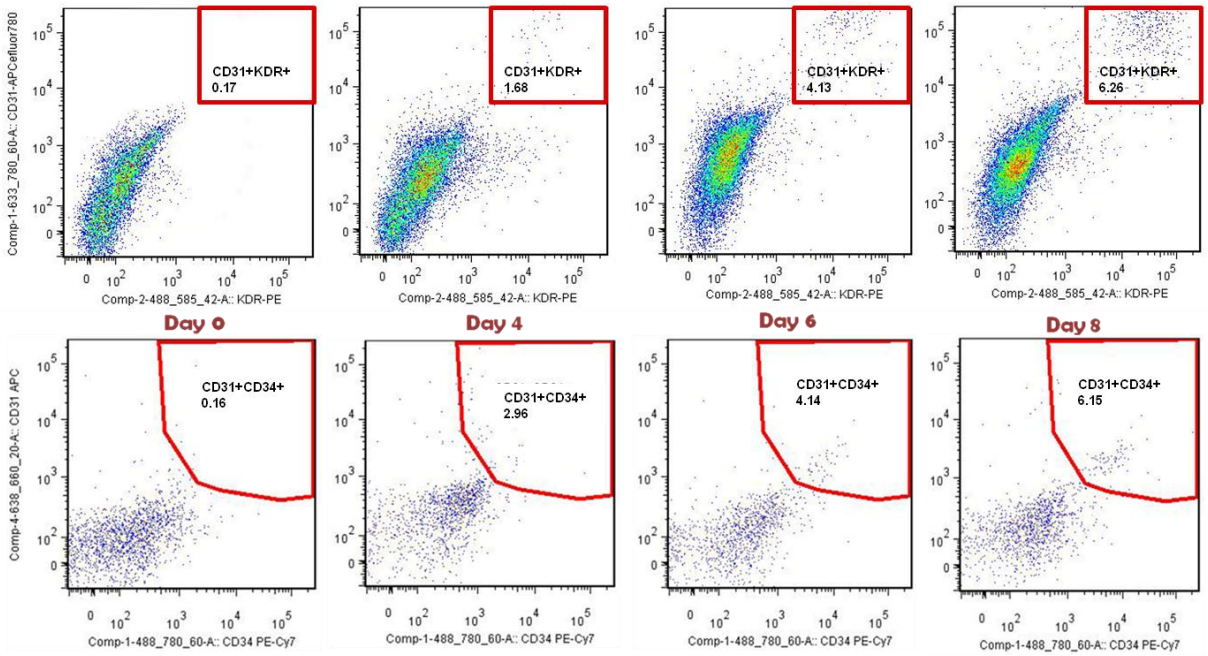


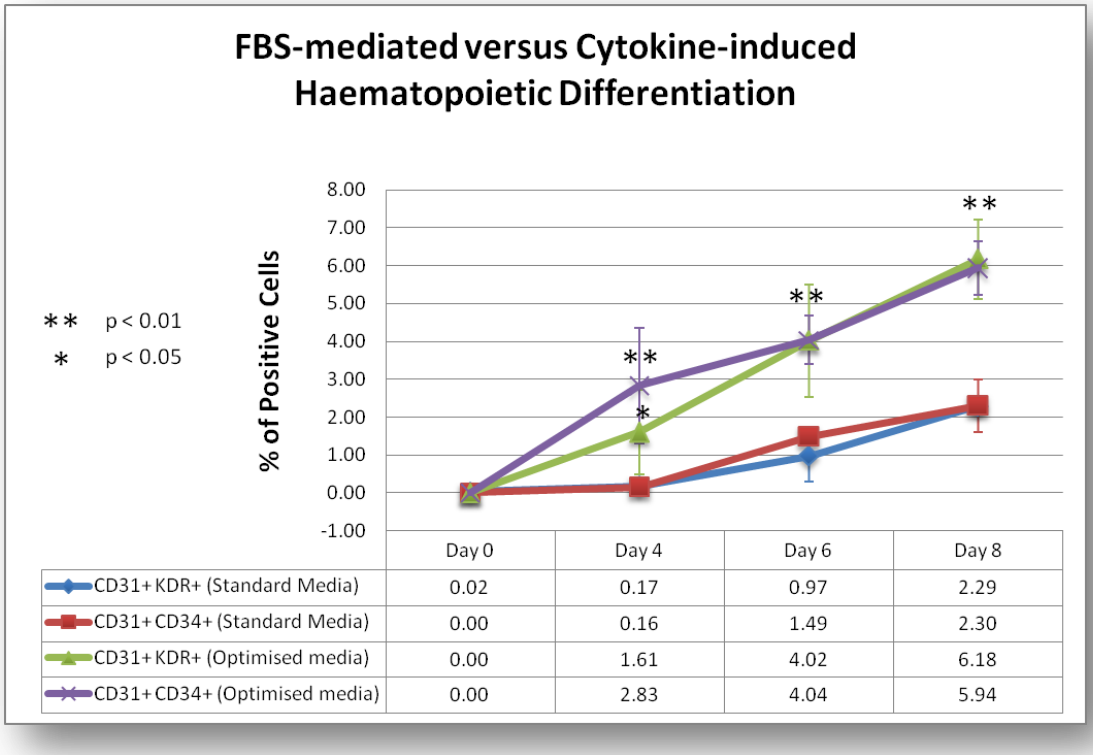
Figure 24: Kinetics of KDR+CD31+ and CD31+CD34+ population during cytokine-induced differentiation

The percentages of double positive CD31+KDR+ and CD31+CD34+ EB cells were measured by flow cytometry at day 0, 4, 6, and 8 of cytokine-induced differentiation. Isotype controls were used to represent the negative population, and 50,000 events were recorded.

In terms of stage of expression of markers, it appears that CD31 and CD34 appears on a larger number of cells much earlier during this differentiation than KDR (day 4), but by day 6,

KDR expression increases rapidly and recoups the deficit, and by day 8 marginally higher numbers of double CD31+KDR+ cells are observed.

Comparisons between FBS-mediated differentiation (standard differentiation) and cytokine-induced differentiation (optimised differentiation) reveal significant differences in the efficiency of expression of haematopoietic markers during differentiation (Figure 25). The same HESC line (H9) was used in this comparative study for both media types, and flow cytometry results of markers' expression shown in Figure 25 display significant enhancements in haematopoietic differentiation efficiency from standard differentiation media to optimised differentiation media across all differentiated days. The average increase in efficiency in the generation of CD31+KDR+ and CD31+CD34+ populations during differentiation at day 4 was 1245%, at day 6 was 228%, and at day 8 was 164%.



	Day 0	Day 4	Day 6	Day 8
CD31+KDR+ (Standard Media)	0.02	0.17	0.97	2.29
CD31+CD34+ (Standard Media)	0.00	0.16	1.49	2.30
CD31+KDR+ (Optimised media)	0.00	1.61	4.02	6.18
CD31+CD34+ (Optimised media)	0.00	2.83	4.04	5.94

Figure 25: Comparisons between standard and optimised media on the kinetics of double-positive haematopoietic cells during differentiation

The percentages of double positive CD31+KDR+, CD31+CD34+, and CD34+CD45+ EB cells were measured by flow cytometry from day 0 to day 8 of FBS-mediated differentiation. Isotype controls were used to represent the negative population, and error bars were calculated using the standard errors of the mean with a sample size of 3.. Student T-test was used to calculate the p-values which show the degree of variation and provides an assessment of significance between the two types of media.

2.3.2.2 Demonstration of Colony-forming Potential of Embryoid Bodies (EBs)

The colony forming assay provides an essential measure of the haematopoietic potential of the differentiated cells and the results in Figure 26 show that the haematopoietic colony forming units formed were very similar to those described earlier in Figure 19. A histological analysis of the CFUs (Figure 27) reveals Wright-Giemsa stained cells that resemble a variety of primitive haematopoietic cell types that would be found in adult peripheral blood, and these include plasmacytoids, myeloid, monocytoid, erythroblast, granulocytes, and eosinophils. The results show that only 0.07% of the 30,000 cells which were plated from day 4 differentiated EBs produced colony forming units, with this percentage increasing to 0.16% and 0.34% from day 6 and day 8 EBs respectively (Figure 26). There appears to be some correlation between increasing numbers of CD31+, CD34+, and KDR+ cells with increasing colony forming units during the differentiation time course, however these results alone does not prove this. A Pearson product moment correlation coefficient was calculated between the percentage of CD31+CD34+ cells and CD31+KDR+ cells (from FACS analysis with optimised media; Figure 25) against the number of CFUs observed at each day to discover which cell type more accurately predicts CFU potential. Due to the limited number of samples within each dataset, the resultant correlation coefficients from both CD31+CD34+ ($r=0.997$) and CD31+KDR+ ($r=0.972$) cells are very similar and both demonstrate high correlation with CD31+CD34+ being the better from these results.

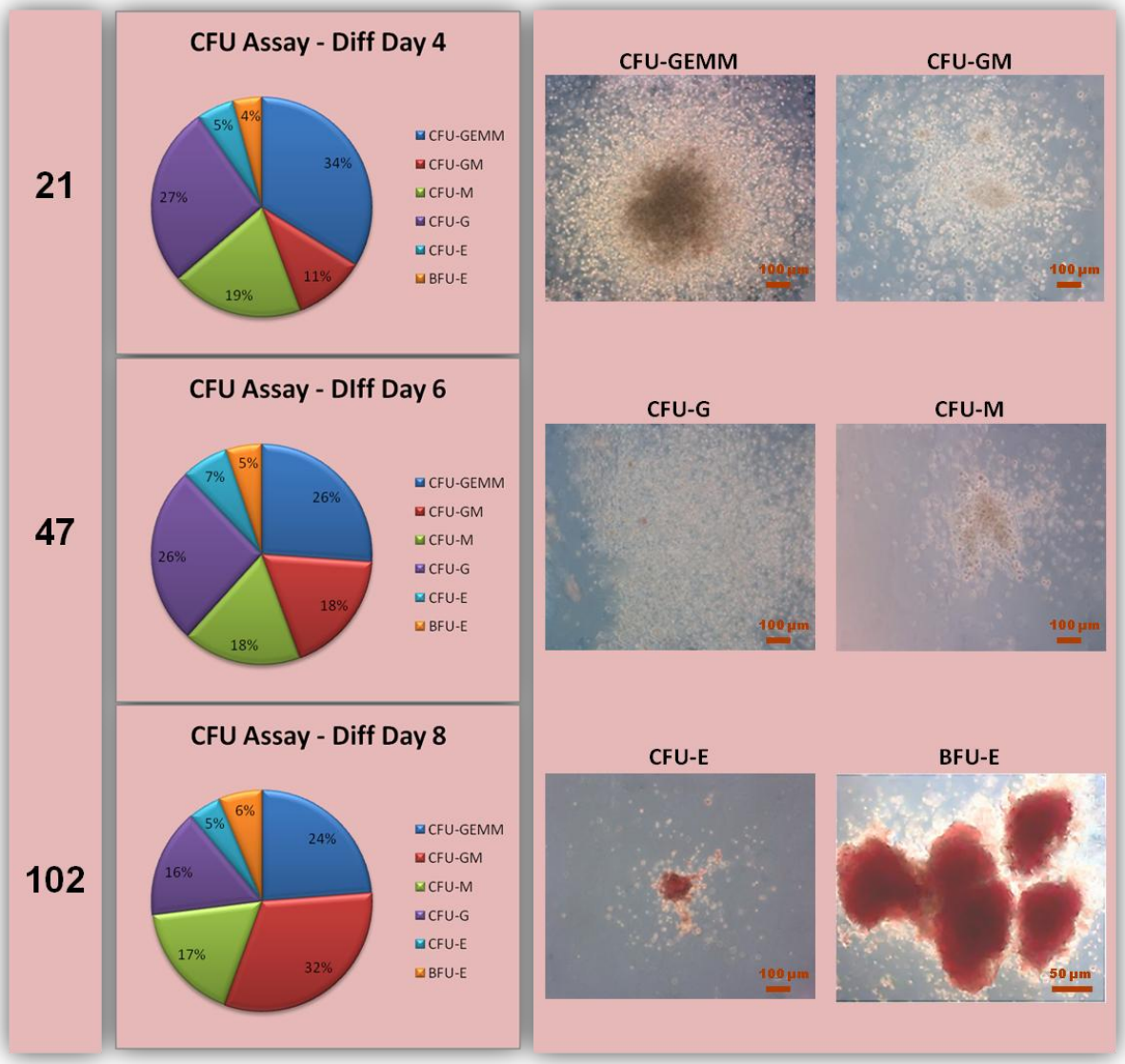


Figure 26: Distribution of Colony Forming Units from EB cytokine-induced haematopoietic differentiation

Results from the colony forming units (CFU) assay showing the recorded numbers of total CFUs (left column), the distribution of CFU types (middle column) produced from 4, 6 and 8 day haematopoietic differentiations from H9 HESCs, and CFU images (right column) were taken on a Zeiss Aviovert 200M inverted microscope with an AxioCam HRC camera. Percentages and numbers where calculated from 2 sample sets of colony forming unit assays.

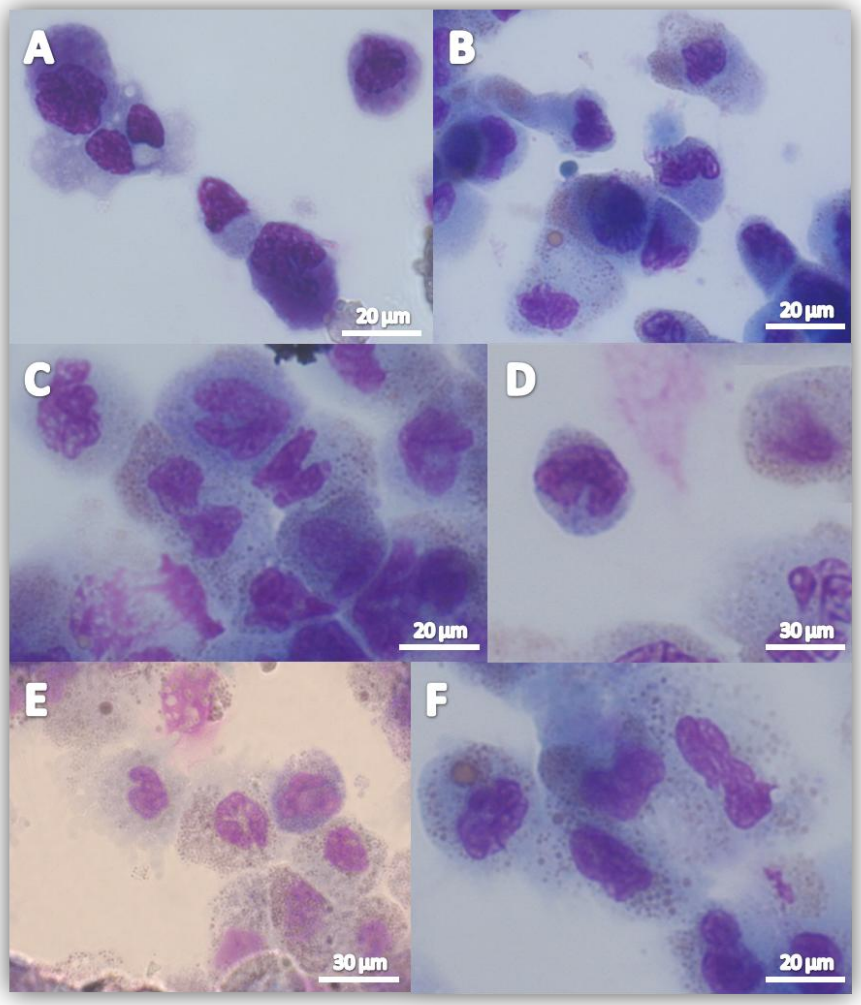


Figure 27: Histology of hESC derived CFU colonies

Slides of colony forming units were produced from the CFU assays (Figure 27) and stained with Wright-Giemsa. These images show a number of cell types of haematopoietic appearance: (A) early plasmacytoids, (B) myeloid , monocytoid and erythroblast, (C) granulocytes and eosinophils, (D) monocytoid and granulocyte, (E) granulocytes, (F) myeloid and eosinphil granulation. These images were taken by a Zeiss Axio microscope using oil immersion x100 lens.

An analysis of the distribution of haematopoietic colony sub-types in each CFU assay reveals a preference of CFU-GEMM type colonies at day 4 (34%), with CFU-G and CFU-GEMM types at day 6 (26% and 26%), and CFU-GM type at day 8 (32%). The appearance and distribution of these colony sub-types do not appear to be representative of primitive haematopoiesis where primitive erythrocytes (CFU-E) would be expected to be the dominant sub-type [248, 249]. This suggests that the optimised differentiation media may not contain the necessary factors required for primitive haematopoiesis and hence primitive erythropoiesis. These results do however suggest that haematopoietic progenitors are formed during differentiation from HESCs, however the percentage of these progenitors together with their efficiency in the production of colony forming units remains very low, demonstrating the fact that additional haematopoietic factors are required.

2.3.2.3 Real-time qPCR Analysis of Haematopoietic Genes

Real-time quantitative polymerase chain reaction (qPCR) was used in the assessment of gene expression of a select number of lineage-specific genes through the 8 days of EB haematopoietic differentiation. The first row of qPCR plots within Figure 28 show the loss of pluripotency during differentiation, by the down-regulation of pluripotency genes *OCT4* and *NANOG*, resulting in 14 and 33 fold down regulation respectively by day 8.

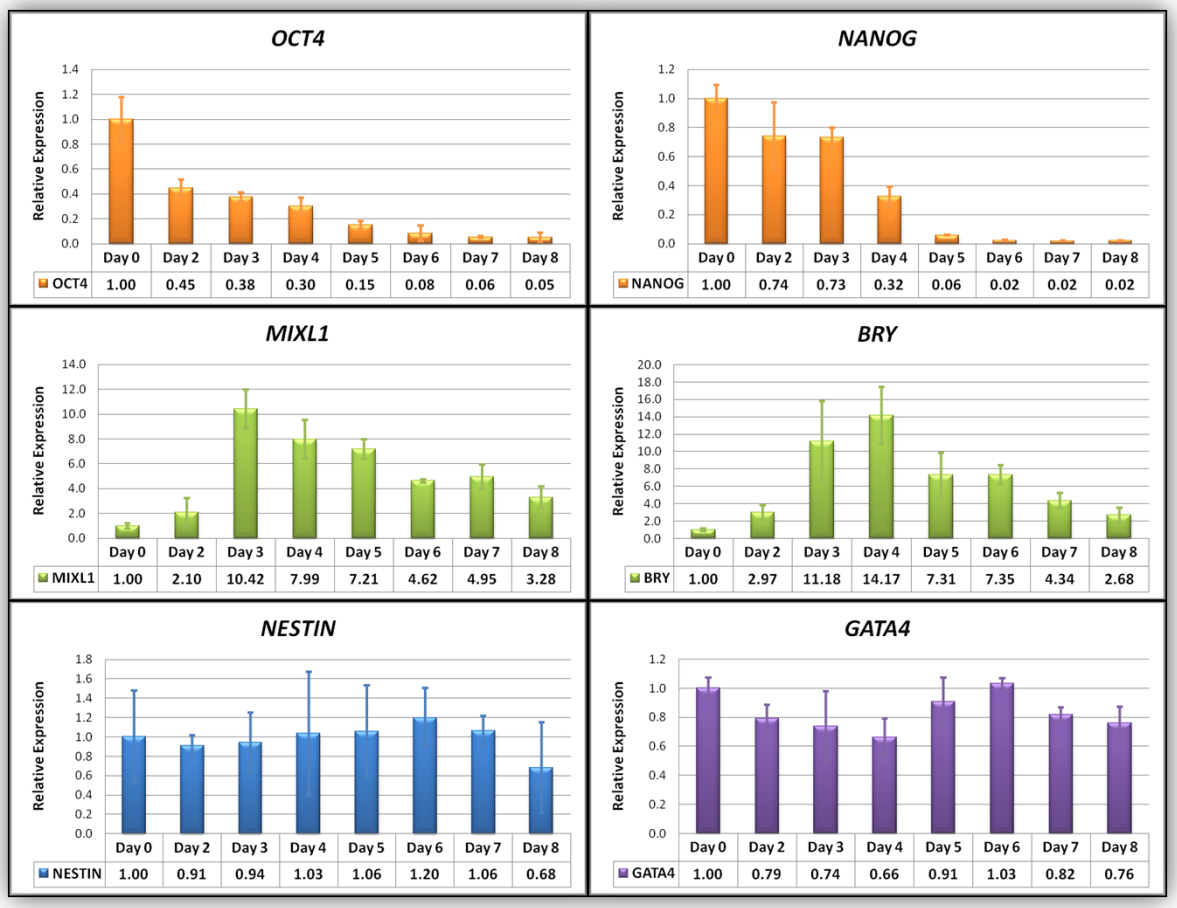


Figure 28: Real-time qPCR analysis of cells from the eight day differentiation – pluripotency, mesoderm, ectoderm, and endoderm markers

A sample of genes were selected in the identification of the different lineage types (mesoderm, ectoderm, endoderm) resulting from differentiation during the 8 day period which includes genes representative of pluripotencv (*OCT4* and *NANOG*), mesoderm (*MIXL1* and *BRY*), ectoderm (*NESTIN*), and endoderm (*GATA4*). The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate and the average gene expression values between H9 and H1 cell lines were calculated for the plots. (Separate H9 and H1 cell line qPCR data are located in the APPENDIX: Figure 83 and Figure 84)

Mesodermal-lineage genes *MIXL1* and *BRY* (Figure 28) appear to increase rapidly and peak at around day 3 and 4 of differentiation induced by the presence of bone morphogenetic protein 4 (BMP4) [250] within the media. The addition of vascular endothelial growth factor (VEGF) to the differentiation media at day 4 assists in the down-regulation of mesoderm genes, and promotes the up-regulation of haematopoietic genes (Figure 29; *CD31, CD34, KDR*, *CD117, CD144*, CD45) [157, 251]. Endodermal and ectodermal lineage markers *GATA4* and *NESTIN* (Figure 28) respectively do not appear to change in expression during the 8 days of differentiation, suggesting that the HESCs under these conditions may not differentiate towards these lineages (Figure 28). Analysis of haematopoietic markers, in particularly those identifying the haemangioblast, *KDR, CD31, CD117, CD34,* and *CD144,* in general exhibited an up-regulation expression profile through the 8 days of differentiation as shown in Figure 29. However, the intensity and rate of change of gene expression of the haematopoietic genes varies to different degrees during differentiation, providing an order of haematopoietic expression during development. From these results it can be seen that *CD34* is one of the first markers to be expressed, followed by *KDR, CD31*, and *CD117*, with *CD144* being expressed at a very low level until day 5 when its expression rapidly increases by about 50 fold at day 8 (compared to day 0) (Figure 29). *CD117* expression barely changes during differentiation with the smallest of increases over the first 8 days of differentiation, and because its expression differs so profoundly from the other haemangioblast markers, its use as a marker for the haemangioblast is questionable. The inclusion of *CD45*, a marker of differentiated haematopoietic cells only (excludes endothelial cells), shows the presence of subset of haematopoietic cells at day 6 of differentiation where its gene expression starts to gradually increase for the remainder of the differentiation.

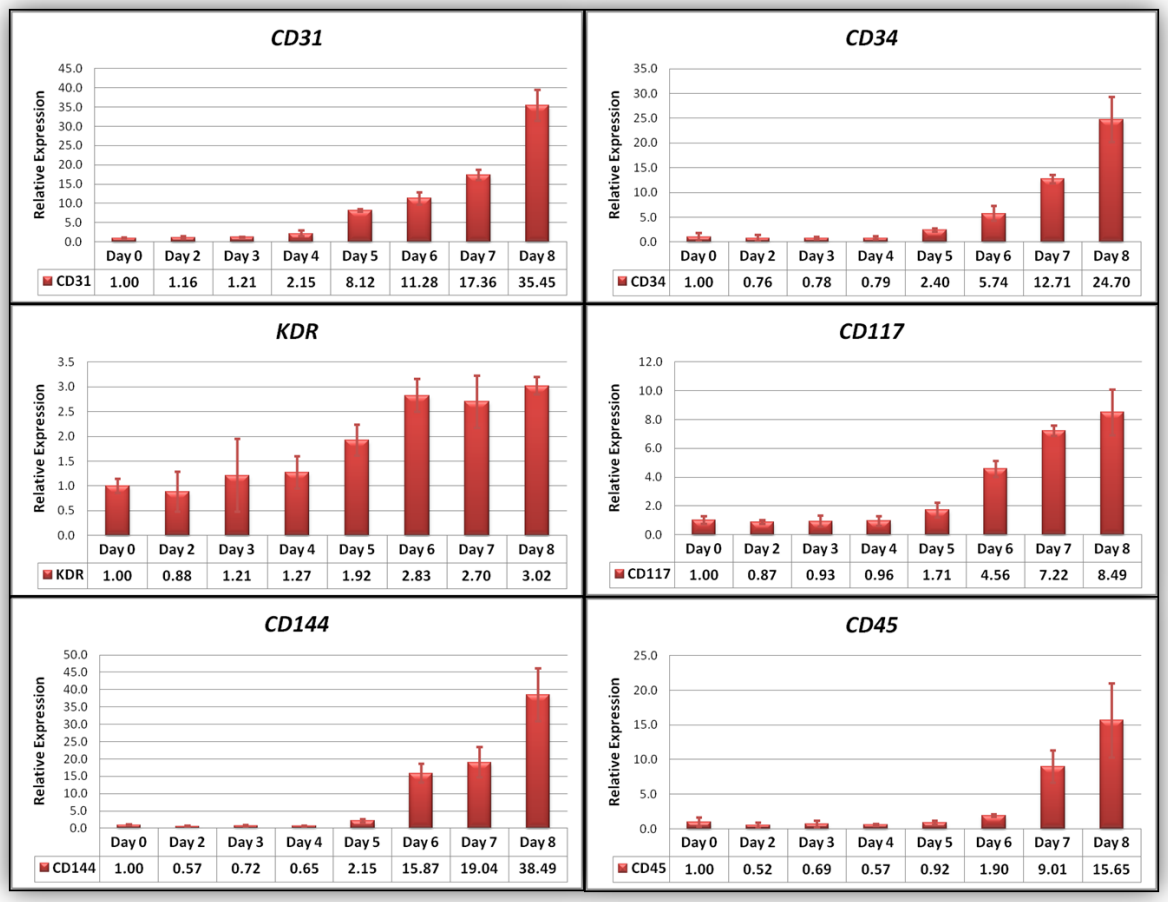


Figure 29: Real-time qPCR analysis of cells from the eight day differentiation – haemangioblast markers

A sample of haematopoietic genes (*KDR, CD117, CD31, CD34, CD144,* and *CD45*) were selected in the analysis of haematopoietic differentiation during differentiation with the optimised differentiation media. The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate and the average gene expression values between H9 and H1 cell lines were calculated for the plots. (Separate H9 and H1 cell line qPCR data are located in the APPENDIX: Figure 85 and Figure 86)

2.3.3 Haematopoietic Differentiation Optimisations

2.3.3.1 Human Embryonic Stem Cell Line

As previously discussed in Chapter 1, human embryonic stem cell lines do not all differentiate with the same haematopoietic potential even when under identical differentiation conditions and media. This varying potential of HESCs was nicely highlighted in a recent study by Ramos-Mejia *et al.* (2010) [252], which compared a number of pluripotent stem cell lines and discovered a novel link between Nodal/Activin signalling and haematopoietic differentiation potential. However, at the time of this experiment, this information was not available and *in vitro* differentiation of both HESC lines were carried out and compared in a bid to discover the optimum line for use in this haematopoietic differentiation experiment.

During early differentiation at day 4, flow cytometry results (Figure 30) reveal that the H9 cell line in general produces greater numbers of CD31+ cells, almost double than with the H1 cell line. KDR antigen expression however appears to be limited at this stage of differentiation, but the H9 cell line still appears to produce 10 fold more positive cells than the H1 cell line albeit at very low levels compared to CD31 expression. This picture changes dramatically as differentiation progresses to day 6, where CD31 expression of the H1 cell line rockets up by 286% (3.8 fold) from day 4, and at this stage H1 generates 3.47% of CD31+KDR+ cells compared with 2.85% by H9. The data at day 8 shows a reversal once again in haematopoietic potential dominance with the H9 cell line generating 7.42% of CD31+KDR+ cells and the H1 cell line producing 5.68%. In the isolation of haemangioblast cells, these are identified by the presence of a number of cell surface antigen proteins with CD31 and KDR being the two principal markers and CD31+KDR- or CD31-KDR+ cells as previously explained will not give rise to haematopoietic cell types. Therefore when examining the results from Figure 30, the best cell line for the generation of the target haemangioblast enriched CD31+KDR+ population within this differentiation setup is H9. The principal reason being because the H9 cell line generates the highest population of dual CD31+KDR+ cells at day 4, which is the time point at which the lowest population of dual positive cells are produced from the three differentiation time points (day4, day 6, day 8). The necessity of obtaining as many dual positive cells as possible at day 4 is based largely on the technical requirement of physically sorting enough for microarray analyses, with a minimum of 100,000 dual positive cells required for each time point.

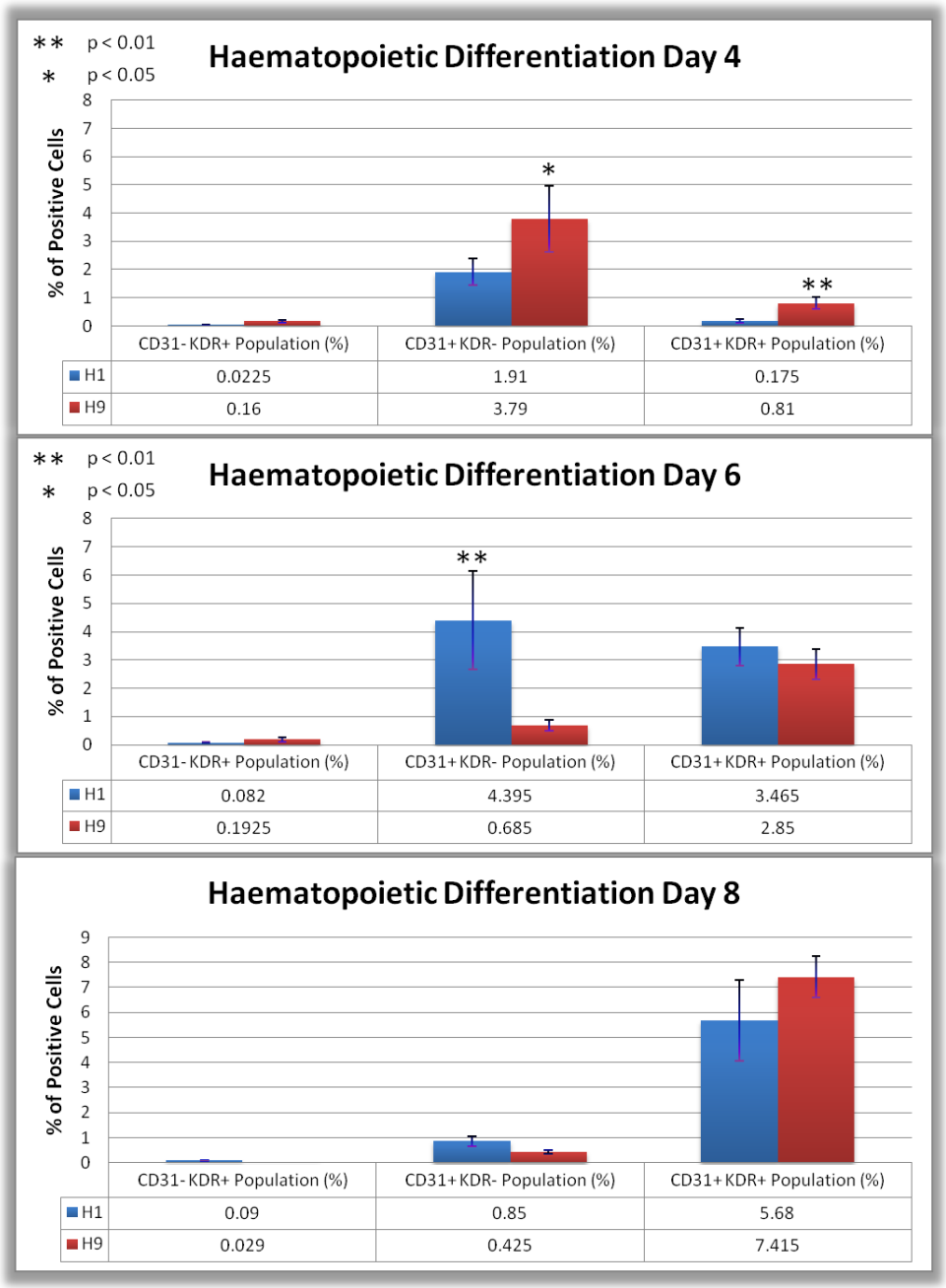


Figure 30: Comparisons between H9 and H1 lines on the effect on haematopoietic differentiation

The percentages of CD31-KDR+, CD31+KDR-, and CD31+KDR+ populations were measured by flow cytometry from the two differentiated HESC lines, H9 and H1. Isotype controls were used to represent the negative population, and error bars were calculated by standard deviation with a sample size of 3.

2.3.3.2 Embryoid Body (EB) density

The recent flow cytometry results from haematopoietic differentiation experiments using the same cell line, differentiation media, and under the same incubation conditions, remarkably produces quite diverse results. In an attempt to try to minimise experimental variables which may attribute to this diversity of data, the most obvious variable which was noticed during differentiation is EB size and EB density. The effects of EB size on the differentiation potential of HESCs have already been published [253] and attempts to repeat this experiment in the creation of custom-size EBs through the dissociated of HESCs within this experiment resulted in the total collapse of the EB after a few days of differentiation. Possible reasons for this result will be discussed in greater detail within the discussion section, and due primarily to time constraints this avenue to investigation was postponed for future work. However within this section, EB density will be investigated as the effect of this variable has not been identified within published literature but it is one which may be controlled within limited parameters.

Published literature of haematopoietic differentiation experiments typically omits the details of EB densities (and size) during differentiation, however within our laboratory, one well of a 6-well plate of EBs is composed from three confluent wells of a 6-well plate of HESCs (Figure 31). It is hypothesised that cell-to-cell contact between EBs may enhance haematopoietic differentiation as a recent monolayer co-culture differentiation study from our group has shown that cell-to-cell contact is important for haematopoietic differentiation [229] and key haematopoietic markers such as CD31 and CD34 glycoproteins are known to interact with other transmembrane proteins through cell-to-cell contact [254].

The plating and differentiation of EBs at double density (Figure 31 and Figure 32; one well of a 6-well plate of EBs is composed from six confluent wells of a 6-well plate of HESCs) ensures that each EB is always in physical contact with one or more EBs during differentiation, and the effect of this is analysed by flow cytometry.

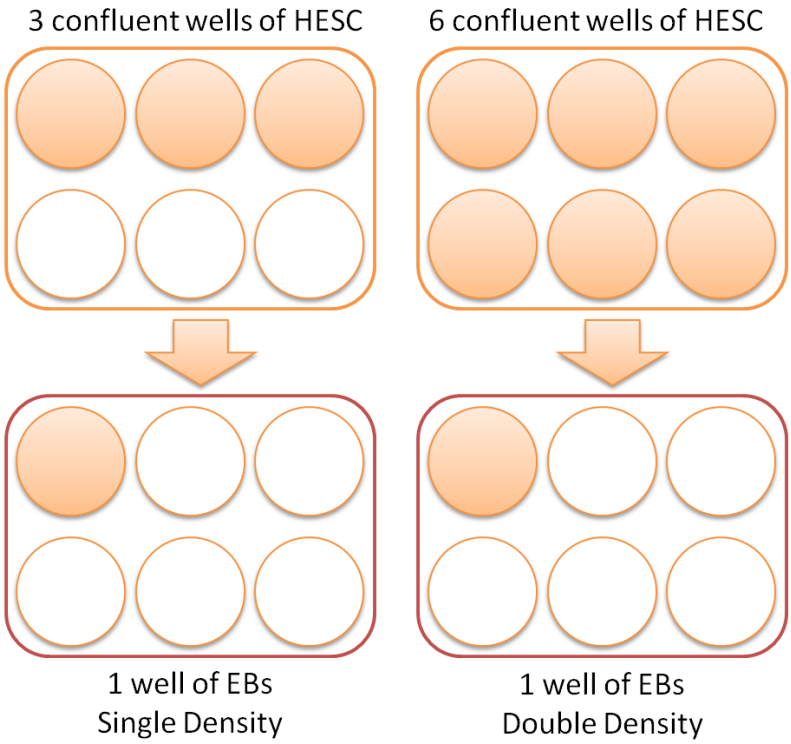


Figure 31: Composition of EB plates – Single Density against Double Density

A single well of EBs plated at single density are composed from 3 wells of HESCs, and a single well of EBs plated at double density are composed from 6 well of HESCs.

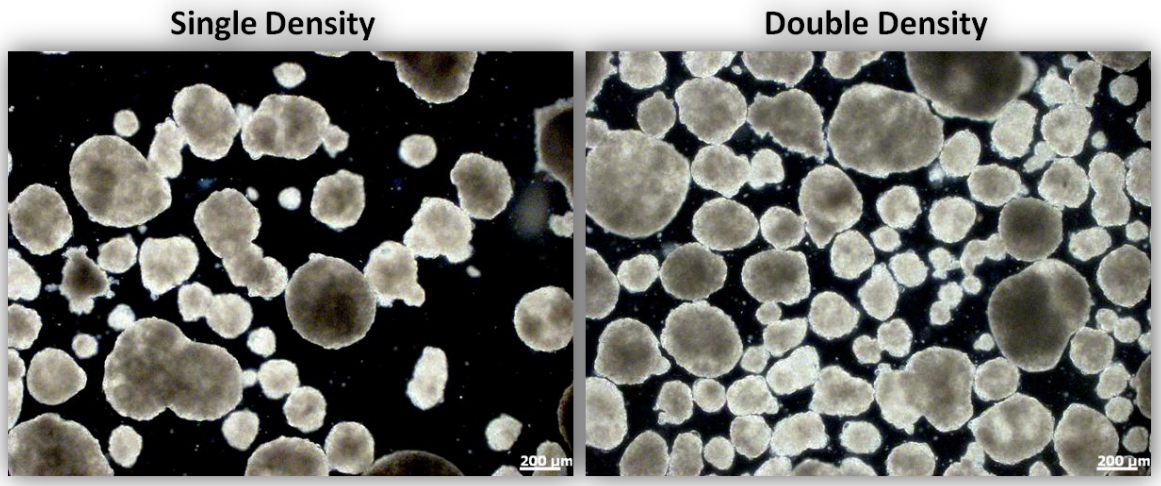


Figure 32: Single versus Double Density Embryoid Bodies (EBs)

Images showing the two differing densities of EBs during haematopoietic differentiation taken on a Zeiss Aviovert 200M inverted microscope with an AxioCam HRC camera at 10x magnification.

At day 4 of differentiation, it appears that the sparser plated single density EBs (SDEBs) produce the greatest number of CD31+KDR+ (0.60%) compared with the double density EBs (DDEBs; 0.39%). This result is largely attributed to the high percentage of CD31+ cells of which SDEBs produce 1.85 fold more than DDEBs, with percentages of KDR+ cells that are produced being very similar (Figure 33).

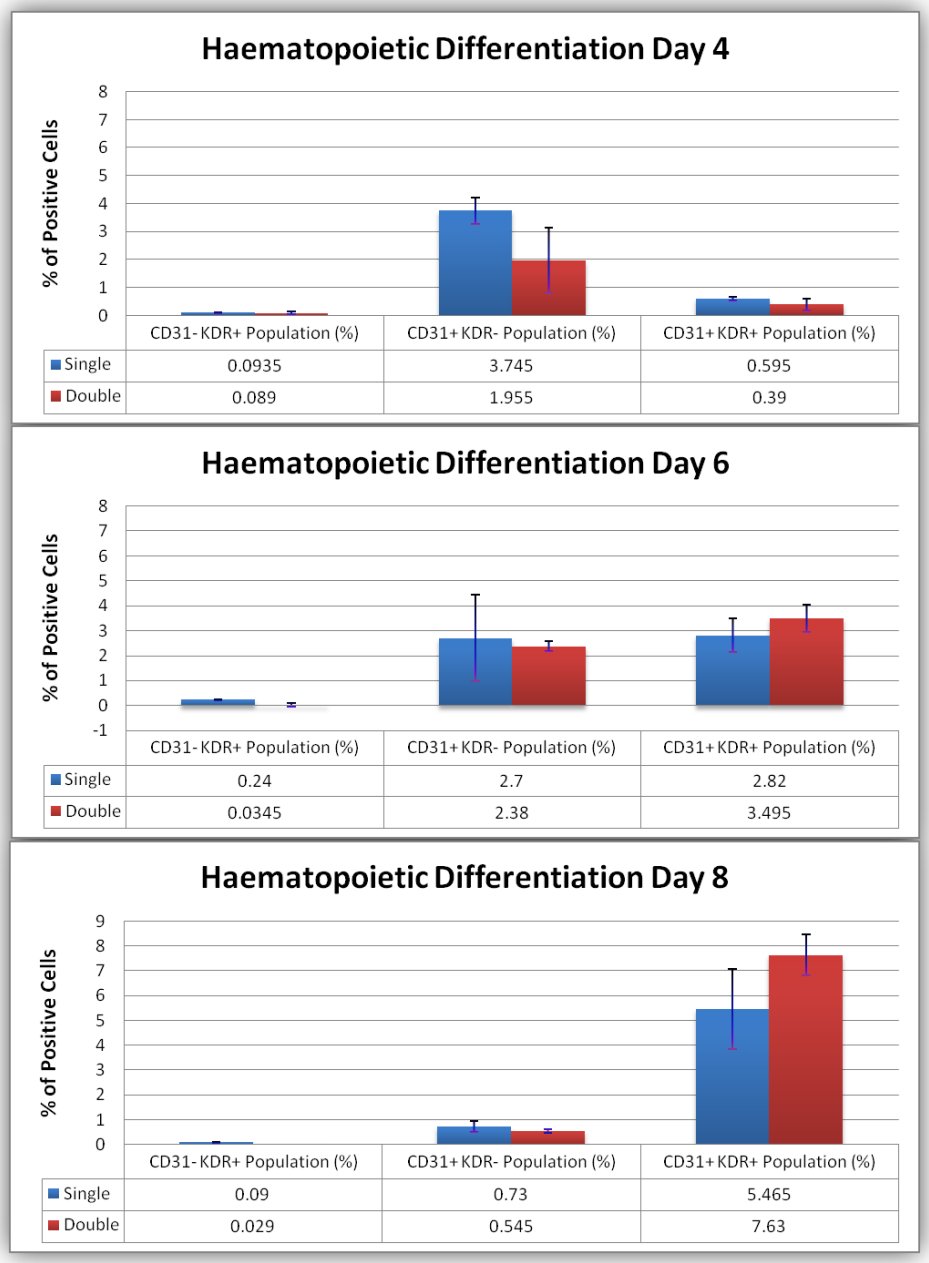


Figure 33: Comparisons between EB densities on the effect of haematopoietic differentiation

The percentages of CD31-KDR+, CD31+KDR-, and CD31+KDR+ populations were measured by flow cytometry from the two EB densities, single and double. Isotype controls were used to represent the negative population, and error bars were calculated using the standard error of the mean with a sample size of 3.

Differences in haematopoietic marker expression at day 6 between SDEBs and DDEBs appear to be less apparent, however it seems that SDEBs generates marginally more CD31+ cells and DDEBs produces more KDR+ cells at this stage. The differences in haematopoeitic expression at day 8 reveal a clearer distinction between the varied haematopoietic potential of the two densities of EBs. Following the same pattern as seen at day 6, this result adds support to hypothesis that SSEBs favour CD31 expression and DDEBs favour KDR expression, and as KDR expression appears to be the limiting factor in CD31+KDR+ cells, DDEBs generate on average 7.63% (CD31+KDR+) cells compared with 5.47% from SBEBs (Figure 33).

The results of EB density were acquired and pooled from cell lines H9 and H1, and therefore it would be interesting to investigate if there is a possible correlation between EB densities with HESC lines. Analysis of the flow cytometry results of day 4 within Figure 34 identifies H9 SDEBs as the most efficient combination in giving rise to CD31+ cells (5.99%) and KDR+ cells (1.11%), which is approximately twice as much as produced by the other three combinations. At day 6, the haematopoietic capacity of H9 SDEBs decreases in comparison with the other combinations, with CD31+ cells decreasing to 3.57% whilst KDR+ cells increase to 3.05% (Figure 34). However, the greatest increase in dual KDR+CD31+ haematopoietic marker expression belongs to DDEBs and in particular H1 DDEBs which at this stage generates 8.25% of CD31+ cells and 4.02% of KDR+ cells. At day 8, the highest number of dual positive CD31 and KDR cells is produced by H9 DDEBs at 8.96%, with the other combinations generating similar numbers.

Taking into account the results of from each differentiation time point, there does not appear to be any significant and consistent correlation between HESC line and EB density during the time course. In general, SDEBs of both cell lines produce the highest number of haematopoietic positive cells up to day 4, then upon VEGF induction from this point, it appears that higher density of EBs do enhance the promotion of haematopoietic marker expression for the remainder of differentiation. The selection of the optimal HESC line for haematopoietic differentiation remains debateable, however these results do appear to marginally favour H9s over H1s, with H9s generating the greatest number of CD31+KDR+ cells on day 4 and day 8. In addition, H9 cells in culture proliferate (under current condition) at a much faster rate than H1s, and therefore the generation of large qualities of HESCs for differentiation in the shortest space of time for microarray analysis favours this HESC line.

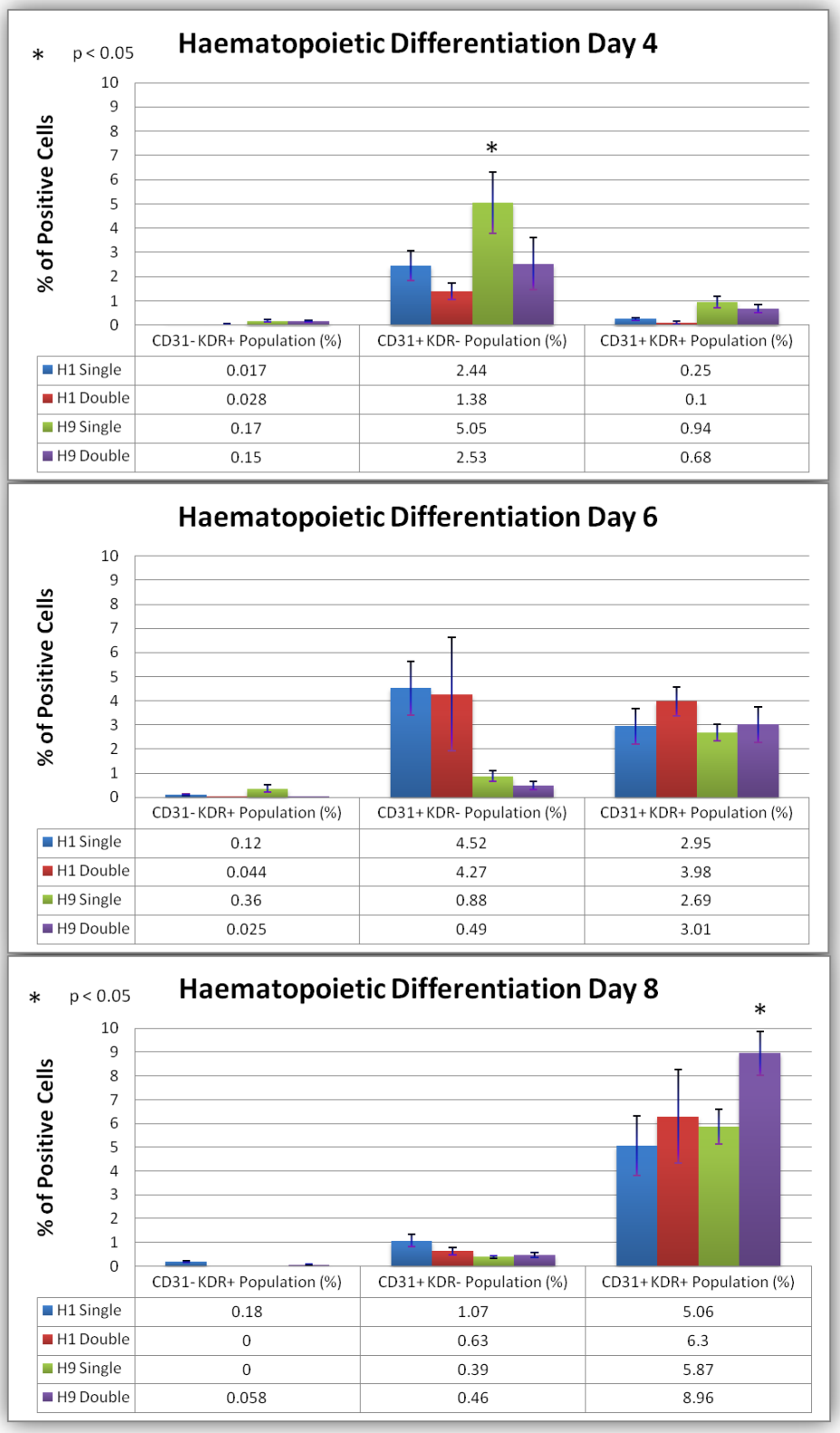


Figure 34: Comparisons between HESC line and EB density on the effect on haematopoietic differentiation

The percentages of CD31-KDR+, CD31+KDR-, and CD31+KDR+ populations were measured by flow cytometry from the H1 and H9 cell lines at two EB densities, single and double. Isotype controls were used to represent the negative population, and error bars were calculated by standard deviation with a sample size of 3.

2.3.3.3 Activin A Supplement

A recent haematopoietic study in mouse by Pearson *et al.* (2008) [108] has suggested that the cytokine Activin A in combination with the three other factors (Bmp4, bFGF, VEGF), that are already included as key components of the optimised media used within this project, is required to drive the selective and efficient differentiation of mouse ES cells (MESCs) to haematopoiesis. It was of interest to investigate if the addition of Activin A to the optimised differentiation media would replicate the enhancement of haematopoietic efficiency as observed in the mouse ESC model.

This experiment investigated the effects of Activin A on haematopoietic differentiation using just H9 EBs, which were cultured with and without Activin A (10ng/ml) supplement in the differentiation media, and the cells were analysed by flow cytometry at day 8. The results in Figure 35 show that the addition of Activin A, instead of enhancing haematopoietic marker expression, produced a slight negative effect on CD31+KDR+ presentation. Due to time and financial constraints, further investigations into the role of Activin A in human haematopoiesis was not pursued. The results from this one differentiation study alone does not provide enough information to assess the full role of Activin A, but what may these results do concluded is that Activin A does not enhance haematopoietic differentiation at this one concentration in combination with this optimised media, at this differentiation time point, and under these set experimental conditions.

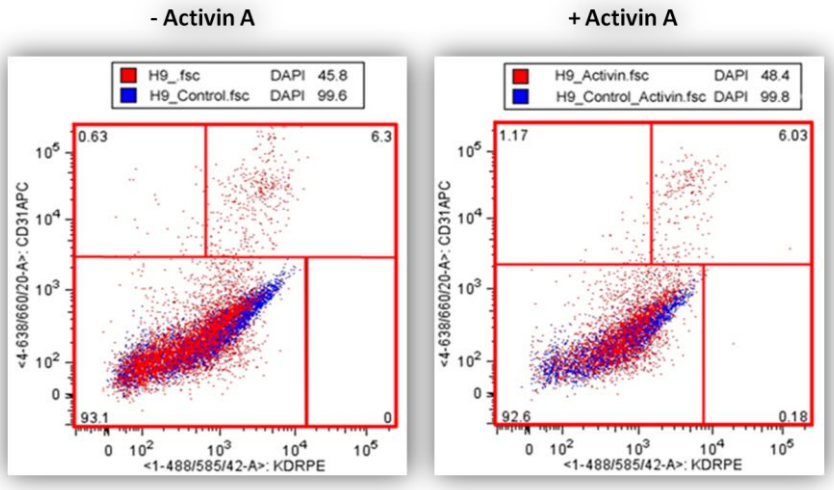


Figure 35: Effect of Activin A on haematopoietic differentiation

The percentages of the CD31+KDR+ haemangioblast enriched population was measured by flow cytometry from the day 8 differentiated EBs with and without Activin A supplement. Isotype controls were used to represent the negative population, with a sample size of 1.

2.3.4 Isolation of the Haemangioblast

2.3.4.1 Practicalities of the Isolation of Haemangioblast-enriched cells

Haematopoiesis studies have identified a number of cell surface markers including KDR, CD31, CD34, CD117, and CD114, which in combination may be used in the identification of the earliest haematopoietic and endothelial precursor, the haemangioblast [143, 180, 182-185, 242-246]. As previously discussed, the principal aim is the isolation of the haemangioblast at various stages of haematopoietic differentiation from HESCs for microarray analysis, in order to identify novel transcripts that may play an important role in haematopoiesis.

As a rough guide for Affymetrix microarray genechip applications, the minimum number of cells required is in the approximate region of 100,000 cells to allow for the isolation of a minimum of 1 µg of high quality of total RNA [255]. As previously commented, the limiting factor with this experiment is the isolation of the haemangioblast population identified by cell surface markers (KDR, CD31, CD34, CD117, and CD114). However, in practice, in this EB haematopoietic differentiation from HESCs using this optimised cytokine media and method, and this pure haemangioblast population represent just 0.001% of the total EBs (Figure 36). Sorting and isolating this pure haemangioblast is not impossible, as flow cytometry results in Figure 36 has confirmed the existence of 2 cells from the 200,000 analysed which carry the cell surface antigens for all five haemangioblast markers. The task of isolating enough of these cells for microarray is again not impossible but rather very difficult, with calculations suggesting that a minimum number of 30,000 confluent 6-well plates of HESCs will be required. This figure does not take into account cell loss, which will occur throughout differentiation, for example during EB formation, gravity-mediated media changes during differentiation, EB dissociation and subsequent filtering, and wash steps at various point during staining. These cell losses resulting from differentiation, in combination with non-viable stained cells identified during the flow cytometry stage, may mean that as little as 30% of the cells from the start may be analysed. As only live stained cells will be analysed and sorted for, which means that the original number of 30,000 plates required for the isolation haemangioblasts realistically becomes closer to 100,000 plates. Unfortunately, within this short time frame and limit resources available it was not possible within this project to culture this number of HESC plates, and the only alternative was to sacrifice the purity of the haemangioblast cell by using fewer markers. The reduction of haemangioblasts markers clearly results in higher percentages of cells (Figure 36), however the elimination of one marker still requires thousands of HESC

plates, and the elimination of two markers also in general requires too many plates into the hundreds. Therefore, the elimination of three markers appears to be the only alternative and studies have shown that the two most important early markers of haemangioblast are CD31 and KDR [75, 143, 165, 246].

Population	Events	% of Cells	Number of cells required to isolation 100,000 cells	Number of plates of EBs	Number of plates of HESCs
ALL EVENTS	200,000	100.000	100,000.0	0.1	0.3
CD34	16,280	8.140	1,228,501.2	1.2	3.7
CD31	10,440	5.220	1,915,708.8	1.9	5.7
CD117	2,740	1.370	7,299,270.1	7.3	21.9
KDR	4,180	2.090	4,784,689.0	4.8	14.4
CD144	1,900	0.950	10,526,315.8	10.5	31.6
CD31 AND CD34	7,220	3.610	2,770,083.1	2.8	8.3
CD31 AND CD117	1,880	0.940	10,638,297.9	10.6	31.9
CD31 AND KDR	4,180	2.090	4,784,689.0	4.8	14.4
CD31 AND CD144	1,360	0.680	14,705,882.4	14.7	44.1
CD34 AND CD117	1,620	0.810	12,345,679.0	12.3	37.0
CD34 AND KDR	3,920	1.960	5,102,040.8	5.1	15.3
CD34 AND CD144	1,160	0.580	17,241,379.3	17.2	51.7
CD117 AND KDR	980	0.490	20,408,163.3	20.4	61.2
CD117 AND CD144	120	0.060	166,666,666.7	166.7	500.0
KDR AND CD144	560	0.280	35,714,285.7	35.7	107.1
CD31 AND CD34 AND CD117	260	0.130	76,923,076.9	76.9	230.8
CD31 AND CD34 AND KDR	1,396	0.698	14,326,647.6	14.3	43.0
CD31 AND CD34 AND CD144	580	0.290	34,482,758.6	34.5	103.4
CD31 AND CD117 AND KDR	20	0.010	1,000,000,000.0	1,000.0	3,000.0
CD31 AND CD117 AND CD144	20	0.010	1,000,000,000.0	1,000.0	3,000.0
CD31 AND KDR AND CD144	120	0.060	166,666,666.7	166.7	500.0
CD34 AND CD117 AND KDR	220	0.110	90,909,090.9	90.9	272.7
CD34 AND CD117 AND CD144	120	0.060	166,666,666.7	166.7	500.0
CD34 AND KDR AND CD144	160	0.080	125,000,000.0	125.0	375.0
CD117 AND KDR AND CD144	20	0.010	1,000,000,000.0	1,000.0	3,000.0
CD31 AND CD34 AND CD117 AND KDR	4	0.002	5,000,000,000.0	5,000.0	15,000.0
CD31 AND CD34 AND CD117 AND CD144	18	0.009	1,111,111,111.1	1,111.1	3,333.3
CD31 AND CD34 AND KDR AND CD144	13	0.007	1,538,461,538.5	1,538.5	4,615.4
CD31 AND CD117 AND KDR AND CD144	10	0.005	2,000,000,000.0	2,000.0	6,000.0
CD34 AND CD117 AND KDR AND CD144	8	0.004	2,500,000,000.0	2,500.0	7,500.0
CD31 and CD34 AND CD117 AND KDR AND CD144	**2**	**0.001**	**10,000,000,000.0**	**10,000.0**	**30,000.0**

Figure 36: Flow cytometry results from day 4 EBs using optimised differentiation media and conditions

Tabulated results from flow cytometry results showing all combination of haemangioblast markers together with recorded percentage of positive cells. Addition information including the number of cells of EBs and HESCs that are required for the isolation of each of these populations was calculated.

As previously described in Chapter 1, the roles of KDR and CD31 during early haematopoiesis are both diverse and critical, and although the differentiated cells identified and isolated by these markers may not be representative of pure haemangioblasts, they do

however represent a purer population (compared to that of the EB) that are enriched for haemangioblasts (Figure 37). It is therefore with these enriched haemangioblast populations (at day 4, day 6, and day 8) which may assist in providing the answers to the key mechanisms and pathways that regulate early haematopoiesis. In addition, the identification of novel transcripts may aid in enhancing the efficiency of haematopoietic differentiation, with the goal of isolating pure haemangioblast cells from fewer HESC plates.

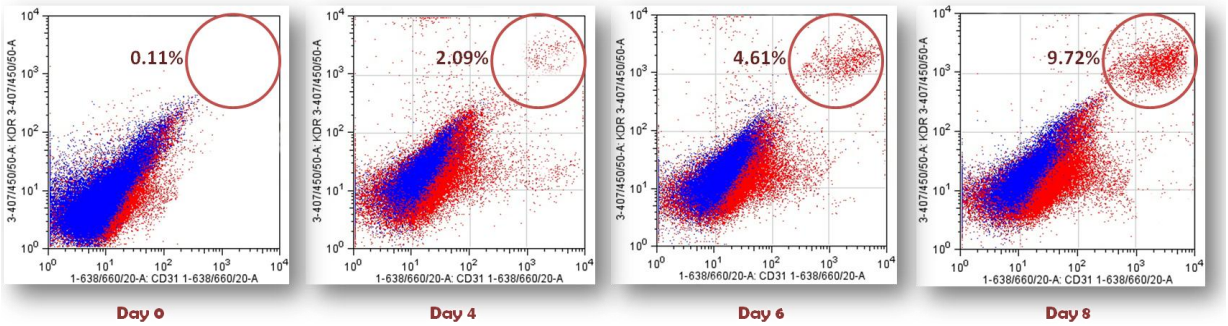


Figure 37: Flow cytometry results of KDR and CD31 positive cells from optimised differentiation media and conditions

Utilising the optimised media complete with cytokines added at the appropriate stage, H9 EBs were cultured at single density from day 0 to day 4 when the density was doubled to day 8.

2.3.4.2 Demonstration of Enhanced Colony-forming Potential of Sorted CD31+KDR+ cells

As observed from previous results, the generation efficiency of colony-forming units (CFUs) from unsorted cells of the haematopoietic differentiated EB is extremely low, with a maximum of 0.34% of the methylcellulose plated cells producing CFUs (Figure 38). In contrast, the colony-forming efficiency of KDR+CD31+ cells is significantly greater, producing as much as 13 times more CFUs (at day 8) when compared with unsorted cells. The results shown in Figure 38, shows a maximum number of 984 CFUs produced at day 8 by KDR+CD31+ sorted cells, which accounts for 3.28% of the original 30,000 sorted cells that were plated. However, in theory this percentage would actually be much higher than which was observed in the colony-forming results (Figure 38) as post-sorting cell survival is never 100%. This was classically demonstrated during the first attempt to culture KDR+CD31+ sorted cells, which resulted in almost 100% cell death, and only the addition of Rho-associated kinase (ROCK) inhibitor (Y-27632; protects cells from apoptosis [226]) with subsequent attempts did survival improve.

It was expected that the overall number of CFUs would increase with the distribution of subtypes remaining fairly constant, but upon closer inspection however, comparisons between the distributions of colony-forming unit subtypes at each day of differentiation appear to differ by a quite a large degree, with noticeable increases of CFU-E and BFU-E types within the KDR+CD31+ sorted population. These differences may simply reflect the spontaneous and uncontrolled differentiation of haematopoietic progenitors towards terminal cell types of CFUs as previous colony forming assays of unsorted EBs has shown great variability in CFU distribution. The CFU-GEMM population which was the major subtype in the results of unsorted EBs, is no longer replicated in the sorted population with CFU-GM, CFU-M, and CFU-G all sharing a larger slice of the pie. This suggests that the sorted population contains cells which are already relatively mature (compared with those in the unsorted EB) and this may explain why the CFU-GEMM precursor (to CFU-GM, CFU-G and CFU-M) is considerably smaller. In addition, the loss of haematopoietic factors that may have been present within the EB and possibly play a role in the survival and emergence of CFU-GEMM population may be a reason for this decrease in the sorted population.

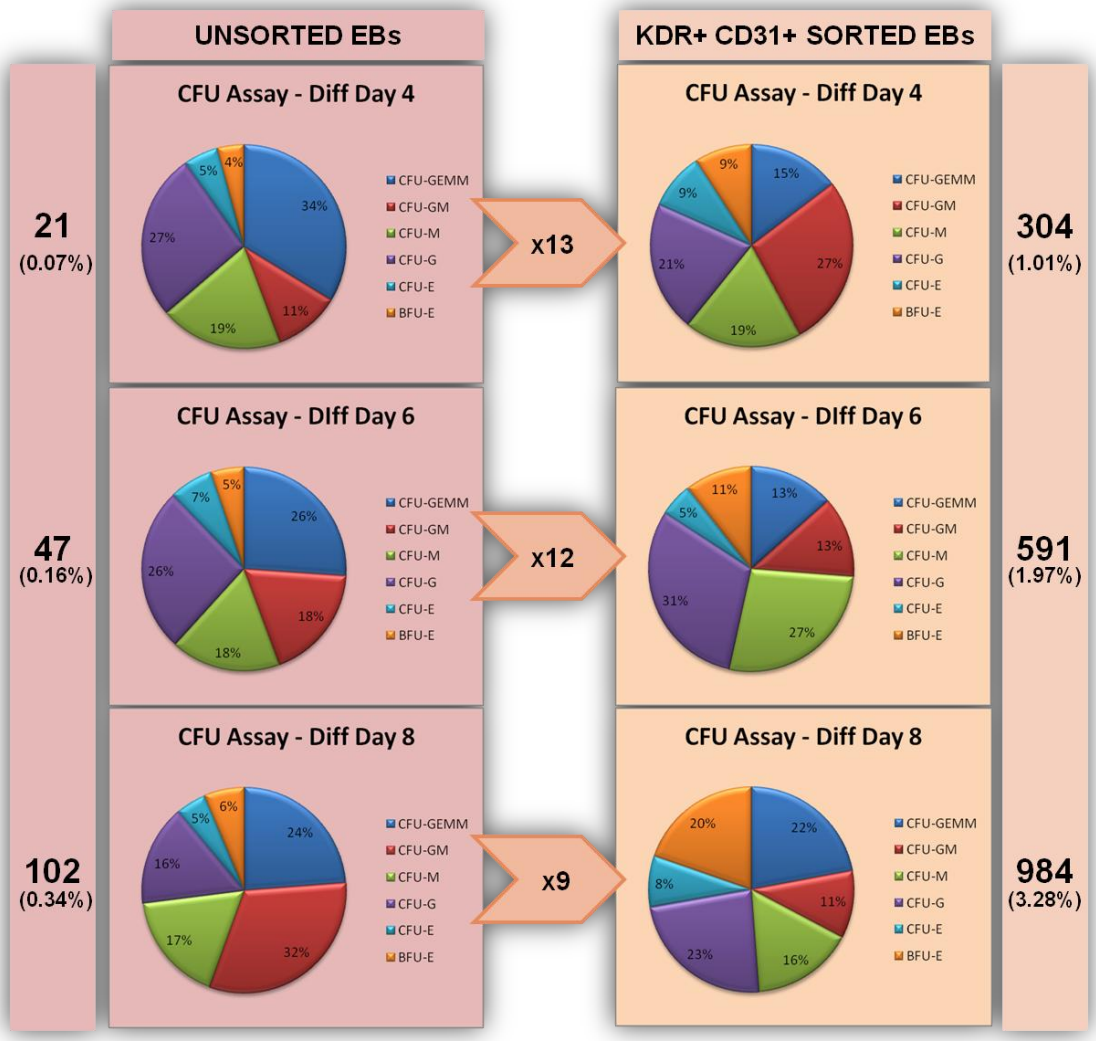


Figure 38: Enhanced colony-forming potential from sorted KDR+CD31+ cells

The efficiency of colony-forming potential increased 9-13 times between unsorted EBs and KDR+CD31+ sorted EBs, showing clear evidence of haematopoietic progenitor enrichment within the sort population. A larger sample of colony-forming units provides a more accurate distribution of the subtypes with the clear expansion of BFU-E and CFU-E types which form the basis of erythropoiesis.

2.4 Discussion

2.4.1 The Optimised Differentiation media

The key component in the differentiation of HESCs into haematopoietic progenitors is the media and as shown by differentiation results in Figure 25, the use of a serum-free cytokine-specific media dramatically increases the efficiency of haematopoietic marker presentation. The principal constituents within the optimised media are clearly the cytokines themselves, but equally important was the timing at which these were added to the differentiation experiment. As described in Chapter 1, the bone morphogenetic protein 4 (BMP4) plays a vital role in mesoderm formation [107, 256, 257], which within this optimised media it was hypothesised that the majority of HESCs may be persuaded to differentiate towards the mesoderm lineage. The addition of basic fibroblast growth factor (bFGF) supplement into the media from day 2 of differentiation appears to further enhance mesoderm formation, which is in direct contrast to its role in the maintenance of pluripotency in HESC culture. However, in support of this contrasting role of bFGF, a study by Yamada *et al.* (1994) [258] had shown that addition of bFGF during EB differentiation resulted in the up-regulation of the mesodermal marker, *Brachyury*. At day 4, the addition of vascular endothelial growth factor (VEGF) to the media promotes the differentiation of mesodermal precursors to the haemangioblast fate and committed haematopoietic progenitors [108, 259]. These factors (and more) may also be present in the FBS-mediated media as demonstrated by its ability to generate haematopoietic cells albeit at a low efficiency compared with the optimised media. Although the optimised media is able to produce higher percentages of haematopoietic cells (compared to the FBS-mediated media), it is quite clearly still not perfect and requires further optimisations, most likely in the form of addition cytokines at specific times of differentiation. It is hoped that through the analysis of the results of the microarray experiment, comparisons between KDR+CD31+ populations at different time points of differentiation may reveal novel roles for cytokines within this media at various stages of differentiation.

2.4.2 Embryoid Body Size and Density

Custom-made uniformed-sized EBs were produced by Ng *et al.* (2008) [253] and it was demonstrated that small EBs (<1000 cells) do not differentiate well and found to take up brown pigmentation, where as large EBs (>4000 cells) fail to differentiate towards the mesoderm lineage in the presence of BMP4 and are prone to differentiate towards ectoderm fates. The optimal sized EB for efficient mesoderm induction should contain approximately 3000 cells, as qPCR results have shown that at day 2 of differentiation, the gene expression level of the mesoderm marker *Brachyury* increases by almost 100 fold compared to the level at the start of differentiation [253]. Unfortunately EB size was not controlled within this experiment for a number of reasons, with the most important being the inability to produce EBs from dissociated single cell H1 and H9 HESCs. Numerous attempts to follow the protocol of EB formation as specified by Ng *et al.* (2008) [253] have resulted in the EBs completely falling apart over the course of differentiation which is due to most likely due to the method of sub-culture of the HESCs as colonies rather than single cells. The ability of these custom-sized EBs to generate cells with haematopoietic potential in the described optimised media has not been thoroughly assessed in this project due largely to time constraints. The production of custom sized EBs requires the complete dissociation of HESC colonies which must be counted before being pipetted into rounded 96-well plates and spun down, and as one standard plate of EBs requires twelve 96-well plates of EBs, it quickly becomes very time consuming especially if a large number of EBs are required. However, the control of EB size may further enhance haematopoietic differentiation and remains a key area for future investigation in the future.

2.4.3 Role of Activin A in Haematopoietic Development

Nostro *et al.* (2008) [203] have shown in mouse studies that Activin and Nodal are necessary for the formation of the primitive streak and for later functions in the induction of KDR+ mesoderm. Although the result from this project appears to contradict what was observed in MESCs, it does not necessarily prove that Activin A is nonessential for human haematopoiesis. This unexpected result may be due to a variety of reasons, with the first being that the mouse system may use a slightly different pathway to human, and perhaps Activin A may not be as crucial for human mesoderm development. As commented on earlier, the timing of the addition of Activin A is very important, together with the correct quantity (and

concentration). Unfortunately within this project, investigating the role of Activin A in human haematopoiesis will involve too much time and resources, and the work involved may easily amount to a PhD within itself. However, the gains of this knowledge may actually be worthwhile and may indeed under the correct conditions enhance differentiation, and analysis of the microarray results may reveal more on the role of Activin A in haematopoietic development.

2.4.4 Enriched Haemangioblast marked by KDR+CD31 cells

The decision to proceed with the utilisation of just two early haematopoietic markers (KDR and CD31) for the isolation of the haemangioblast population was purely dependent on the efficiency of differentiation. As previously discussed in order to culture enough HESCs to sort the minimum required number of haemangioblast cells that express the haematopoietic markers (KDR, CD31, CD34, CD144, and CD117), one would require in the region of one hundred thousand 6-well plates to begin differentiation just for day 4 alone. Although KDR+CD31+ cells do not solely contain the haemangioblast population, studies have shown that this haematopoietic precursor population must however express KDR and CD31 [181, 182, 260], and therefore this KDR+CD31+ sorted population may contain 'enriched haemangioblasts'. However, the exact degree of enrichment of the haemangioblast within the sorted population is unknown, but a calculation based on the differences of percentage of haemangioblast markers may provide an estimation of the 'pureness' of the enriched population. When calculating the percentage of haemangioblasts at day 4 using the data given in Figure 36, it was assumed that KDR, CD31, CD34, CD117, and CD144 identifies the pure haemangioblast population (100%), then this haemangioblast population within the sorted KDR+ CD31+ population would represent just 0.02% ((2/4180)*100), and the same population within the unsorted EB would represent a minuscule 0.001% ((2/200000)*100). Although the percentage of haemangioblasts within the sorted cells appear very low, but when compared with the same percentage within unsorted cells however, it confirms that the sorted population is 20 fold more enriched with haemangioblast cells than the unsorted population.

The results of the colony-forming assays appears to question the validity of the haemangioblast markers, showing an overwhelming 1.01% (day 4), 1.97% (day 6), and 3.28% (day 8) of colony-forming units, which are certainly several magnitudes greater than the 0.02% of haemangioblast cells predicted to be present in the sorted population. This may suggest

that cells with haematopoietic potential do not necessarily require the expression of all five haematopoietic markers or perhaps that during the course of the differentiation experiment, a portion of haemangioblasts may have matured into primitive haematopoietic cells which no longer express all of the markers. In summary, in seems that the sorted population may contain a percentage of haemangioblast-derived cells (or cells with haematopoietic potential) as well as a small percentage of pure haemangioblast cells which is 20 fold more enriched than in whole EBs. The use of these sorted cells for microarray analysis in the assessment of gene expression change through the course of differentiation may uncover novel genes of haematopoietic importance. The theory behind this assumption relies on the principle that the population assessed by microarray is enriched for haemangioblast (and haematopoietic cells), therefore in turn it is predicted that that a greater number of genes, especially those with significant changes in the level of expression during differentiation may be associated with haematopoiesis.

CHAPTER 3 – MICROARRAY

3.1 Introduction

A microarray enables the simultaneous assessment of gene expression across an entire genome and thus allowing cellular behaviour to be studied not just on a gene-specific level, but on a genome-wide level. This technology is not new and has been described in literature as early as the 1980s in a study of the effects of interferon on human fibrosarcoma cells, and it was discovered that interferons acted through key genes which contribute to cell proliferation arrest [261]. This same technology may assist in the discovery of novel genes that are important in early haematopoiesis by allowing the entire genome to be assessed at various different haematopoietic conditions.

This microarray experiment consists of two biological replicates for each time point (Day 0, Day 0, Day 4, Day4, Day 6, Day 6, Day 8, and Day 8) giving a total of 8 Affymetrix GeneChips, however with unlimited funds and resources, this would increase to a minimum of 3 biological repeats with 3 technical repeats for each biological duplicate, giving a grand total of 36 GeneChips. This increase in the number of repeats will provide added confidence to the results and the p-values will also be likely be more meaningful as a consequence.

There have already been a number of haematopoietic differentiation studies (whole unsorted embryonic bodies) which have utilised microarrays in screening for haematopoietic-relevant genes [262-264]. However, there has not been to this date a microarray study comprising of this degree of purified population of H9 differentiated haematopoietic progenitors (dual KDR+ CD31+ cells). It must be iterated that the original aim was to use a purer haemangioblast population identified by five key cell surface markers (CD31, KDR, CD34, CD117, and CD144), but for practical reasons as discuss in the previous chapter, KDR and CD31 were the only markers used in the selection. Kinase insert domain receptor (KDR) and platelet endothelial cell adhesion molecule (PECAM1 / CD31) are both important early haematopoietic marker genes and their essential roles have been well established in literature [157, 182, 265-268]. The use of just two markers represent a less pure haemangioblast population compared with all five markers, which means that there may be more background non-haematopoietic specific gene expression which would could hamper the discovery of novel haematopoietic important transcripts. This will therefore require higher levels of filtering and processing during the analytical stage in order to ‘separate the wheat from the chaff’, which may involve

utilising a combination of different strategies including p-value and fold change thresholds, similarity matrixes, and profile clustering. These approaches will facilitate the discovery of genes that are important in early haematopoietic development, which may lead on to further gene-specific studies in characterising and establishing their functional role in early haematopoiesis.

3.2 Methods

3.2.1 Cells to Microarray

3.2.1.1 RNA Extraction from sorted CD31+ KDR+ (Day 4, 6, 8) and unsorted (Day 0) cells

Dual kinase insert domain receptor (KDR) positive and platelet / endothelial cell adhesion molecule (PECAM-1 / CD31) positive cells from day 4, day 6, and day 8 differentiated H9 embryoid bodies were sorted by fluorescence-activated cell sorting flow cytometry directly into 750 µl of TRIzol LS reagent (Invitrogen). The TRIzol LS reagent immediately lyses the sorted cells, dissolves cell components, and crucially maintains and protects the integrity of the RNA [39]. A maximum volume of 250 µl of sorted cells were added to each 1.5 ml microfuge tube containing 750 µl of TRIzol LS reagent, and thus giving a maximum total combined volume of 1 ml. Day 0 undifferentiated and unsorted H9 cells were spun down and resuspended in 250 µl of phosphate buffered saline (PBS) and combined with 750 µl of TRIzol LS reagent. The samples were homogenised by incubation at room temperature for 5 minutes to facilitate the complete dissociation of nucleoprotein complexes. 10 µg of glycogen (Invitrogen) were added to the homogenised sample as a carrier before 200 µl of chloroform (VWR) were added for phase separation. The microfuge tube was shaken vigorously by hand for 15 seconds and incubated at room temperature for 15 minutes, before centrifuging at 12,000 x g for 15 minutes at 4°C. The upper clear aqueous phase of the sample was transferred to a clean 1.5 ml microfuge tube, and the RNA was precipitated out by the addition of 500 µl of isopropyl alcohol (VWR). The sample was incubated at room temperature for 10 minutes and centrifuged at 12,000 g for 10 minutes at 4°C. The supernatant was removed and the RNA pellet was washed with 1 ml of 75% ethanol which was mixed by vortexing, and followed by centrifugation at 7,500 x g for 5 minutes at 4°C. The supernatant was discarded and the tube was carefully inverted and placed on a paper towel for 5 minutes, and then the RNA pellet was air dried at room temperature for 5 minutes. The RNA pellet was dissolved in 10 µl of RNase-free water and incubated at 60°C for 10 minutes to increase solubility of the pellet. Once dissolved the RNA was checked for its quality and quantity using the Thermo Scientific Nanodrop and Agilent Bioanalyzer machines.

3.2.1.2 RNA Quantity Control - Nanodrop

Prior to amplification of RNA, the concentration and quantity were assessed using a Thermo Scientific Nanodrop 1000 spectrophotometer as previously described in Chapter 2.

3.2.1.3 RNA Quality Control - Bioanalyser

The assessment of RNA integrity and quality is the most important first step when attempting to generate meaningful gene expression data for a reliable microarray experiment. The method of RNA quality control described below used an expensive Agilent 2100 Bioanalyser (Agilent Technologies) in the determination of whether a particular RNA sample was of suitable quality to produce a reliable microarray. The Agilent 2100 Bioanalyser assesses the RNA quality and quantity using electrophoretic separation on microfabricated chips. The standard method to assess RNA integrity is to use denaturing agarose gels but the major drawback is the quantity of RNA required for visualisation and this generally is at least 200 ng [269] compared to the Bioanalyser which require just 20 ng. Within this project, total RNA was isolated from limited numbers of sorted cells and it was impractical to spare such a large quantity for quality control, and therefore the Bioanalyser approach was favoured.

Only RNA samples that contained RNA concentration of 20 ng / µl or higher were selected and quality assessed using a RNA Nanochip within an Agilent 2100 Bioanalyser. The RNA samples were kept on ice while the reagents and RNA size markers were allowed to equilibrate at room temperature for 30 minutes. The gel was prepared by pipetting 550 µl of RNA 6000 Nano gel matrix (Agilent) into a spin filter and centrifuged at 1500 x g for 10 minutes. The filter was removed and 65 µl of this gel were allocated into an RNase-free microfuge tube for each sample. RNA 6000 Nano dye concentrate (Agilent) was vortexed for 10 seconds and 1 µl was added to the tube containing the filtered RNA 6000 Nano gel matrix. The microtube containing the gel matrix and dye was vortexed until the dye was no longer visible, and centrifuged at 13,000 x g for 10 minutes at 20°C. A new RNA Nano chip was placed in the Chip Priming Station (Agilent) and the connected syringe was adjusted to the 1 ml position. 9 µl of the gel-dye mix were pipetted into the well marked with a white on a black circle, then the priming station was closed and the plunger was pressed for 30 seconds. 9 µl of the gel-dye mix were pipetted into remaining two wells marked with black G (illustrated above the wells). 5 µl of RNA 6000 Nano Marker were pipetted into all remaining wells including the ladder well. 1 µl of RNA ladder (a set of six RNA transcripts with lengths of 0.2, 0.5, 1.0, 2.0, 4.0 and 6.0 kb) was pipetted into the well marked by a picture of small ladder (marked above the

well), and 1 µl of each sample was pipetted into the remaining wells. The chip was centrifuged at 2,400 x g for 1 minute and then placed in the Bioanalyser machine where the RNA samples were subjected to electrophoretic separation and detected by laser induced fluorescence.

Results were shown in the form of electropherogram which provided a detailed visual assessment of the quality of the RNA, displaying a marker peak followed by two ribosomal peaks (28S and 18S) for each sample, and from which a ribosomal RNA ratio (28S : 18S) and a RNA Integrity Number (RIN) was calculated. Studies have shown that the use the ribosomal RNA ratio only does not offer a reliable standard for the assessment of the integrity of RNA purely due to the fact that ribosomal ratios will vary between different species, tissue types and even the method of RNA extraction [270]. The RIN was calculated not just from the ribosomal RNA ratio but from the entire electrophoretic trace of the RNA sample using adaptive learning algorithm (Bayesian learning technique) derived using data from a large collection of electrophoretic RNA measurements (approximately 1300 total RNA samples from a variety of tissue and species types all with varying levels of RNA integrity), and its output is based on a simple numbering system from 1 (most degraded) to 10 (most intact) based on the electrophoretic profile of the sample [270]. Only samples with a ribosomal RNA ratio between 1.8 and 2.0 and a RIN of 8 and greater were considered to be of high enough quality and used for microarray in this project.

3.2.2 Microarray - Two-Cycle Target Labelling

3.2.2.1 First-Cycle First-Strand cDNA Synthesis

The RNA isolated from the sorted and unsorted cells underwent two complete cycles of cDNA synthesis instead of just one cycle due to the low quantity of total RNA that was available at the start. The minimum quantity of RNA required for a One-Cycle cDNA synthesis protocol for Affymetrix GeneChip target labelling is 1 µg [255], and in this project the average quantity of total RNA isolated from sorted cells was just 300 ng. An Affymetrix Two-Cycle cDNA synthesis protocol permits successful double strand cDNA synthesis from as little as 10 ng (to a maximum of 100ng) of starting total RNA [271].

First-cycle, first-strand cDNA synthesis used a total quantity of 100 ng of RNA in a 3 µl volume combined with 2 µl of T7-Oligo(dT) Primer/Poly-A Control Mix (
Table 22) and these were mixed thoroughly within a 0.2 ml PCR tube. Each sample (5 µl) was incubated at 70°C in a thermal cycler for 6 minutes, then cooled immediately to 4°C for 2 minutes and collected by centrifugation at 5,000 x g for 1 minute. 5 µl of First-Cycle, First-Strand Master Mix (
Table 23) were added to the sample and mixed thoroughly. The sample was briefly centrifuged at 5,000 x g for 1 minute and incubated at 42°C for 1 hour, then at 70°C for 10 minutes, and finally cooled at 4°C for 2 minutes before again briefly centrifuged at 5,000 x g for 1 minute.

T7-OLIGO(DT) PRIMER/POLY-A CONTROLS MIX

Components	Quantity	Supplier
T7-Oligo(**dT) Primer (50 µM/µl**),	2.0 µl	Affymetrix
Dil Poly-A RNA Controls	2.0 µl	Affymetrix
RNase-free water	16.0 µl	Affymetrix

Table 22: Composition of T7-Oligo(dT) Primer/Poly-A Controls Mix

FIRST-CYCLE, FIRST-STRAND MASTER MIX (2x)

Components	Quantity	Supplier
5x 1st Strand Reaction Mix	4.0 µl	Affymetrix
0.1 M DTT	2.0 µl	Affymetrix
RNase Inhibitor	1.0 µl	Affymetrix
10 mM dNTP	1.0 µl	Affymetrix
SuperScript II	2.0 µl	Affymetrix

Table 23: Composition of First-Cycle, First-Strand Master Mix (2x)

3.2.2.2 First-Cycle Second-Strand cDNA Synthesis

The cDNA synthesised from the first-cycle, first-strand cDNA reaction was used as the template for a first-cycle, second-strand cDNA reaction and combined with 10 µl of First-Cycle, Second-Strand Master Mix (
Table 24). The sample was centrifuged and incubated at 16°C for 2 hours, then at 75°C for 10 minutes, before cooled to 4°C for 2 minutes and centrifuged again.

FIRST-CYCLE, SECOND-STRAND MASTER MIX (2x)

Components	Quantity	Supplier
RNase-Free Water	9.6 µl	Affymetrix
17.5 mM MgCl2	8.0 µl	Affymetrix
10 mM dNTP	0.8 µl	Affymetrix
E.coli DNA Polymerase	1.2 µl	Affymetrix
RNase H	0.4 µl	Affymetrix

Table 24: Composition of First-Cycle, Second-Strand Master Mix (2x)

3.2.2.3 cDNA Purification

Prior to *in vitro* transcription, the cDNA product from the first cycle, second-strand reaction was purified and cleaned of components which otherwise would inhibit downstream reactions. 600 µl of cDNA binding buffer (Affymetrix) were added to the total double stranded cDNA produced from the first-cycle, second-strand reaction and mixed by vortex for 3 seconds. 500 µl of the combined sample were transferred to a cDNA clean up spin column (Affymetrix) and centrifuged at 10,000 x g for 1 minute. The flow through was discarded and the remaining

sample was loaded onto the spin column and centrifuged at 10,000 x g for 1 minute. The spin column was removed and transferred to a new collection tube, where 750 µl of cDNA wash buffer were pipetted into the spin column, and the tube was spun down at 10,000 x g for 1 minute. The flow through was discarded and with the lid of the spin column left open, centrifuged at maximum speed for a further 4 minutes. The spin column was transferred to a clean 1.5 ml collection tube and 14 µl of cDNA elution buffer (Affymetrix) were carefully pipetted directly onto the spin column membrane, and incubated at 20°C for 1 minute and centrifuged at 12,000 x g for 1 minute.

3.2.2.4 In Vitro Transcription (IVT)

The cDNA purified from the first-cycle, second-strand reaction was used in an *in vitro* transcription (IVT) reaction in the generation of cRNA. The second-strand cDNA reaction sample was combined with 30 µl of First-Cycle, IVT Master Mix (
Table 25). This mixture was centrifuged at 5,000 x g for 30 seconds and incubated at 37°C for 16 hours.

FIRST-CYCLE, IVT MASTER MIX (2x)

Components	Quantity	Supplier
10X Reaction Buffer	10.0 µl	Ambion
ATP Solution	10.0 µl	Ambion
CTP Solution	10.0 µl	Ambion
UTP Solution	10.0 µl	Ambion
GTP Solution	10.0 µl	Ambion
Enzyme Mix	10.0 µl	Ambion

Table 25: Composition of First-Cycle, IVT Master Mix (2x)

3.2.2.5 cRNA Purification

Prior to second-cycle, first-strand cDNA synthesis the cRNA product from the *in vitro* transcription reaction was purified free of nucleases and other contaminates which may affect the efficiency of the downstream cDNA synthesis reactions. 50 µl of RNase-free water (Affymetrix) were added to the IVT reaction and mixed by vortex for 3 seconds. 350 µl of IVT cRNA binding buffer (Affymetrix) were added to the mixture and vortexed for a further 3 seconds. 250 µl of 100% ethanol (VWR) were added to the lysate and mixed thoroughly by pipetting and then transferred into an IVT cRNA cleanup spin column (Affymetrix) which was centrifuged at 10,000 x g for 15 seconds. The spin column was removed and transferred to a new collection tube, where 500 µl of cRNA wash buffer were pipetted onto the spin column, and the tube was spun down at 10,000 x g for 15 seconds. The flow through in the collection tube was discarded and 500 µl of 80% ethanol were added to the spin column, which was then centrifuged at 10,000 x g for 15 seconds. The flow through was again discarded and with the lid of the spin column left open it was centrifuged at maximum speed for 5 minutes. The flow through was discarded again and centrifuged at maximum speed for a further 5 minutes. The spin column was transferred to a clean 1.5 ml collection tube and 13 µl of RNase-free water were carefully pipetted directly onto the spin column membrane, and centrifuged at 12,000 x g for 1 minute. The concentration of purified cRNA was measured with a Thermo Scientific Nanodrop 1000 spectrophotometer as described previously in 'RNA Quality Control – Nanodrop' protocol.

3.2.2.6 Second-Cycle First-Strand cDNA Synthesis

Second cycle first strand cDNA synthesis required 600 ng of purified cRNA in a 9 µl volume which were incubated at 70°C in a thermal cycler for 10 minutes, then cooled immediately at 4°C for 2 minutes and collected by centrifugation at 5,000 x g for 10 seconds. 9 µl of Second-Cycle, First-Strand Master Mix (

Table 26) were added to the sample and mixed thoroughly. The sample was briefly centrifuged at 5,000 x g for 10 seconds and incubated at 42°C for 1 hour before centrifuged again at 5,000 x g for 10 seconds. 1 µl of RNase H (Affymetrix) was added to the sample, which was incubated at 37°C for 20 minutes, then at 95°C for 5 minutes, before cooled at 4°C for 2 minutes and centrifuged at 5,000 x g for 10 seconds.

SECOND-CYCLE, FIRST-STRAND MASTER MIX (3x)

Components	Quantity	Supplier
5x 1st Strand Reaction Mix	12.0 µl	Affymetrix
0.1 M DTT	6.0 µl	Affymetrix
RNase Inhibitor	3.0 µl	Affymetrix
10 mM dNTP	3.0 µl	Affymetrix
SuperScript II	3.0 µl	Affymetrix

Table 26: Composition of Second-Cycle, First-Strand Master Mix (3x)

3.2.2.7 Second Cycle Second Strand cDNA Synthesis

4 µl of 5 uM T7-Oligo primer were added to the second-cycle, first-strand cDNA reaction sample, which was mixed thoroughly and centrifuged at 5,000 x g for 10 seconds. The sample was incubated at 70°C for 6 minutes then cooled at 4°C before centrifuged at 5,000 x g for 10 seconds. 125 µl of Second-Cycle, Second-Strand Master Mix (
Table 27) were added to the sample, which was thoroughly mixed and briefly centrifuged at 5,000 x g for 10 seconds. The sample was incubated at 16°C for 2 hours, and then 2 µl of T4 DNA polymerase (Affymetrix) was added to the reaction, and incubated at 16°C for 10 minutes before cooling to 4°C for 2 minutes and centrifuged again at 5,000 x g for 10 seconds. The second-cycle, second-strand cDNA product was purified as previously described in "cDNA purification" protocol prior to *in vitro* transcription and biotin labelling.

SECOND-CYCLE, SECOND-STRAND MASTER MIX (3x)

Components	Quantity	Supplier
RNase-Free Water	264.0 µl	Affymetrix
17.5 mM MgCl2	90.0 µl	Affymetrix
10 mM dNTP	9.0 µl	Affymetrix
E.coli DNA Polymerase	12.0 µl	Affymetrix

Table 27: Composition of Second-Cycle, Second-Strand Master Mix (3x)

3.2.2.8 Biotin-Labeled cRNA Synthesis

The final *in vitro* transcription (IVT) reaction involved synthesis of Biotin tagged cRNA, which was carried out following the same method as in the previously described IVT protocol with a minor alteration in which all uracil bases were tagged with a Biotin molecule (supplied by Affymetrix). 12 µl of purified second-cycle, second-strand cDNA were combined with 28 µl of Biotin-labelled cRNA IVT Master Mix (Table 28). This mixture was briefly centrifuged at 5,000 x g for 30 seconds and incubated at 37°C for 16 hours. The Biotin-labelled cRNA was purified as previously described in "cRNA purification" protocol prior to fragmentation.

BIOTIN-LABELLED CRNA, IVT MASTER MIX (3x)

Components	Quantity	Supplier
RNase-Free Water	24.0 µl	Affymetrix
10x IVT Labelling Buffer	12.0 µl	Affymetrix
IVT Labeling NTP Mix	36.0 µl	Affymetrix
IVT Labeling Enzyme Mix	12.0 µl	Affymetrix

Table 28: Composition of Biotin-Labelled cRNA, IVT Master Mix (3x)

3.2.2.9 Biotin-Labeled cRNA Fragmentation

In this protocol the purified Biotin-labelled cRNA was randomly fragmented into pieces between 30 to 400 base pairs in length in preparation for microarray hybridisation. 20 µg of purified biotin labelled cRNA in a 32 µl volume were combined with 8 µl of 5 x fragmentation buffer (Affymetrix), which was incubated at 94°C for 35 minutes and immediately transferred to 4°C for a final incubation of 5 minutes.

3.2.2.10 Biotin-Labeled cRNA Eukaryotic Target Hybridisation

30 µl of fragmented biotin labelled cRNA (15 µg) were combined with 270 µl of Hybridisation Master Mix (Table 29). This hybridisation cocktail was incubated at 94°C for 5 minutes, then incubated at 45°C for 5 minutes, and then briefly centrifuged at 15 to 30°C for 5 minutes. The hybridisation cocktail was incubated at 99°C for 5 minutes whilst an Affymetrix HG-U133 Plus 2 chip was pre-hybridised with 200 µl of Hybridisation Buffer (Table 30) and incubated at 45°C for 10 minutes. The hybridisation cocktail was incubated at 45°C for 5 minutes and then centrifuged at 5,000 x g for 5 minutes. The Hybridisation Buffer (Table 31) was removed from the chip and was replaced with 200 µl of the hybridisation cocktail, and the chip was incubated at 45°C for 16 hours.

HYBRIDISATION MASTER MIX

Components	Quantity	Supplier
Control Oligonucleotide B2 (3 nM)	5 µl	Affymetrix
20x Eukaryotic Hybridisation Control	15 µl	Affymetrix
Herring Sperm DNA (10 mg/ml)	3 µl	Promega
Acetylated BSA (50 mg/ml)	3 µl	Invitrogen
DMSO	30 µl	Sigma-Aldrich
Distilled Water	64 µl	-

Table 29: Composition of Hybridisation Master Mix

MES STOCK BUFFER (12x)

Components	Quantity	Supplier
MES hydrate	64.61 g	Sigma-Aldrich
MES Sodium Salt	193.3 g	Sigma-Aldrich
Molecular Biological Grade water	800 ml	Cambrex

Table 30: Composition of MES Stock Buffer (12x)

HYBRIDISATION BUFFER (2x)

Components	Quantity	Supplier
12x MES Stock Buffer (Table **30**)	8.3 ml	-
5 M NaCl	17.7 ml	Ambion
0.5 M EDTA	4.0 ml	Sigma-Aldrich
10% Tween-20	0.1 ml	Pierce Chemical
Water	39.8 ml	

Table 31: Composition of Hybridisation Buffer (2x)

3.2.2.11 Probe Array Wash, Stain and Scan

The hybridisation cocktail was removed from the chip, replaced with 200 µl of pre-wash buffer and the chip was incubated at 50°C for 30 minutes. A priming cycle was carried out on the fluidics station (Affymetrix) which rinses the machine and stabilises the fluid dynamics ready for washing and staining. The hybridised chip was placed inside the fluidic station together with Stain 1: Streptavidin Solution (Table 35) and Stain 2: Antibody Solution (Table 37). A program was started which began with 10 cycles of 2 mixes / cycle with Wash Buffer A: Non-Stringent Wash Buffer (Table 32) at 30°C, followed by 4 cycles of 15 mixes / cycle with Wash Buffer B: Stringent Wash Buffer (Table 33) at 50°C. The probe array was stained for 300 seconds with Stain 1: Streptavidin Solution (Table 35) at 35°C, followed by 10 cycles of 2 mixes / cycle with Wash Buffer A: Non-Stringent Wash Buffer (Table 32) at 30°C. The probe array was stained for 300 seconds with Stain 2: Antibody Solution (Table 33) at 35°C, followed by 300 seconds with Stain 1: Streptavidin Solution (Table 35) at 35°C. The array was washed with 15 cycles of 4 mixes / cycle with Wash Buffer A: Non-Stringent Wash Buffer (Table 32) at 35°C, before finally the probe array was filled with Array Holding Buffer (Affymetrix). The probe array was scanned using an Agilent GeneArray Scanner and the resultant .dat image was analysed for probe intensities with Affymetrix Microarray Suite software. The probe intensities were exported as .CEL files and analysed with GeneSpring GX 10 (Agilent) microarray software.

WASH BUFFER A: NON-STRINGENT WASH BUFFER

Components	Quantity	Supplier
20x SSPE (3 M NaCl, 0.2 M NaH2PO4, 0.02 M EDTA)	300 ml	Cambrex
10% Tween-20	1 ml	Pierce Chemical
Water	699 ml	-

Table 32: Composition of Wash Buffer A: Non-Stringent Wash Buffer

WASH BUFFER B: STRINGENT WASH BUFFER

Components	Quantity	Supplier
12x MES Stock Buffer	83.3 ml	-
5 M NaCl	5.2 ml	Ambion
10% Tween-20	1 ml	Pierce Chemical
Water	910.5 ml	-

Table 33: Composition of Wash Buffer B: Stringent Wash Buffer

2x STAIN BUFFER

Components	Quantity	Supplier
12x MES Stock Buffer	41.7 ml	-
5 M NaCl	92.5 ml	Ambion
10% Tween-20	2.5 ml	Pierce Chemical
Water	113.3 ml	-

Table 34: Composition of 2x Stain Buffer

STAIN 1: STREPTAVIDIN SOLUTION

Components	Quantity	Supplier
2X Staining Buffer (Table **34**)	300 µl	-
Acetylated BSA (50 mg/ml)	24 µl	Invitrogen
Streptavidin (1 mg/ml)	6 µl	Molecular Probes
Water	270 µl	-

Table 35: Composition of Streptavidin Solution

GOAT IGG STOCK (10mg/ml)

Components	Quantity	Supplier
Goat IgG	50 mg	Sigma-Aldrich
150 mM NaCl	5 ml	Ambion

Table 36: Composition of Goat IgG Stock (10mg/ml)

STAIN 2: ANTIBODY SOLUTION

Components	Quantity	Supplier
2X Staining Buffer	300 µl	-
Acetylated BSA (50 mg/ml)	24 µl	Invitrogen
Goat IgG (10mg/ml) (Table **36**)	6 µl	-
Biotinylated anti-strepavidin antibody (0.5 mg/ml)	6 µl	Vector Laboratories
Water	270 µl	-

Table 37: Composition of Antibody Solution

3.3 Results

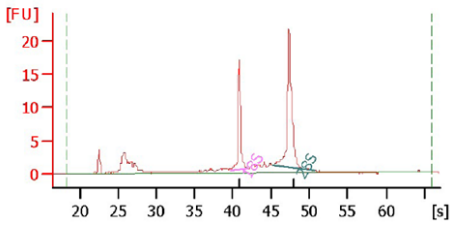
Due to the *in silico* nature of microarrays, the methods of analysis used are chiefly dependent on the results, and therefore within this section it seems logical to include excerpts of result-dependent methods interlaced within some parts of the results for ease of reading.

3.3.1 Microarray Quality Control

3.3.1.1 RNA Quality and Integrity: Pre-Microarray (Bioanalyser Results)

The starting RNA integrity and quality of any microarray experiment is of the upmost importance and as one would expect, poor quality RNA will always give poor quality results. The assessments of RNA integrity and quality for each of the samples were performed on an Agilent 2100 Bioanalyser (Agilent Technologies) before they were amplified and target labelled for hybridisation onto microarrays. The results of all microarray samples from the Bioanalyser analysis are shown in the table of Figure 39 displaying the RNA concentration, the rRNA ratio (28s / 18s), and the RNA integrity number. The output results from each sample were displayed in the form of an electropherogram from which the ribosomal peaks of 28S and 18S were measured and the rRNA ratio calculated. The results table (Figure 39) shows the RNA quantity for the majority of samples to be too low for one-cycle cDNA synthesis protocol (minimum total quantity of 1 µg required) as the average total quantity of RNA which was available was about 300 ng, and so a two-cycle cDNA synthesis protocol was used. Despite the low quantities of RNA in each of the samples, the quantity and integrity were very good as indicated by the rRNA ratios and RNA Integrity Numbers which must be 1.8 - 2.0 and >8 respectively. The table in Figure 39 shows the quality control results only from samples which were used in the microarray, and does not include samples which were of insufficient quality and quantity for microarray progression. As such was the high stringency of this test that if all sorted and unsorted RNA samples produced within this project were tested on the Bioanalyser and passed RNA quality control, then there would be close to six replicates for each microarray time point / day instead of two.

A



Overall Results

RNA Area:	92.8
RNA Concentration:	98 ng/µl
rRNA Ratio [28s / 18s]:	2.1
RNA Integrity Number (RIN):	9.6 (B.02.02)

Fragment table

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.73	42.03	15.5	16.7
28S	45.14	50.55	32.1	34.6

B

Sample	RNA Concentration (ng/µl)	rRNA Ratio (28s / 18s)	RNA Integrity Number (RIN)
Day 0B	98	2.1	9.6
Day 0C	122	1.9	9.4
Day 4A	72	2.2	8.6
Day 4B	89	1.8	9.1
Day 6A	53	1.8	8.3
Day 6B	58	1.7	8.1
Day 8A	76	2.0	9.0
Day 8B	62	1.7	8.3

Figure 39: Bioanalyser Results – RNA Integrity and Quality

A: Results from the Agilent 2100 Bioanalyser are shown in the form of an electropherogram together with a RNA concentration measurement, an rRNA ratio, and RNA Integrity Number. B: Table of results from all microarray samples showing the RNA concentration (required minimum concentration of 10 ng / µl), rRNA ratio (required between 1.8 and 2.0), and RNA Integrity Number (required 8.0 or higher).

3.3.1.2 Introduction to Affymetrix Microarray and Agilent GeneSpring

The Affymetrix array system utilises a probe set system which assigns multiple probes to each transcript and thus this approach of multiple independent measurements translates into greater accuracy and reproducibility of array data. The Gene Chip Human Genome U133 Plus 2 microarrays used in this experiment are produced by Affymetrix and contain a Brobdingnagian 54,675 probe sets which provide comprehensive cover of the entire human genome including gene fragments and multiple splice variants. The 54,675 probe sets translates into approximately 47,400 unique transcripts (with a number of transcripts represented by multiple probe sets), and these transcripts of multiple splice variants match up to over 38,500 genes [272].

The GeneSpring microarray analysis software suite was designed firmly with biologists in mind and enables rapid analysis of expression data, whilst simplifying the complicated statistics thus allowing the more important biological questions to be answered. The GeneSpring software is currently produced by Agilent Technologies and further information

together with download links may be found on their website [273]. GeneSpring contains a vast variety of statistical tools for testing differential expression which include paired and unpaired t-tests, one-way and multi-way analysis of variance (ANOVA), and permutative p-value computations. GeneSpring also offers a number of tools (Principle Components Analysis, Clustering Algorithms, Find Similar Entities tool) which may assist in the discovery of patterns within the expression data between samples which would effectively help to reduce the number of genes which may be of potential interest allowing for deeper analysis, and put these findings into a biological context by exploring which biological process, molecular function and biological pathways are involved or affected.

3.3.1.3 Experiment Setup

The following method of experiment setup for GeneSpring was written in such detail as to allow the reader to replicate this analysis with ease and certainty. The microarray probe intensities data were exported as .CEL files ready to be analysed using GeneSpring GX 10 which began with the creation of a 'new experiment' (Figure 40). An experiment name was given for the identification of array samples and the experiment type – 'Affymetrix microarray' was selected. The .CEL files which describe the intensities determined for every feature on a chip were located and imported into GeneSpring and the 8 individual array sets were normalised using the Robust Multiarray Average (RMA) summarisation algorithm for background correction and the baseline was set to the median of all samples. The purpose of this normalisation step allows multiple Affymetrix arrays to be reliably compared against each other, as the presence of background data noise is a problem which may have been attributed by a number of sources with the most common being non-specific binding of cRNA to a particular probe. Affymetrix arrays, instead of spotting the full sequence of the gene on a chip (Spotted arrays), selects between 11 - 20 probes (~25 bases) for each gene and this probe set is known as a Perfect Match (PM) probe. Next to each of these probes are another set of probes which represent the same 25 base sequence but these probes crucially contain a single base mutation in the middle of the sequence, and thus this probe set is a called Mis-Match (MM) probe. The reason for the inclusion of these MM probes is for the detection of cross-hybridisation / non-specific binding and the intensities of MM probes subtracted from PM probes may offer more accurate overall signal intensities (MAS5 Algorithm). However, in practice it has been shown that the signal intensity of MM probes may often be larger than those of PM probes, and thus results in overall negative expression values, which are nonsensical. The RMA algorithm background corrects without involving the subtraction of the MM probe values and uses only the PM data (a combination of true signal and background

non-specific signal) for which it fits a combination of 'noise' (normal) and 'signal' (exponential) distributions for each array separately . A baseline of zero was set to the median of all samples across the arrays (which was a default option as part of the GeneSpring software) to provides a neater looking profile plot with similar numbers of high intensities as there are low intensities (Figure 40).

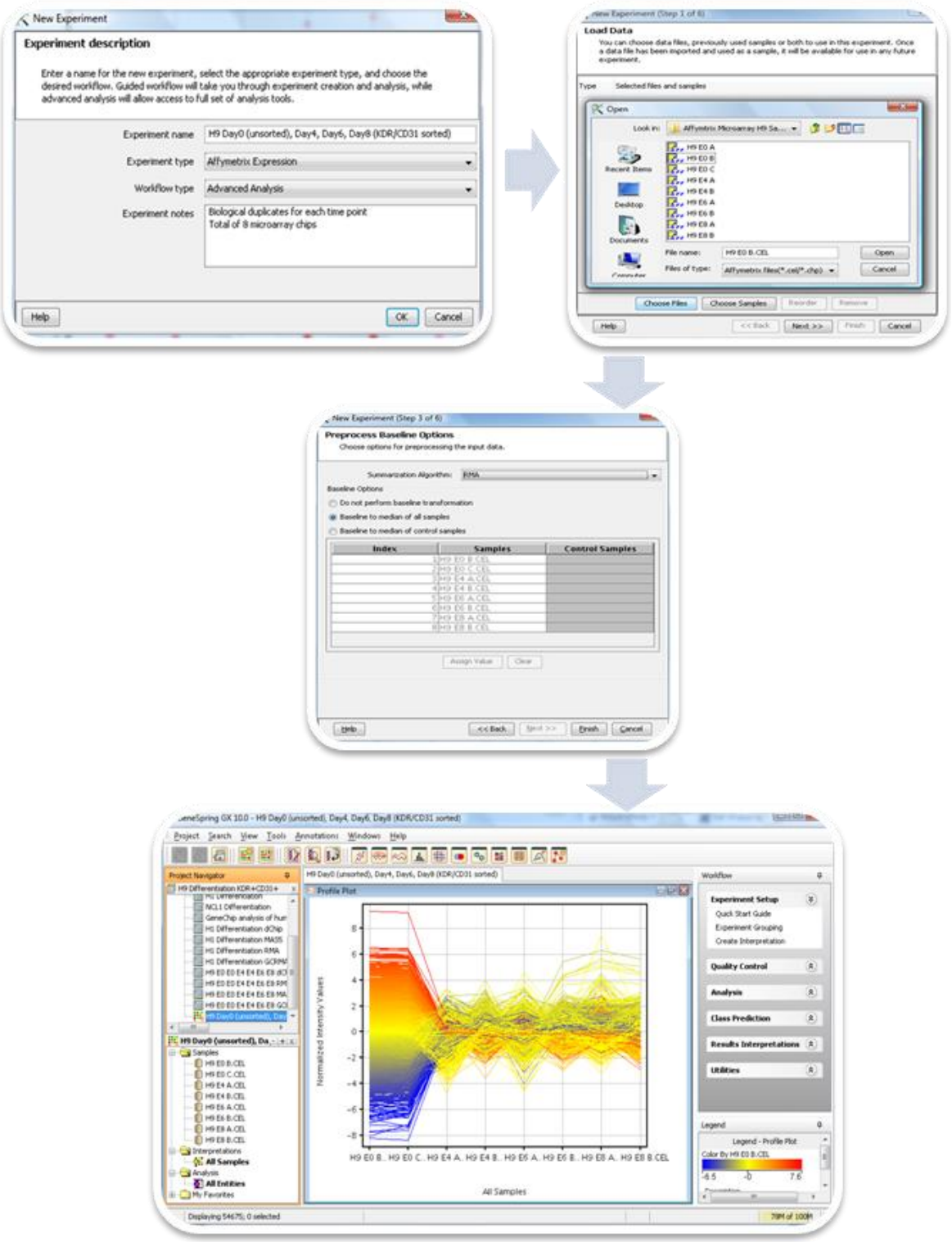


Figure 40: GeneSpring – Experiment Setup

These screenshots illustrate the experiment setup of a microarray project which began from the completion of an experiment description include an experiment name, experiment type, and workflow type. The .CEL files were located and imported together as a batch and a RMA summarisation algorithm was applied during normalisation with the baseline (0) set to the median of all samples. Once the experiment set up was completed, the gene expression profiles of all transcripts from all samples were displayed in a profile plot ready for analysis.

With the probe intensity values of each of the 8 samples imported and normalised, experiment parameters which assist to define the grouping / replicate structure of the experiment were added using 'Experiment Grouping'. A new parameter named 'Day' was added to allow the replicate samples to be grouped together and these grouped samples were also organised into chronological order. An experiment interpretation was created using 'Create Interpretation', in which samples from the experiment parameter (Day) were selected (Day 0, Day 4, Day 6, Day 8) and averaged, and thus the average intensity values across the replicates were used for visualisation and analysis (Figure 41). Within this report, interpretations are underlined, for example Day 0 against Day 4 against Day 6 against Day 8 (Figure 41), to improve the readability of this report. Lists of genes (entities) created within GeneSpring as a result of the analyses are highlighted in **bold** within this report, for example **Up-Regulating Genes from Day 0 to Day 4**.

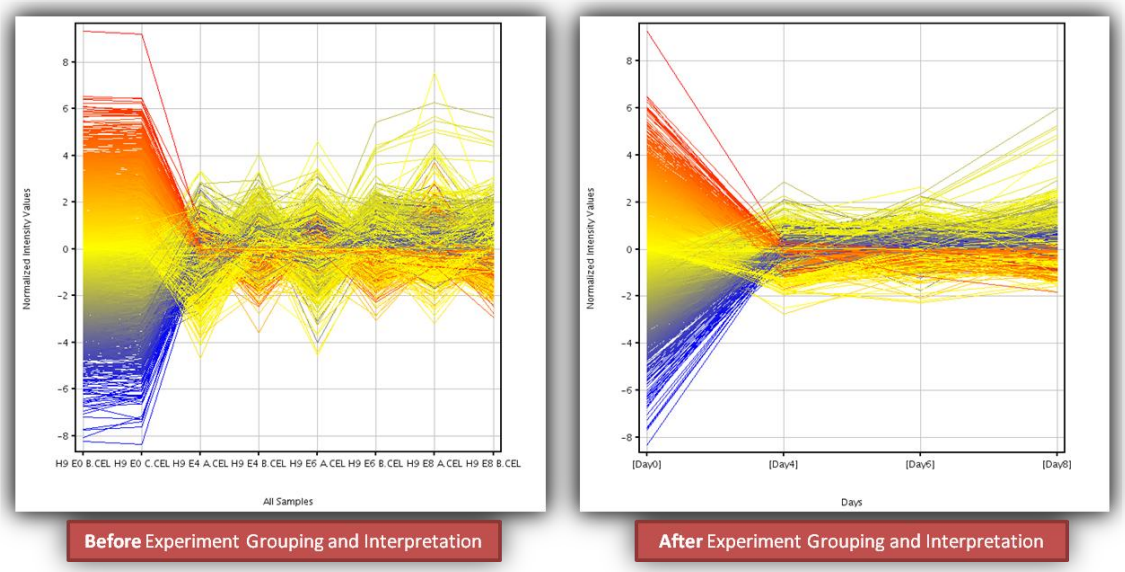


Figure 41: GeneSpring GX – Experiment Grouping and Experiment Interpretation

The profile plot before experiment grouping and interpretation (left) shows the gene expression of all of the individual biological replicate samples (8), and after experiment grouping and interpretation the replicates were averaged and renamed.

3.3.1.4 RNA Quality and Integrity: Post-Microarray (GeneSpring Results)

Before analysis may commence it is crucial that the data provided by the microarrays are true and their reliability must therefore be scrutinised by a series of quality control measures, and any samples which appear aberrant must be removed at this stage.

The first of these measures is the Principle Component Analysis (PCA) which tests if the sample replicates within the same day / time point correlate closer with each other than with samples from different days / time points (Figure 42C). The Principle Component Analysis (PCA) is a widely used tool for exploratory data analysis to uncover unknown trends and attempts at evaluating correlation between variables by condensing the data into a number of principle components / dimensions (and more information on PCA may be located within a PDF file available from Agilent's website [274]). The results from the PCA plot (Figure 42C) show that overall the replicates of each day do cluster relatively close to each other with Day 0 replicates (red) situating the furthest away from the others and clearly displaying a noticeable difference between undifferentiated and differentiated samples. It does make logical sense that the differentiated samples are grouped together as these samples had undergone differentiation and flow cytometry sorted for KDR+ CD31+ and thus exposed to more variables than Day 0 samples. On a gene expression level, it is expected that changes between Day 0 to Day 4 would be larger than subsequent changes between Day 4 to Day 6 simply due to the obvious difference being that the Day 0 samples are from undifferentiated pluripotent cells and Day 4 samples are from differentiated sorted KDR+ CD31 cells. When comparing Day 4 to Day 6 with Day 6 to Day 8, the PCA shows greater difference between the former transition than the latter, and one reason for this may be due to the introduction of Vascular Endothelial Growth Factor (VEGF) into the differentiation media after Day 4 time point, and thus the resulting effect of the cytokine may be a possible cause as the media composition for Day 6 and Day 8 samples were identical.

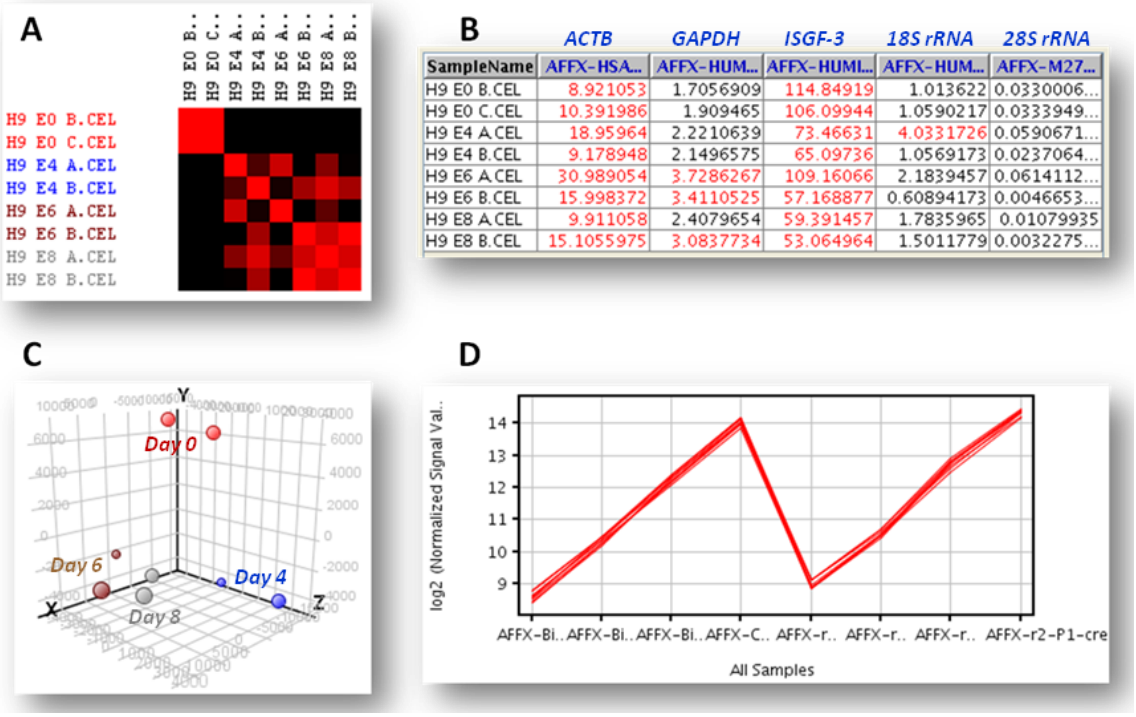


Figure 42: GeneSpring GX – Quality Control on Samples

The heat map (A) was produced from pairwise comparisons between all of the samples using the Pearson correlation coefficient to produce a colour gradient from black to red, showing the samples which highly correlate with each other in red and those with little or no correlation in black. The table of 3' / 5' ratios (B) shows the results of a number of internal controls which were accessed for RNA quality using including beta-actin (ACTB), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), interferon-stimulated gene factor 3 (ISGF-3), 18S ribosomal ribonucleic acid (18S rRNA), and 28S ribosomal ribonucleic acid (28S rRNA), and any values above 3 for any of the genes may suggest degradation with the sample. The Principle Component Analysis (PCA) plot (C) shows the correlation between all of the samples (Day 0 samples are coloured in red, Day 4 samples are coloured in blue, Day 6 samples are coloured in brown, and Day 8 samples are coloured in grey) and may be used to detect aberrant sample replicates. This hybridisation control plot (D) shows the expression of a number of biotin labelled cRNA transcripts (which were added in staggered qualities during the hybridisation step) from all samples to detect any anomalies which may have arisen from the hybridisation and washing processes.

The next control measure involved the calculation of Pearson correlation coefficients for each pair of samples / arrays and these results were displayed in the form of a heatmap (Figure 42A) showing by colour intensity (red) the degree of correlation between samples. In very much the same way as how PCA would group sample replicates together, the correlation coefficient heatmap would be expected to also show high correlation values between sample replicates. The results of the Pearson Correlation Coefficient heat map (Figure 42A) showed very high correlation between the replicate samples of Day 0 (unsorted cells) and these samples show very low correlation to any sample of any of the other days. The results of the other samples / days are more complicated, as although the replicates (A and B) do correlate with each other to a degree, the level of correlation between different samples (for example E4A, E6A, and E8A) which were differentiated together was higher than replicates of the same day. These results illustrate very clearly the batch to batch differences between differentiation

experiments, and highlight the requirement for tighter control on minimising variation between replicates and contribute to the identification of true biological gene expression change data.

RNA sample quality was assessed again to confirm that the RNA which was used in the microarrays was of sufficiently high quality and not degraded during the amplification and hybridisation processes. The RNA was measured by a quality control sampling tool called 'Internal Controls' (Figure 42B) which displays the 3'/5' ratios for a number of specific probesets (included as standard in the Affymetrix microarrays) including beta-actin (ACTB) (AFFX-HSAC07/X00351__at), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), (AFFX-HUMGAPDH/M33197__at), interferon-stimulated gene factor 3 (ISGF-3) (AFFX-HUMISGF3A/M97935__at), 18S ribosomal ribonucleic acid (18S rRNA) (AFFX-HUMRGE/M10098__at), and 28S ribosomal ribonucleic acid (28S rRNA) (AFFX-M27830__at). The 3' / 5' ratio of the internal controls may offer an indication of the integrity and quality of the starting RNA as well as the efficiency cDNA synthesis and *in vitro* transcription of cRNA, and therefore a 1:1 molar value of 3' to 5' transcript should ideal give a signal ratio of 1 for perfect RNA. Affymetrix has stressed that there is no single cut-off threshold in the assessment of RNA quality with these internal controls due of the many different isoforms of these internal control genes which exist and their varied tissue specific expression patterns, but as a rough guideline they do refer to a threshold ratio of less than 3 for the majority of tissues and organisms [275]. The results from this assessment are displayed in the table in Figure 42B, and all values that are greater than 3 are highlighted in red, which one can immediately see that two entire columns of ACTB (second column of the table) and ISGF-3 (fourth column of the table) are red. These results suggest that there may be some problems associated with the samples, but as previously explained Affymetrix have suggested that the threshold value of 3 for a number of internal controls may not be suitable, and this test should rather be used to flag these deviant samples for further investigation. There are three red samples (H9 E6 A, H9 E6 B, and H9 E8 B) within the GAPDH column (third column of the table), but the values are only just over the arbitrary threshold value of 3 which does not present much concern. The final two remaining columns from 18S rRNA (fifth column of the table) and 28S rRNA (sixth column of the table) are almost perfect which correlates very well with the earlier results of the Bioanalyser (Figure 39).

Hybridisation controls (Figure 42D) were used as a measure of the quality of the hybridisation / washing steps and to pinpoint any inconsistencies associated with the samples by the expression level assessment of a number of spiked-in control probes of biotin-labelled

cRNA transcripts (bioB, bioC, bioD, and cre), which is of particular importance if the microarrays were not all processed (hybridised / washed) at the same time. The concentrations of these transcripts were known and when the samples / arrays were compared, the levels of these transcripts would be expected be the same and any differences would indicate inconsistencies with the hybridisation / wash stages. The results of the hybridisation control in Figure 42D show the expression of the biotin-labelled cRNA probes of each sample on a $\log^2$ scale, with the expression profile of all samples lining up almost perfectly with each other. This result suggests that any differences in gene expression between the samples should not be attributed to the hybridisation and washing stages, and thus adds credibility to any differences to being the result of biological effects.

In summary, the quality and integrity of the starting RNA from the beginning was very close to perfect and showed clean rRNA bands with very little protein / solvent contamination which resulted in high 28s / 18s rRNA ratios. The Pearson Correlation Coefficient heat map (Figure 42A) indicated a possible irregularity concerning the Day 6 replicates with Day 6B samples correlating closer to the Day 8 samples, but the results of Principle Component Analysis (PCA; Figure 42C) suggested that samples from Day 6 and Day 8 were actually all quite similar to each other. The internal control of 3' / 5' ratio assessment of a number of stably expressed genes showed mixed results but there is no perfect stably expressed gene for all tissue types, thus this result was only used as a suggestion that there may be some problems but depended too much on tissue / cell type. The hybridisation control showed samples which were hybridised and washed at different times to be identical in the expression level of a number of spiked in control probes, which meant that any variations in expression value of genes between replicates were not introduced by the hybridisation and washing process. The results of this battery of quality control tests did suggest some inconsistencies with the samples but nothing major was flagged up, and such variations between replicates may be attributed to the large number of variables associated with the multiple protocols for which the replicates were subjected to. In conclusion, all microarray samples successfully satisfied the interrogation of the quality control tests on a samples level, and thus progression was permitted for the assessment of quality control assays on a transcript level.

3.3.1.5 Quality Control on Validity of Gene Expression

A quantitative polymerase chain reaction (qPCR) approach was used as an independent method to validate gene expression levels of a number of transcripts selected from the microarray. The transcripts which were selected were a handful of stably expressed genes including *GAPDH, RPL13A, SDHA,* and *TBP*, pluripotency and self-renewal genes *OCT4* and *NANOG*, and genes that are specific to all three germ layers including ectoderm genes *NESTIN*, *CER1, FOXA2,* endoderm genes *CCND1, GFAP, GATA4*, and mesoderm genes *BRY* and *MIXL1*, and haematopoietic genes including *CD31, KDR, CD34, CD144,* and *CD117*. The qPCR protocol for this quality control assay on genes was undertaken using the same protocol as previously described in Chapter 2 using cDNA from both H9 (Day 0), H9 KDR+ CD31+ (Day 4, Day 6, Day 8) and H1 (Day 0), H1 KDR+ CD31+ (Day 4, Day 6, Day 8) as template. The expression profile of the genes obtained from qPCR from both H9 and H1 cell lines were compared against the microarray expression profile of the same gene. The addition of qPCR data from the H1 HESC line adds confidence to the experiment that the results displayed for the HESC line H9 (in both microarray and qPCR) were not attributed solely to H9 and that the results of this study is reproducible and representative of any HESC line. The Ensembl Genome Brower identification code was obtained for each gene of interest from the National Center for Biotechnology Information (NCBI) Entrez Gene database [276] and this code was used within GeneSpring to search for the transcript rather than using the gene name or gene symbol, with the reason being that any one gene may also include multiple cDNA / mRNA fragments which may not be representative of the full gene and aids in the identification of the correct splice variant if applicable. The results from the search were saved through the creation of a new gene list for each transcript and raw data values were retrieved from the transcript inspector which was accessed by double clicking on the transcript within the profile plot browser. The data obtained from the microarray were plotted and compared with those data obtained from qPCR analysis, and any major differences in expression at this stage would render the array data unreliable and thus results obtained from subsequent downstream analysis would be questionable.

The qPCR results of the four stably expressed house-keeping genes (*GAPDH*, *RPL13A*, *SDHA*, and *TBP*) which were used in the qPCR normalisation of each transcript are shown in Figure 43, and as one would expect their expression remained relatively constant throughout the time course with only negligible changes. There is no known gene within the human genome whose expression is constant in all cells of all tissues at all developmental stages

[277], and thus this is the reason why a number of known stably expressed genes (*GAPDH*, *RPL13A*, *SDHA*, and *TBP*) were used in combination for the normalisation of other genes.

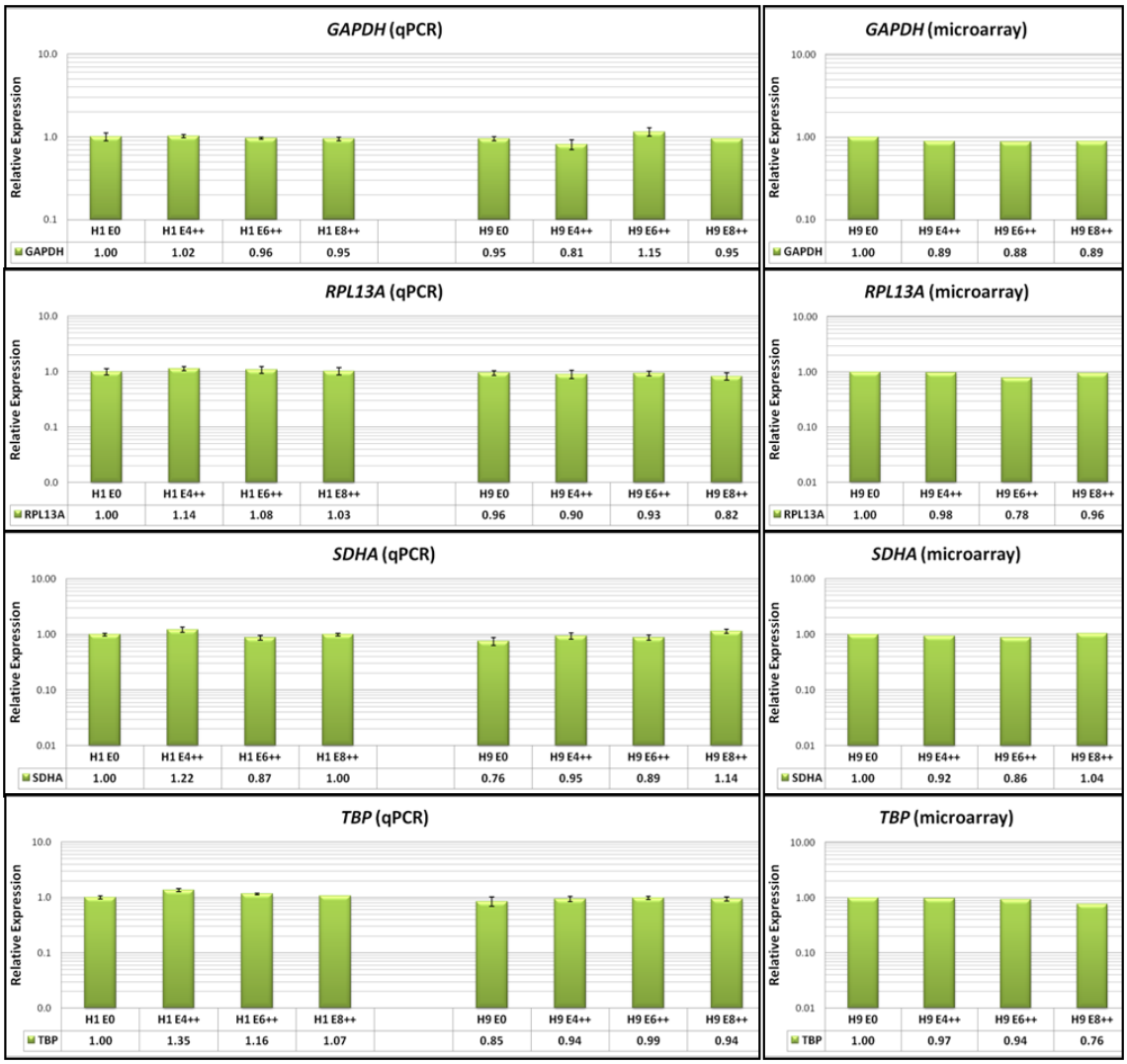


Figure 43: Microarray Quality Control – Gene expression of House Keeping Genes

The gene expression level of a number of common stably expressed house-keeping genes (including *GAPDH, RPL13A, SDHA, TBP*) were compared between the results acquired from microarray and qPCR. These genes were all normalised against each other and the combined result was normalised with all microarray and qPCR transcripts. The qPCR results from an independent cell line H1 (cultured, differentiated, and sorted using the exact same protocols as with H9 cell line) were included to add confidence that the results shown were not specific to the H9 cell line. The relative expression value for H1 Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this.

The classic embryonic stem cell pluripotent marker *OCT4* and self-renewal marker *NANOG* were selected from this experiment as a control for differentiation and as shown by their profile plots (Figure 44), the expression level at Day 0 began very high but rapidly decreased with progression of differentiation through the time course as expected. Lineage-specific genes (*FOXA2, CER1, NESTIN, BRY, MIXL1, CCND1, GFAP, GATA4*) were selected to ascertain the direction of differentiation within this project, providing an indication as to

whether this differentiation had progressed down the mesoderm (*BRY* and *MIXL1*), the endoderm (*CCND1, GFAP,* and *GATA4*), the ectoderm (*FOXA2, CER1,* and *NESTIN*) lineage, or a combination all three. Analyses of the gene expression profiles of these lineage-specific genes shown in Figure 44 and Figure 45 revealed no significant changes in expression lineage preference through the differentiation time course. However it was expected that genes from the mesoderm lineage (*BRY, MIXL1;* Figure 45) would be up-regulated through this haematopoietic differentiation and this may indeed have been the case if the samples of Day 4, Day 6, and Day 8 were from embryoid bodies, but the RNA for this microarray were from KDR and CD31 positive cells. Cells which are positive for both KDR and CD31 are enriched with haematopoietic and endothelial potential as previously described, and therefore it was important to screen for known haematopoietic and endothelial genes.

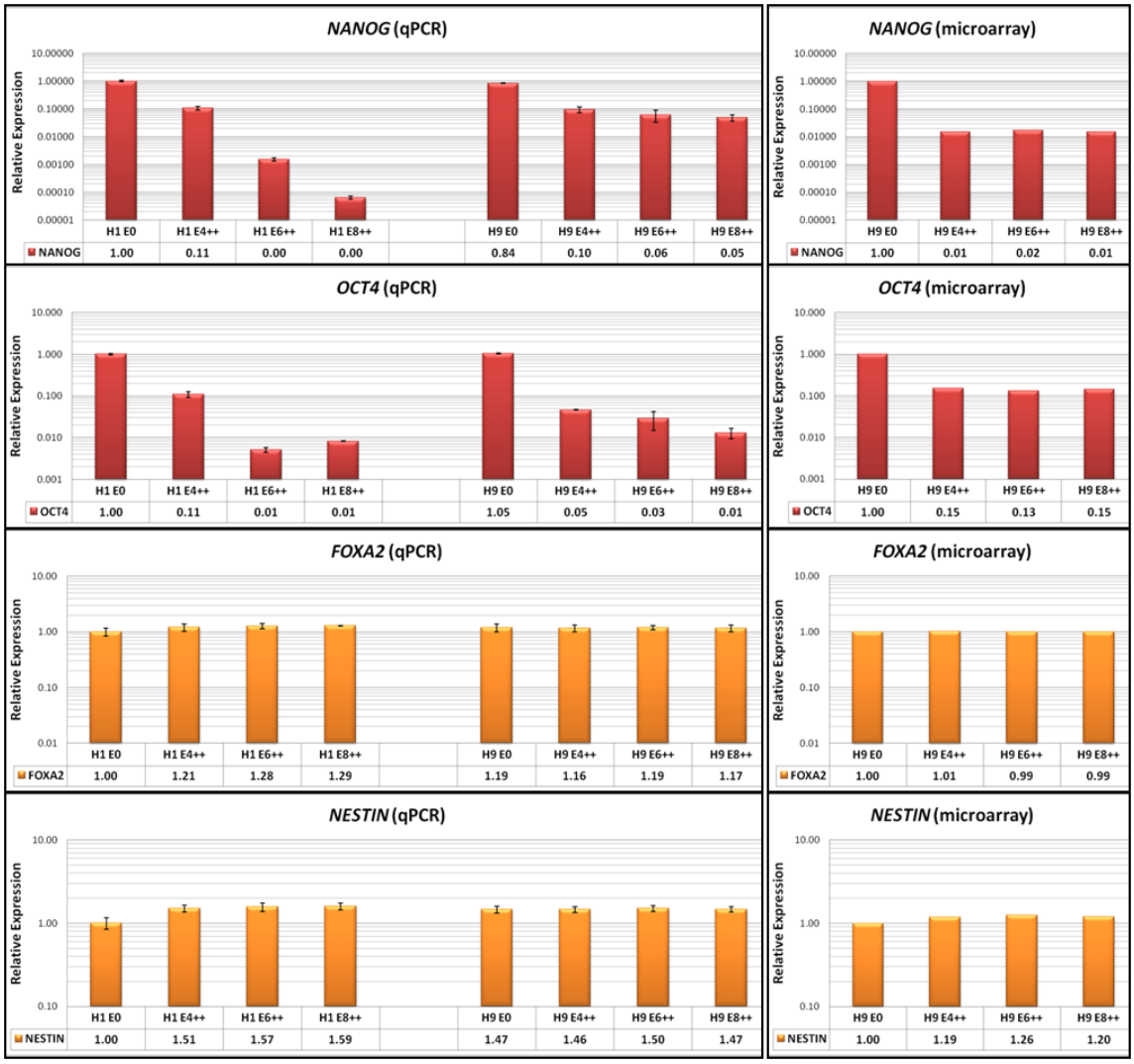


Figure 44: Microarray Quality Control – Gene expression of Embryonic Stem Cell and Ectoderm Genes

These plots show the gene expression levels of a number of classical embryonic stem cell pluripotency and self-renewal markers (*OCT4* and *NANOG*) and ectodermal genes (*FOXA2* and *NESTIN*). The plots contain results acquired from microarray and qPCR experiments from the H9 hESC cell line but also include qPCR results from the H1 hESC cell line (which were cultured, differentiated, and sorted using the exact same protocols as with H9 cell line) which adds confidence to the overall experiment and shows that these haematopoietic candidates are not exclusively generated by the H9 hESC cell line. The relative expression value for H1 Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. The expression value of each sample for each gene were normalised against stably expressed genes *GAPDH, TBP, SDHA,* and *RPL13A*.The error bars were generated using the standard error of the mean of a 3 replicates.

Figure 45: Microarray Quality Control – Gene expression of Mesoderm and Endoderm Genes

These plots show the gene expression levels of a number of common mesodermal genes (BRY and MIXL1) and endodermal genes (CCND1, GFAP and GATA4). The plots contain results acquired from microarray and qPCR experiments from the H9 hESC cell line but also include qPCR results from the H1 hESC cell line (which were cultured, differentiated, and sorted using the exact same protocols as with H9 cell line) and adds confidence to the result that these haematopoietic candidates are not exclusive to the H9 hESC cell line. The relative expression value for H1 Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. The expression value of each sample for each gene were normalised against stably expressed genes GAPDH, TBP, SDHA, and RPL13A. The error bars were generated using the standard error of the mean of a 3 replicates.

The microarray gene expression profile plots of the known haematopoietic genes (*CD31, KDR, CD34, CD144*, and *CD117*) are shown in Figure 46 together with results from qPCR of both H9 and H1 hESC lines. The expression profiles in both the microarray and qPCRs of the majority

of the genes (*CD31, KDR, CD34,* and *CD144*) show very large and significant increases in gene expression (>2 fold change) from Day 0 to Day 4 with the exception of *CD117* which display only a 1.3 fold increase in expression from the microarray data, which is in stark contrast to *CD144* which show a 342 fold increase between the same two time points. All of the genes do however exhibit far smaller changes in expression between subsequent days of differentiation from Day 4 to Day 8, and this lack of change may be because the time points are relatively close together and perhaps longer time points such as Day 20 and Day 30 may yield greater expression changes. A possible theory of why there is such small changes between the sorted cells (KDR+ CD31+) may be because the expression of these particular cells are already expressing these genes at maximum expression level and perhaps even at Day 20, the expression levels of these genes may not increase any further. This theory appears to agree with this pattern of results as these sorted cells may have already been committed to haematopoietic / endothelial cell fate and thus these genes (*CD31, KDR, CD34, CD144,* and *CD117*) which are responsible for this early commitment may not be required any longer for more later differentiation. It will be of great interest to later investigate which genes (downstream) are being affected by these highly expressed haematopoietic genes and identify genes which show a large increase in expression between the later days (Day 4 to Day 6 and Day 6 to Day 8). It is important to reiterate that the corresponding qPCR data of gene expression for every transcript which was selected and analysed within Figure 43, Figure 44, Figure 45 and Figure 46 matched remarkable accurately with those from the microarray, and therefore as these results had satisfied quality control, further analysis of the microarray was allowed to proceed.

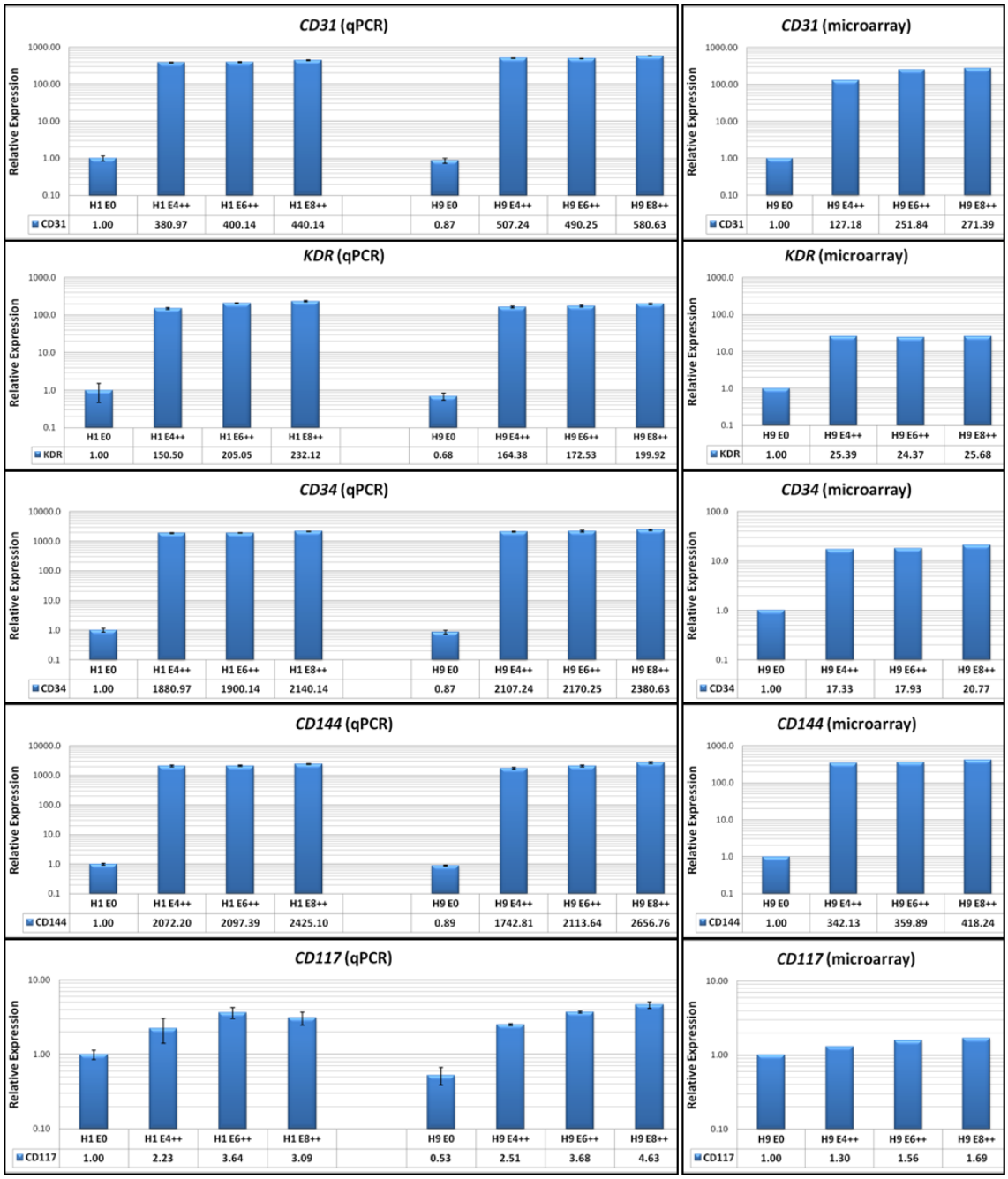


Figure 46: Quality Control – Gene expression comparison between qPCR and microarray of Haematopoietic genes

A number of genes were selected and their microarray gene expression levels from day 0 to day 8 were compared with those observed from quantitative polymerase chain reaction (qPCR) analysis. The plots contain a number of haematopoietic genes which include CD31, CD34, KDR, CD144, and CD117 and these expression levels were all normalised against a battery of housekeeping genes including GAPDH, RPL13A, TBP, and SDHA. The relative expression value for H1 Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. The expression value of each sample for each gene were normalised against stably expressed genes *GAPDH, TBP, SDHA,* and *RPL13A*. The error bars were generated using the standard error of the mean of a 3 replicates.

3.3.1.6 Quality Control on the Validity of the Experiment

It was of interest on a genome-wide gene expression level to investigate the various different human embryonic stem cell (HESC) lines and discover which line is most efficient in haematopoietic differentiation. As previously discussed not all HESC lines are created equally and cell line to cell line differences in gene expression do exist which result in differences in differentiation potentials between the lines. A variety of differentiated HESC (H1, H9, NCL1) whole embryoid bodies (EB) microarray data [278] were kindly provided by Dr. Owen Hughes and Prof. Linda Lako for a haematopoietic genes comparison. In addition, the merits of this experiment was assessed by comparing enriched sorted population of KDR+ CD31+ cells with un-sorted population of whole embryoid bodies (EB) and to observe any similarities / differences in haematopoietic potential on the transcript level. Studies have already shown that the KDR+ CD31+ cell population *in vitro* are enriched for haematopoietic progenitors [157, 181, 182], and colony-forming assays (MethoCult) have demonstrated that these cells have the enhanced ability to give rise to a spectrum of haematopoietic colony types (CFU-E, BFU-E, CFU-GM, CFU-G, CFU-M, and CFU-GEMM) (Chapter 2), and thus a host of haematopoietic genes would be expected to be highly expressed within the microarray data.

Microarray data from HESC lines, H1, H9, and NCL1 were imported into GeneSpring as a new experiment using the protocol outlined and described. Transcripts which were up-regulated with a fold change of 2 or greater from Day 0 to Day 4 were selected from EB data of H1, H9, and NCL1 cell lines for comparison. Within GeneSpring, up-regulating transcripts were selected using the 'Fold Change' tool and upon selection, within this tool 'All Entities' was chosen for the 'Entity List' and 'Day 0 vs. Day 4' was chosen for the 'Interpretation'. If the interpretation named 'Day 0 vs. Day 4' was not available, it was created (to include samples from Day 0 and Day 4 only) using the 'Create Interpretation' tool within the 'Experiment Setup' tab. As the interpretation only contains samples Day 0 and Day 4, the 'Pairs of Conditions' within 'Pairs Options' which were to be fold change interrogated were set as follows: 'Condition 1 = [0]' and 'Condition 2 = [4]', and the fold change 'Change cut-off' was set to '2'. Within the 'Fold Change' tab the fold change was sorted by direction (up / down) by clicking on the column header 'Regulation([0] vs [4])'. All probe set IDs which displayed an 'up' in the 'Regulation([0] vs [4])' column were selected and saved as '**H1** or **H9** or **NCL1 + Day0toDay4Up2fold**' by clicking on the 'Save custom list' button.

The lists of transcripts which up-regulated by a fold change of 2 or greater from each of the cell lines microarray data (**H1 + Day0toDay4Up2fold, H9 + Day0toDay4Up2fold, NCL1 + Day0toDay4Up2fold**) were first analysed within GeneSpring and then exported and analysed

within an ontology enrichment tool called Metacore (GeneGo; [279]). The GeneSpring analysis involved the discovery of the number of haematopoietic markers which were present within each gene list, as this was used as a basic indication of haematopoietic differentiation potential for each cell line. The results shown in Figure 47 suggest that the H1 cell line may be more superior in terms of haematopoietic differentiation potential than the others as it includes the greatest number of haematopoietic progenitor markers (*CD144*, *KDR*, and *CD117*). This result correlates very well with external studies by other groups in the same field whereby the H1 cell line was frequently preferred and used in haematopoietic differentiation experiments [49, 157, 181]. Deeper analysis which involved all transcripts present in each gene list was carried out by a Metacore analysis as previous described, which interrogates the transcripts as a whole and identifies the top genes pathways and process network from which the transcripts belong. The results from this analysis are shown in Figure 48 (process networks) and Figure 49 (gene ontology), and a comparison between the different unsorted cell lines (whole EBs) reveal differences in the developmental processes but these processes (for example System Development, Developmental Process, Cell Differentiation, etc.) are very vague and very general with no obvious differential preference. These results were however, based on the nature of this analysis and differentiation experiment, not surprising as one would not expect a preference in differentiation towards a particular lineage using a standard differentiation protocol. The standard differentiation protocol as previous described in Chapter 2 consists of standard human embryonic stem cell media but with Knockout Serum Replacement (Invitrogen) substituted with Foetal Bovine Serum (Autogen Bioclear) which causes HESCs to differentiate towards cells of all three germ layers [278].

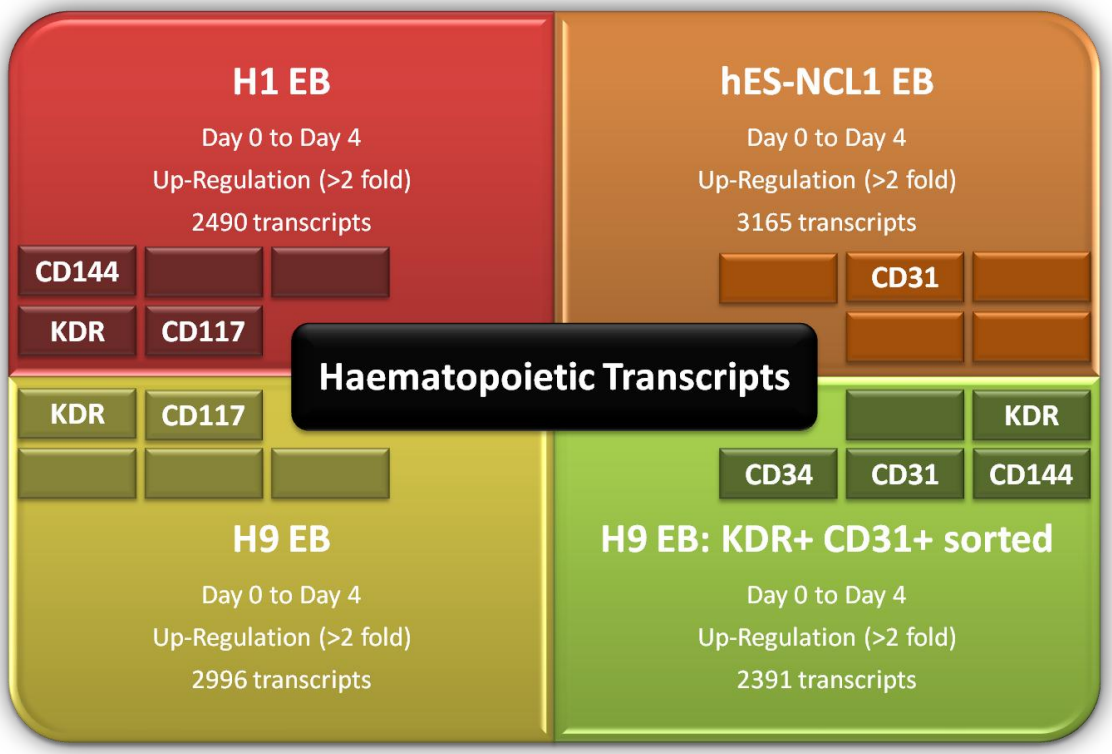


Figure 47: Haematopoietic Transcripts - Comparison between different hESC lines and Sorted / Unsorted Cells

An analysis of the key haematopoietic transcripts (CD34, CD31, KDR, CD117, CD144) within the unsorted cell lines H1, H9, and NCL1 reveal distinct differences in their potentials in haematopoietic differentiation. The unsorted cell lines were differentiated using standard differentiation media and it appears from the number of haematopoietic transcripts present that the haematopoietic potential of H1 cell line is superior, followed by H9 and NCL1. CD117 however is crucially not expressed in the H9 KDR+ CD31+ cell line but are expressed in H9 and H1 unsorted cells which may suggest the differentiation protocol was not favourable towards CD117+ cells or that KDR+ CD31+ cells do not express CD117+ at a high level.

It was also of interest to compare the microarray gene expression data of unsorted H9 EBs with H9 KDR+ CD31+ sorted cells and reiterate the importance of KDR and CD31 expression for the differentiation of haematopoietic cells. The importance of these haematopoietic markers had already been demonstrated *in vitro* by the formation of colony forming units in Methylcellulose colony assay experiments. However it was of interest to further investigate the possible reasons for this enhanced haematopoietic differentiation efficiency on a genome level and discover the key gene maps and network processes which contributed to this outcome. As previous described, gene lists containing transcripts which increased in expression from Day 0 to Day 4 by a fold change of 2 fold or greater were created from each data set, and the preliminary analysis involved a comparison of the number of haematopoietic markers present in each list, followed by a more detailed Metacore analysis. The results from the haematopoietic transcripts comparison (Figure 47, Figure 48, Figure 49)

between unsorted cell lines (H1, H9, NCL) and sorted H9 (KDR+ CD31+) population show that the optimised differentiation protocol combined with enrichment for KDR+ CD31+ markers do enhance the differentiation towards haematopoiesis.

Figure 48: Process Networks – Comparison between different Cell lines and Sorted / Unsorted Cells

All transcripts which were up-regulated in their gene expression level by a fold change of 2 or greater between Day 0 and Day 4 of differentiation where selected and their process networks analysed. A comparison between these process networks reveals differences between unsorted cells belonging to different cell lines and the results of their process networks shows process representative from all three germ cell layers. The process network from the sorted (KDR+ CD31+) H9 population does show a marked difference from those belonging to unsorted cells and is strong biased towards mesoderm and haematopoiesis.

Figure 49: Gene Ontology Processes – Comparison between different Cell lines and Sorted / Unsorted Cells

The gene ontology processes of a number of different human embryonic stem cell lines were compared and analysed on transcripts which were up-regulated in their gene expression level by a fold change of 2 or greater between Day 0 and Day 4 of differentiation. A comparison between the gene ontology processes reveal differences between unsorted cells belonging to different cell lines and the processes identified are often vague and quite general (for example system / cell development). The gene ontology processes from the sorted (KDR+CD31+) H9 population however is rather more specific and does list a number of haematopoiesis-related processes which suggest that the sorted population of KDR+ and CD31+ cells do appear to be enriched for haematopoietic progenitors.

A larger number of haematopoietic progenitor markers were detected in the KDR+CD31+ sorted population compared to the unsorted cell lines (Figure 47) including CD34, CD31, CD144, and KDR, with just CD117 missing. It is however interesting that CD117 is present within the unsorted cell lines of H9 and H1, and a number of hypotheses may offer an explanation to this outcome, with one being that the differentiation protocol requires optimising to promote and up-regulate haematopoietic gene expression at the correct time. Another hypothesis may simply be that the KDR+ CD31+ cells are not CD117+ and this gene is not up-regulated within this population at these time points (Day 4, Day 6, and Day 8). Studies have shown that CD117 is expressed on other cells besides haematopoietic stem cells which include some types of cancers (gastrointestinal stromal tumors [280] and chronic myelogenous leukaemia [281]).

A process networks analysis (Figure 48) between unsorted (H1, H9, NCL) and sorted cell lines (H9 KDR+CD31+) revealed stark differences between the specificity of developmental processes, with the results of unsorted cells displaying general processes of no specific developmental lineage (mesoderm, endoderm, and ectoderm). The results of the sorted cells however displayed a strong bias towards developmental processes from the mesodermal lineage, more specifically haematopoietic development (platelets, endothelium-leucotye interactions, blood vessels, neutrophil, blood coagulation) and cardiac development (angiogenesis, wnt, b-catenin, notch, VEGF, IP3, integrin signalling). A gene ontology analysis (Figure 49) also uncovered a similar trend within the results of the sorted cells population which contain more haematopoietic-related process than any unsorted cell line.

To conclude, it does appears that these preliminary results of unsorted differentiated HESCs with standard differentiation media against sorted differentiated HESCs cells with haematopoietic differentiation media do support the merit of using the later for haematopoietic studies. These results do suggest that these sorted KDR+ and CD31+ cells are not pure haematopoietic progenitors as confirmed by the inclusion of a number of non-haematopoietic developmental processes, and a greater number of haematopoietic markers must be used if a purer haematopoietic population is desired. As previously explained, one could fulfil this in attempting to select and sort for all known cell surface haematopoietic markers, but this would come at the cost of the number of cells which one could in practise successfully sort for. This would have a knock on effect with regards to the quantity and quality of RNA which one could effectively isolate from the sorted cell population for reliable, repeatable microarray experiments. The conclusion of this study may result in the discovery of novel transcripts which may aid in the *in vitro* differentiation process or assist in the selection of haematopoietic progenitor cells. This would thus allow for higher haematopoietic

differentiation efficiencies and permit the use of additional markers without sacrificing cell numbers and affecting downstream applications.

3.3.2 Microarray Profiles Analysis

3.3.2.1 Microarray Expression Profile Screen

Within a human Affymetrix array there are 54,675 probesets (oligonucleotide probes complementary to corresponding transcripts) and therefore a logical starting point of an analysis of this magnitude would be to examine the gene expression profile progression through the experimental time points (Day 0 to Day 8). There are 4 time points per sample (Day 0, Day 4, Day 6, Day 8) in this experiment and the 3 pairwise chronological comparisons were investigated which are comprised of Day 0 against Day 4, Day 4 against Day 6, and Day 6 against Day 8. When focusing on a comparison between two time points (for example Day 0 to Day 4), any one particular transcript has one of three possible general expression options, with the first being that the transcript level may increase from Day 0 to Day 4, with the second being that the transcript level may decrease from Day 0 to Day 4, and the third being that the transcript level may remain unchanged from Day 0 to Day 4. A complete expression profile for a particular transcript would include all 3 pairwise comparisons and multiplied by 3 possible general expression options (increase, decrease, and no-change) for each, will give a total of 27 (3x3x3) different possible general expression profiles (Figure 50).

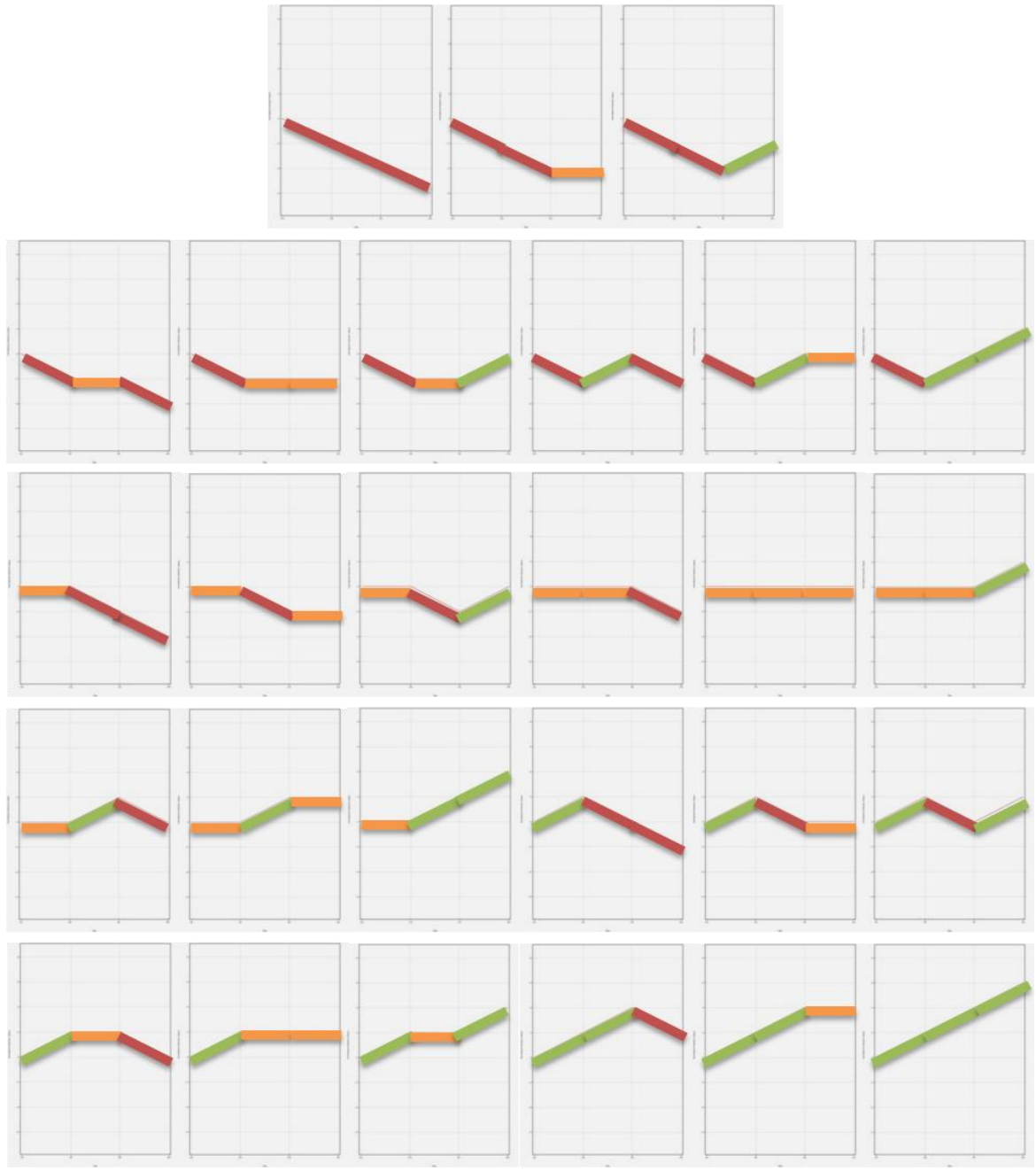


Figure 50: Microarray Expression Profiles - Theoretical Combinations Clustering

This figure of 27 expression profile plots showing the total number of combinations that will result from a 4 sample microarray experiment. This method of clustering is very general, but it will offer a rough idea of the number of transcripts which have similar expression patterns and possibly similar roles during differentiation. Colour coded profiles show down-regulation (red), up-regulation (green), and no-change (orange).

These general profile plots (Figure 50) were created in GeneSpring, where three new interpretations were created using the 'Create Interpretation' tool so that each interpretation would include just two days for pairwise comparisons instead of the original four. These interpretations (Day 0 against Day 4, Day 4 against Day 6, and Day 6 against Day 8) were used

in the creation of gene lists for each pairwise comparison and the first gene list contained transcripts with expression levels that increased from one time point to the other (for example '**Day 4 to Day 6 – Up-Regulation**'), the second contained transcripts which decreased (for example '**Day 4 to Day 6 – Down-Regulation**'), and the third contained transcripts which did not changed (<1.1 fold change, for example '**Day 4 to Day 6 – No-Change**'). The 'Fold Change' tool was used in the selection of transcripts which increased / decreased / unchanged in expression from Day 4 to Day 6, and within this tool '**All Entities**' was chosen for the 'Entity List' and 'Day 4 vs. Day 6' was chosen for the 'Interpretation'. As the interpretation only contains samples Day 4 and Day 6, the 'Pairs of Conditions' within 'Pairs Options' which were to be fold change interrogated were set as follows: 'Condition 1 = [4]' and 'Condition 2 = [6]'. In the selection of transcripts which increased / decreased from Day 4 to Day 6, the digits '1.1' was entered in the 'Change cut-off' option. All transcripts with an expression level fold change of 1.1 or greater were labelled as up-regulating and all transcripts which were less were considered as non-changing between the two days. A fold change of 1.0 is the same as no fold change and because the minimum incremental increase within GeneSpring was 0.1, 1.1 was therefore arbitrarily chosen as a minimum threshold by which non-changing genes (between 1 and 1.1) and up-regulating genes (>1.1) may be distinguished. Within the 'Fold Change' tab the fold change was sorted by direction (up (up-regulation) / down (down-regulation)) by clicking on the column header 'Regulation([4] vs [6])'. All probe set IDs which displayed an 'up' in the 'Regulation([4] vs [6])' column were selected and saved as '**Day4toDay6Up1.1fold**', and all probe set IDs which displayed a 'down' in the 'Regulation([4] vs [6])' column were selected and saved as '**Day4toDay6Down1.1fold**'. These lists only contain transcripts which have an expression fold change of 1.1 or greater, and therefore in order to select non-changing transcripts which are less than 1.1, a Venn diagram must be used to subtract the **Day4toDay6Up1.1fold** and **Day4toDay6Down1.1fold** lists from a list that contains all transcripts with any expression level change regardless of size and hence a fold change value of 1 (Figure 51A&B). Gene lists containing all transcripts which increased / decreased between the two days were selected (as previously described) using the 'Fold Change' tool with the same settings, but the fold change was set to '1' in the 'Change cut-off' option. Within the 'Fold Change' tab the fold change was sorted by direction and all probe set IDs which displayed an 'up' in the 'Regulation([4] vs [6])' column were selected and saved as '**Day4toDay6Up1.0fold**', and all probe set IDs which displayed a 'down' were selected and saved as '**Day4toDay6Down1.0fold**'. These gene lists were placed into Venn diagrams (Figure 51) (as previously described) and the non-changing gene lists were selected from the subtraction of gene lists of 1.1 fold change from 1.0 fold change from both up regulating and

down regulating lists. These two selected gene lists of non-changing transcripts were combined using a Venn diagram (Figure 51C) and saved as '**Day4toDay6NoChange**'.

The gene lists created from the pairwise comparisons were systematically combined together to create a complete time course (Day 0 to Day 4 to Day 6 to Day 8) using the Venn diagram tool to create all possible combinations as shown in Figure 50 (for example '**Day 0 to Day 4 – Up-Regulation**' combined with '**Day 4 to Day 6 – Down-Regulation**' combined with '**Day 6 to Day 8 – No-Change**'). The number of transcripts present for each combination of Figure 3 were tallied up and sorted, with the highest number of transcripts at the top and the lowest at the bottom (Figure 52). As microarray samples Day 4, Day 6, and Day 8 were isolated from sorted KDR+ and CD31+ cells which are known markers for early haematopoietic cells, a few hypotheses may be speculated. It was hypothesised that due to this enriched population (KDR+CD31+), a large proportion of genes could be haematopoietic-related genes and therefore a significant increase in expression from Day 0 to Day 4 may be observed. However equally, as these cell types are hypothesised to be haematopoietic-like, this would suggest that a similarly large proportion of genes would decrease in expression, for example those genes involved in the maintenance of pluripotency and self-renewal, from Day 0 to Day 4.

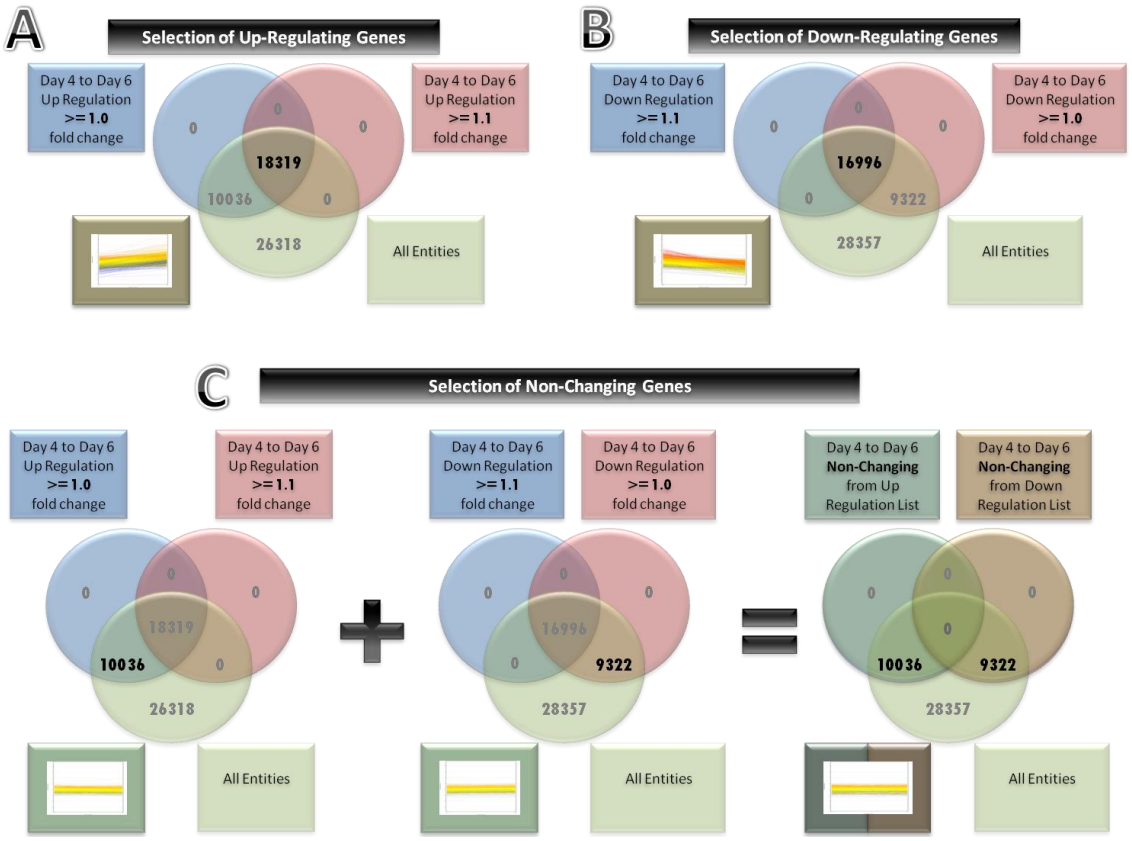


Figure 51: Selection of Up-Regulating, Down-Regulating, and Non-Changing Transcripts

Figure 51A: For the selection of transcripts which are only up-regulated from Day 4 to Day 6, the following three gene lists were combined into a Venn diagram – '**Day 4 to Day 6 Up regulation >=1.0 fold change**', '**Day 4 to Day 6 Up regulation >=1.1 fold change**', and '**All Entities**'. The intersection of all three combined gene lists (in **bold**) contain genes which are up-regulated by 1.1 fold change or more but excludes genes that are up-regulated by 1.1 fold or less. Figure 51B: Using the same Venn diagram method, the selection of genes which are only down-regulated from Day 4 to Day 6, the following three gene lists were combined into a Venn diagram – '**Day 4 to Day 6 Down regulation >=1.0 fold change**', '**Day 4 to Day 6 Down regulation >=1.1 fold change**', and '**All Entities**'. The intersection of all three combined gene lists (in bold) contain genes which are down-regulated by 1.1 fold change or more but not genes that are down-regulated by 1.1 fold or less. Figure 51C: Utilising the same Venn diagram from the selection of genes which up-regulate and down-regulate between the samples, the intersection between the gene lists of '**Day 4 to Day 6 Up regulation >= 1.0 fold change**' and '**All Entities**' (in bold) were used. This was combined with the gene list obtained from the intersection of '**Day 4 to Day 6 Down regulation >= 1.0 fold change**' and '**All Entities**' within a Venn diagram. This means that genes with a fold change difference of 1.1 fold change or less are said to be non-changing between the two days.

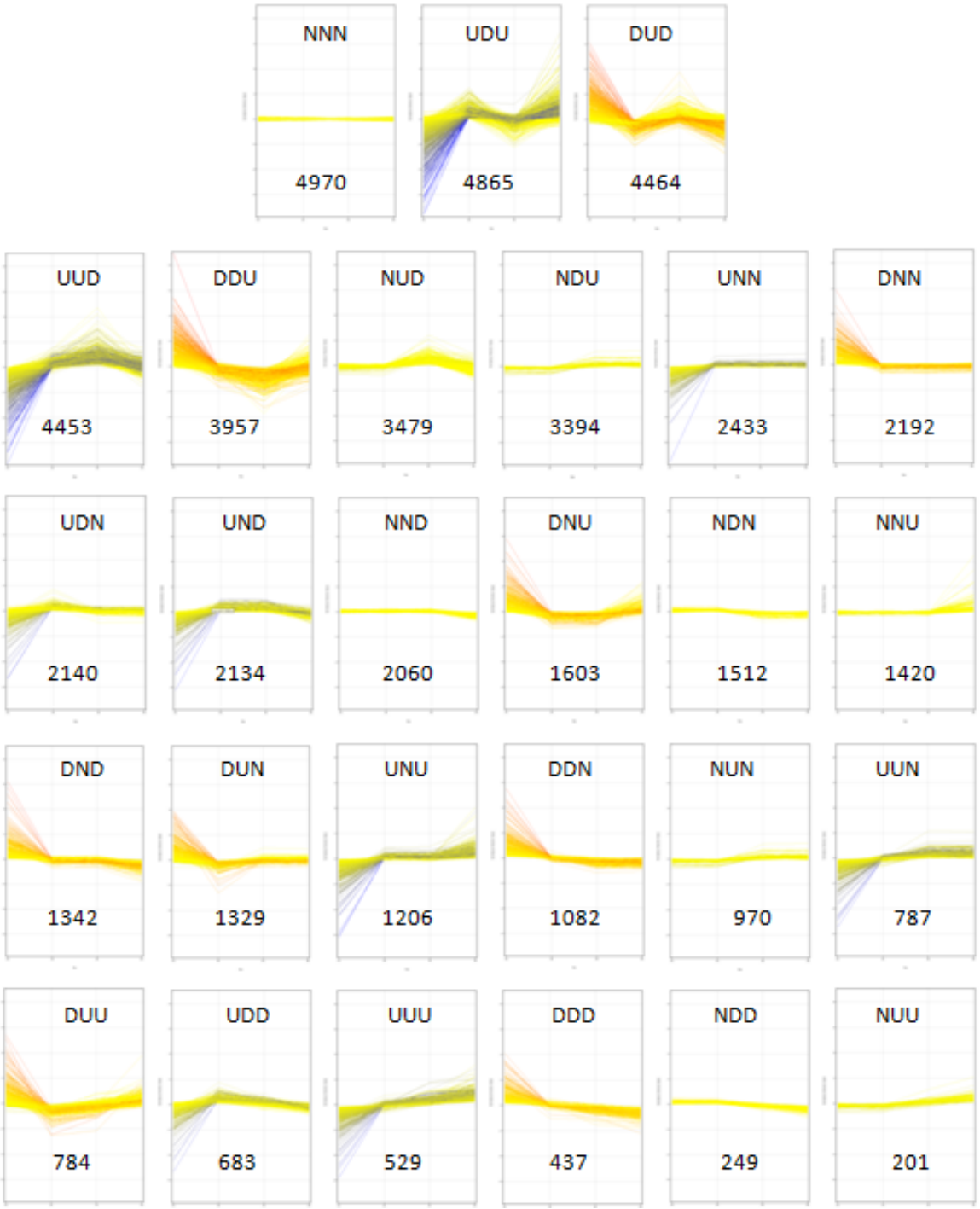


Figure 52: Microarray Expression Profiles - Data Combinations Clustering

Within this figure the 27 expression profiles plots show the distribution of the 54,675 transcripts across the 4 sample microarray experiment. The expression profiles which contain the highest number of transcripts appear at the top and with the lowest at the bottom. Within each expression profile plot there is a three letter code (U=Up-regulation, D=Down-regulation, N=No-change) which describes the expression pattern of the transcripts between samples 1 and 2, samples 2 and 3, and samples 3 and 4. For example, NNN means that there is no-change in expression from Sample 1 to Sample 2, there is no-change in expression from Sample 2 to Sample 3, and there is no-change in expression from Sample 3 to Sample 4.

A quick visual inspection of the expression plots in Figure 52 shows the presence of a wide spectrum of expression patterns and the microarray experiment appears to contain transcripts of all 27 possible profiles. The expression profiles which contain the greatest number of transcripts are shown in the top row of Figure 52 and these are '**No-Change, No-Change, No-Change (NNN)**' (4970 transcripts), '**Up-Regulation, Down-Regulation, Up-Regulation (UDU)**' (4865 transcripts), '**Down-Regulation, Up-Regulation, Down-Regulation (DUD)**' (4465 transcripts). These top three expression combinations which contain the largest number of transcripts were further analysed using a network / pathway and gene ontology tool called 'Metacore' [279] in order to discover common threads within the various gene expression profiles.

3.3.2.2 Metacore Analysis – Protocol

GeneGo - Metacore is a comprehensive, up-to-date, manually curated, knowledge database designed for pathway analysis and data mining based on gene expression [279]. It was used within this experiment to analyse gene lists produced from GeneSpring through the generation of networks and the identification of functional processes and cellular pathways based on statistical relevance. Metacore is a web-based tool and therefore all of the complicated processing is performed on powerful remote servers so the end user's computing power not a limiting factor when using this tool. Further information regarding Metacore including a free trial of the software may accessed through the GeneGo website [279].

Before Metacore analysis may proceed, GeneSpring-created genes lists were first exported from GeneSpring by the selection of the required interpretation (to identify the samples required in the gene list), then right clicking on the gene list itself (within 'Entities') and selecting 'Export List'. Within export options, only 'Raw Signal Values' were selected, and the gene list was exported as a standard .txt file. The exported gene list .txt file was opened within Microsoft Office Excel 2007 and the first 50 rows of information regarding the array were deleted until the first row began with heading 'Probe Set ID', and the file was saved as .xls format ready to be imported into Metacore for analysis. Due to its simple universal parsing import tool, Metacore is able to accept gene data from a wide range of data formats including Affymetrix, Agilent, single nucleotide polymorphism (SNP) arrays, small interfering ribonucleic acid (siRNA) and proteomics data. Metacore as previous described is web-based and was

accessed at the following location – https://portal.genego.com [282] with the necessary login credentials. The 'Data Analysis Wizard (General Parser)' was accessed by clicking on 'Upload Experiments with Gene or Protein IDs', and the .xls file containing the gene list was located by clicking on the 'Browse' button. The 'Next' button was selected and after the file upload was completed, the column types in the uploaded file were verified so that the first column was labelled as 'Affymetrix tag IDs' and the remaining columns were labelled as 'Intensity'. The 'Next' button was selected and 'Homo sapiens' was chosen as the species of interest. The 'Experiment name' was selected and the data was activated ready to be used in subsequent analyses.

The gene ontology processes within the gene list were retrieved and analysed by the selection of the 'Analyse Data' link within the 'Start Page', and through the sub-menu, the 'Gene Ontology Processes' link was selected. Metacore's public gene ontology processes were retrieved from the EMBL-EBI Gene Ontology database [283] and the output was sorted in order of process specificity with the most general of processes at the beginning and the more specialised processes at the end. The numbers of genes associated with any one process were recorded (within the brackets) from the top gene ontology processes and analysed through the creation of networks. The top 50 most significant process networks and pathway maps were retrieved through the same method and these were selected for within the 'Analyse Data' link on the 'Start Page'.

3.3.2.3 Metacore Analysis – General Expression Profiles Results

The results from the general expression profile analysis (Figure 52) show that transcripts which exhibit no change '**NNN**' throughout the experimental time course account for 9.1% of the total number of transcripts (54,675 transcripts) within the microarray. A gene ontology analysis within Metacore of these transcripts reveal roles in a number of general signalling processes, which include immune responses and cyclase activity regulation (Figure 53), and primarily composing of genes which are required for the basic function of the cell. Therefore, stably expressed genes would be expected to appear in this list and a quick search reveals common housekeeping genes *GAPDH* and *Beta Actin* (Appendix: Table 38). A more thorough analysis of this list together with the application of t-test p-values may reduce this list of stably expressed transcripts down to a size whereby novel house-keeping genes or genes which are stably expressed during haematopoietic differentiation may be more reliably selected and used in for the normalisation of future haematopoietic gene expression studies.

The expression profile '**Up-Regulation, Down-Regulation, Up-Regulation (UDU)**' contain the second largest number of transcripts (8.9%), and the roles of these include various metabolic processes, cell proliferation, heart development, and muscle cell differentiation (Figure 53). The presence of transcripts which are known to function in heart and muscle development suggests that the initial up-regulation from Day 0 to Day 4 may be due to mesoderm induction by the BMP4 cytokine, and the subsequent down-regulation from Day 4 to Day 6 may be due to the effect of the addition of cytokine VEGF during this period as mesodermal genes are down-regulated marking the beginning of haematopoietic transcript up-regulation later in development. A study by Yang *et al.* (2008) [284] describes the successful derivation of human cardiovascular progenitors from HESCs using a media cocktail (containing cytokines BMP4, bFGF, Activin A, DKK1, and VEGF) which contains the basic cytokines (BMP4, bFGF, and VEGF) present in the optimised media. The addition of one or more key cytokines (into the existing optimised media) appears to be required for the up-regulation / activation of haematopoietic-specific pathways and for the down-regulation / inhibition of cardiac and muscle pathways in order to obtain a more defined haematopoietic population.

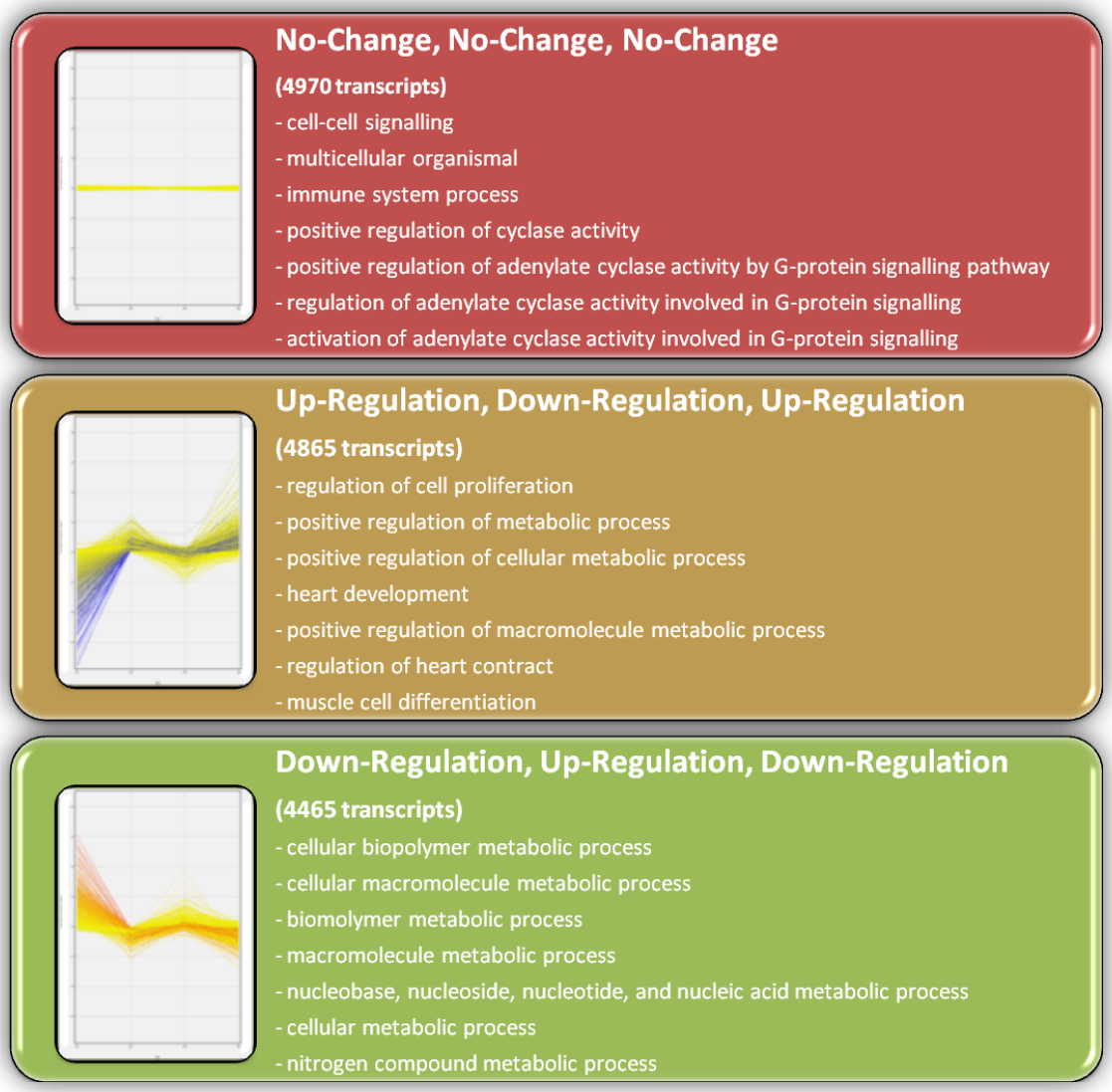


Figure 53: Top Three Microarray Expression Profiles – Gene Ontology Analysis

The three general expression profiles which contain the highest number of transcripts are as follows, NNN (No-Change in expression from Day 0 to Day 4, No-Change in expression from Day 4 to Day 6, and No-Change in expression from Day 6 to Day 8), UDU (Up-Regulation in expression from Day 0 to Day 4, Down-Regulation in expression from Day 4 to Day 6, and Up-Regulation in expression from Day 6 to Day 8) and DUD (Down-Regulation in expression from Day 0 to Day 4, Up-Regulation in expression from Day 4 to Day 6, and Down-Regulation in expression from Day 6 to Day 8). The gene ontology output for the transcripts with an expression profile of NNN appears to suggest that the majority of these are involved in general cell signalling mechanisms and in particular in the positive regulation of adenylate cyclase activity. The results for transcripts with a UDU profile appear to be involved in cell development and the regulation of metabolic processes, with a specific focus on mesoderm differentiation (heart development and muscle cell differentiation). However, transcripts with an expression profile of DUD appear to be involved almost entirely with general metabolic processes with no indication of any specific involvement in development or differentiation.

The third largest expression profile (8.1%) belong to '**Down-Regulation, Up-Regulation, Down-Regulation (DUD)**' whereby these transcripts are involved in many various metabolic processes including cellular biopolymer, cellular macromolecule, nucleobase, nucleoside, nucleotide, nucleic acid, and nitrogen compound (Figure 53). The initial thought to was to

suggest that these transcripts were related to embryonic stem cell genes which are involved in the maintenance of pluripotency and self-renewal due to the rapid down-regulation seen from Day 0 to Day 4. However a search of two classic embryonic stems cell markers *OCT4* and *NANOG* returned a negative result with these genes located within the expression profile of '**Down-Regulation, No-Change, No-Change** (**DNN**; 2192 transcripts; Figure 52)'. It appears that the genes within this list are largely assorted metabolic-related transcripts playing general roles in cellular processes with no particular connection to development.

The results of the microarray expression profile screen (Figure 52) confirm that the largest cluster is composed of non-changing genes with roles in general cell regulation and signalling processes. This is followed by the second largest cluster of genes which initially up-regulate (Day 0 to Day 4) before down regulating (Day 4 to Day 6) and up-regulating (Day 6 to Day 8) again. It was hypothesised that it is within this cluster where haematopoietic-related genes may be present, primarily because the RNA used for analysis were isolated from flow cytometry sorted for haematopoietic / endothelial progenitors markers (KDR+CD31+) on haematopoietic differentiated HESCs, and hence it would be expected that haematopoietic genes would be present in one of the major expression profiles (Figure 53). If this hypothesis is true, it appears to question the suitability of the use of KDR+CD31+ cells, which may not be truly representative of pure haematopoietic progenitors and reinforce the argument that a larger number of markers are required in the selection of highly enriched haemangioblast cells.

3.3.2.4 Haematopoietic Expression Profiles

A rational approach in the discovery of novel transcripts that play an important role in early haematopoietic differentiation from HESCs may be to observe the behaviour of characterised haematopoietic genes. The gene expression profiles of known haematopoietic genes (*CD31, CD34, KDR, CD117,* and *CD144*) may provide clues to the discovery of novel haematopoietic-related genes. The gene names of haematopoietic genes were searched for using the 'Search Entities' function within GeneSpring and the corresponding results were saved and their expression was compared as shown in Figure 54.

The early haematopoietic progenitor genes as previously described are *CD31, CD34, KDR, CD117,* and *CD144* and as one would expect, their expression profiles appear in general, visually very similar to each other but with the exception of *CD117* (Figure 54). In comparison with the 27 general expression profile combinations from Figure 52, the expression profiles of the haematopoietic genes matched up to a number of these generalised profiles including

UNU (**Up-regulation, No-change, Up-regulation;** 1,206 transcripts), **UNN** (**Up-regulation, No-change, No-change;** 2,433 transcripts), **UDU** (**Up-regulation, Down-regulation, Up-regulation;** 4,865 transcripts), and **UUU** (**Up-regulation, Up-regulation, Up-regulation;** 529 transcripts) (Figure 54). It stands to reason that other haematopoietic-related genes may also follow similar profiles to these characterised haematopoietic genes and a starting point would be to analyse these gene lists belonging to UNU, UNN, UDU, and UUU, combining to give 9,033 potential candidates (representing 83% reducing from the starting 54,675).

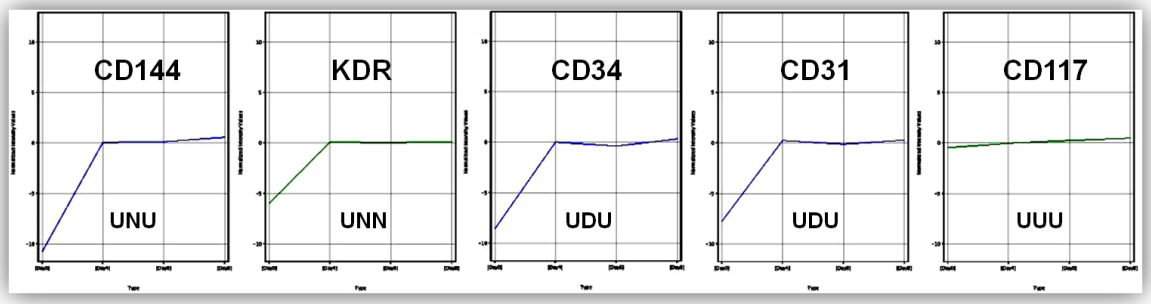


Figure 54: Gene Expression Profile Plots of Haematopoietic Genes

This figure contains microarray expression profile plots of CD144, KDR, CD34, CD31, and CD117 which appear to be quite varied. There are 4 expression profiles shown above and these are **UNU** (**Up-regulation, No-change, Up-regulation**) which is similar to 1206 other transcripts in the microarray, **UNN** (**Up-regulation, No-change, No-change**) which is similar to 2433, **UDU** (**Up-regulation, Down-regulation, Up-regulation**) which is similar to 4865 and **UUU** (**Up-regulation, Up-regulation, Up-regulation**) which is similar to 529. Although the majority of haematopoietic genes match up with different generalised expression profiles, a visual comparison however reveals clear similarities showing a large up-regulation from sample 1 to sample 2, and all subsequent changes in expression between later samples although varied are very small.

As genes of similar function are more likely to be expressed at similar levels and at similar stages of development, haematopoietic genes may be used in the search of other genes with a similar gene expression profile. This search was performed using a tool within GeneSpring called 'Find Similar Entities', which searches through the expression data and attempts to find all transcripts with similar expression profiles. The starting 'Entity List' (gene list from which the search will take place) was '**All Entities**', and the 'Interpretation' (Number of Days / Samples) was set for 'All Samples'. Haematopoietic genes (*CD31, CD34, KDR, CD117,* and *CD144*) were chosen for the 'Choose Query Entity', and the similarity metric was set to 'Spearmann' (as it was found that this correlation metric incorporated the majority of genes from Pearson and Euclidean searches), but the precise algorithm used by GeneSpring within 'Find Similar Entities' was unfortunately not provided. The results from the similarity searches of all haematopoietic genes were combined into a master gene list called **HaemSimList** using a Venn diagram by first combining results from CD31, CD34, KDR into a gene list and then

incorporating the results from CD117 and CD144 into the same list. The results of the haematopoietic similarity searches are shown in Figure 55 and the combination of all similar haematopoietic transcripts into a single gene list (**HaemSimList**) contain a total of 1,186 transcripts.

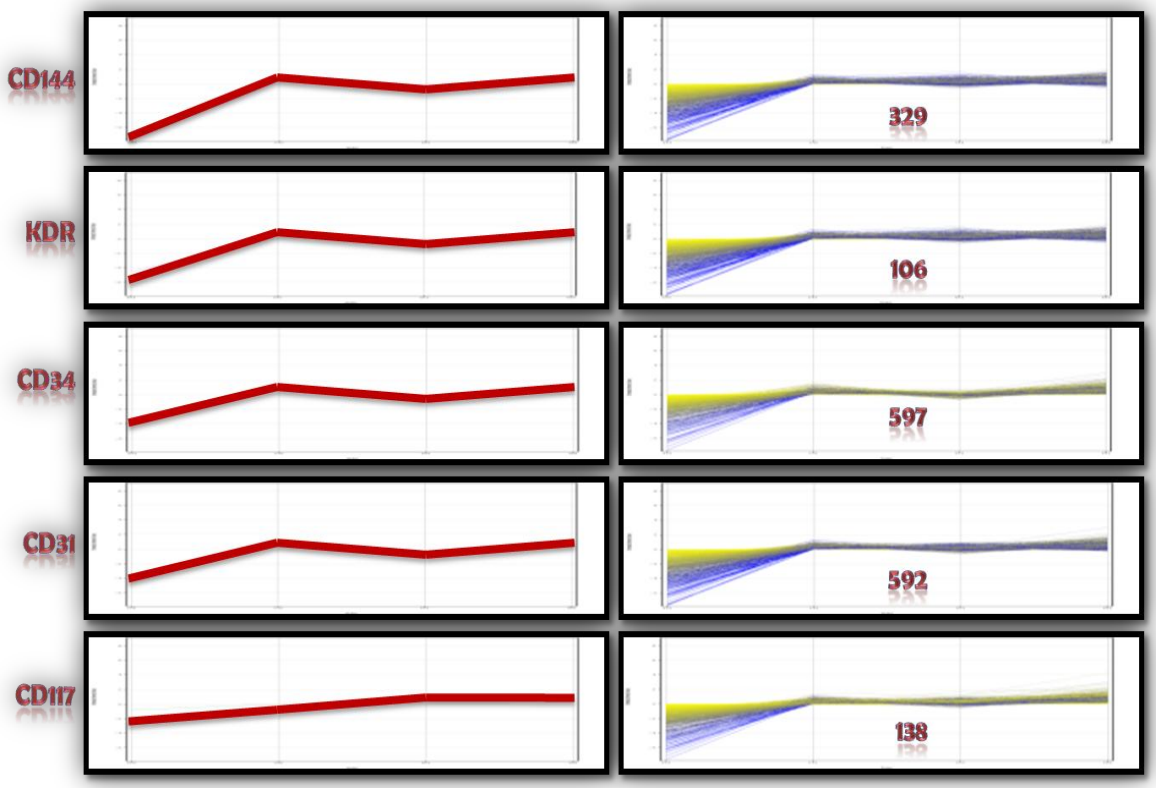


Figure 55: Expression Profile Plots - Haematopoietic Similar Genes

The gene expression profile of the haematopoietic genes (CD144, KDR, CD34, CD31, CD117) (left column) were used as a template for the selection of genes which closely follow a similar expression profile (>95% similarity) (right column) and these were combined together to give a total of 1186 transcripts. These gene lists of similar genes were selected from all transcripts of the microarray with no fold changes or p-value thresholds applied.

3.3.3 Microarray Analysis of Day 0 to Day 4

One of the important aims of a microarray analysis is to identify the useful data from the not so useful, and already a few methods have been described in the search for novel haematopoietic transcripts. These ranges from trying to hypothesis the presence of these novel transcripts within generalised profiles (Figure 52), to suggesting that these may have similar expression profiles to known haematopoietic genes (Figure 55). Although both methods do have their own merits (and as well as flaws), it was however decided through the examination of expression profiles of haematopoietic genes, that the only consensus between these transcripts lie in the large increase in expression (with the exception of *CD117*) from day 0 to day 4 (Figure 54). The inclusion of transcripts with increases in gene expression between day 0 and day 4 in combination with the haematopoietic similar transcripts would appear to be the optimal approach in reducing the gene list from the starting 54,675 transcripts (including multiple splice variants and fragments of genes).

3.3.3.1 Day 0 to Day 4 – Up-Regulation Gene List

The 'Fold Change' tool was used in the selection of genes that increased in expression from Day 0 to Day 4 and within this tool, '**All Entities**' was chosen for the 'Entity List' and 'Day 0 vs. Day 4' was chosen for the 'Interpretation'. As the interpretation only contains samples Day 0 and Day 4, the 'Pairs of Conditions' within 'Pairs Options' which are to be fold change interrogated were set to 'Condition 1 = [0]' and 'Condition 2 = [4]'. The threshold value of 2 for fold change which was set by GeneSpring as default was changed to 1 within the 'Change cut-off' option, and within the 'Fold Change' tab the fold change was sorted by direction (up / down) by clicking on the column header 'Regulation([0] vs [4])'. All probe set IDs which displayed an 'up' in the 'Regulation([0] vs [4])' column were selected and saved as '**Day0toDay4Up1fold**) by clicking on the 'Save custom list' button. A gene list (**Day0toDay4Up1fold**) containing all transcripts which increased in expression from day 0 to day 4 contain a mammoth 31,091 transcripts (Figure 54), but not all of these transcripts are thought to be biologically important and therefore statically significant transcripts with a large enough expression difference were selected for by fold change and p-value.

3.3.3.2 Differential Gene Selection – Fold Change and P-Values

An important analytical aspect of a microarray study is the identification and selection of differentially expressed genes, what this means and how this may be achieved. In the strictest statistical definition, a gene is differentially expressed if its expression level changes between samples of interest regardless of the size of the change. However, a biological definition in the strictest remains open for debate, as a gene in this scientific context is likely to be considered differentially expressed if the change is by a worthwhile amount. Therefore, the question remains, so what size of gene expression change is a worthwhile significant amount? Early array studies from 1990s had simply opted in the selection of differentially expressed gene purely in terms of fold change alone and a 2-fold change was typically considered worthwhile, but the use of this method by itself has drawbacks which include noise, variability and replication problems [285]. At the turn of the century, a number of new methods were introduced using p-values as a measure of significance produced by empirical Bayes and other statistical means, by which the variances were adjusted towards a global estimate.. However, these methods alone also suffered from intensity-level dependent false positives, and tend to nominate genes as statically significant even if their fold changes were arbitrarily small. A possible solution was to use the p-value information in combination with fold change, and a number of studies have shown this approach provides improved numbers of biologically meaningful genes and agreements between arrays than using each separately [285-287].

This combined approach of utilising fold change in combination of p-values in the identification of significantly expressed transcripts appeared to be the optimal solution. However, the question remains of the size of the fold change and p-value that should be applied within this experiment. A number of published microarray papers that use this combined (fold change and p-value) approach appear to set their own limits, with fold changes set low as 1.3 in one experiment [286], and as high as 4 in another [287]. The same variation was also seen with p-values, with a study setting this as high as 0.2 [286], and another setting as low as 0.05 [288]. However, this variation may simply be dependent on the individual microarray experiments, and in particular the samples being compared and analysed. The variability between samples appears to be an important factor, for example, with samples of low variability, statistical significant transcripts may have a low fold change, but between samples of high variability, a higher fold change may be required for the equivalent significance. This demonstrates that the application of a fixed fold change for one set of samples may not be suitable for those of another set, suggesting that perhaps a trial and error approach may be required.

3.3.3.3 Day 0 to Day 4 – Up-Regulation – 2 Fold Change and 0.5 P-Value Gene List

To decide which size of fold change would best suit the data, a number of fold changes (1.0, 1.3, 2.0, 4.0, 8.0) were applied for the samples of Day 0 and Day 4, and the number of genes present within each cut-off were recorded (Figure 56). Gene lists of transcripts which are contained within the various fold change thresholds were produced using the 'Fold Change' tool within Gene Spring as previously described. The analysis of fold change thresholds between samples Day 0 and Day 4 reveals that the majority of transcripts within the microarray experiment do not increase by a large magnitude with over 92% of the transcripts (that increase in expression) increasing by less than 2 fold, and the effect of the application of a 2 fold change threshold equates to a greater than 10 fold reduction in number of transcripts (Figure 56). An increase to a 4 fold change results in a further 3 fold reduction in this number and an equally magnitude of reduction was seen with an additional increase to 8 fold change.

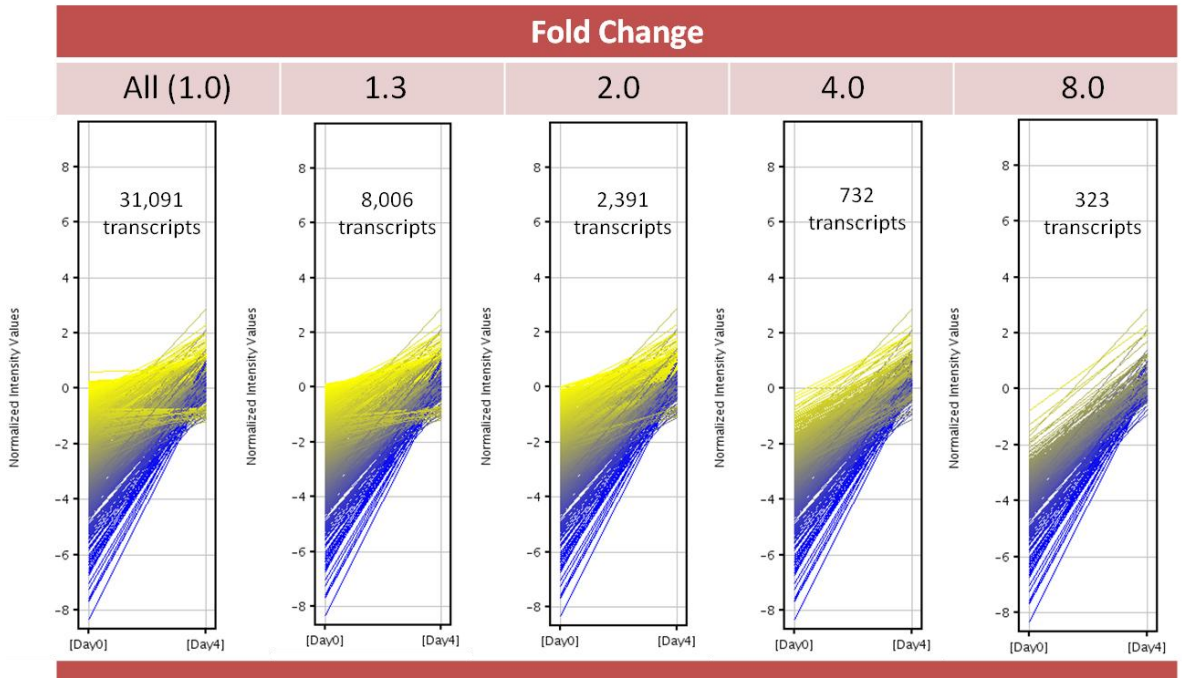


Figure 56: Number of Up-Regulated Genes at Different Fold Changes between Day 0 and Day 4

A variety of fold changes from 1.0 to 8.0 were applied to the transcripts between Day 0 and Day 4 and the total numbers of transcripts which met within each threshold were displayed in the corresponding expression profile plots. The results show that the majority of the transcripts in this experiment do not change by a large amount and less than 8% increase by greater than 2 fold.

To assist in the determination of a suitable threshold by which the fold change may be set, the identification and analysis of the fold change levels of a set of known haematopoietic genes (*CD31, CD34, KDR, CD144,* and *CD117*) may assist to ascertain an optimal level of fold change by which to capture novel haematopoietic genes within as small a gene list as possible. If the comparisons of fold changes between the haematopoietic genes reveal consensual levels of increases of expression between Day 0 and Day 4, then this would greatly assist in the placement of the fold change threshold. These haematopoietic genes were identified and selected within Gene Spring using the 'Entities' tool from the 'Search' menu, and the corresponding data values for each of the haematopoietic transcripts were retrieved from the 'Entity Inspector' which were accessed by double clicking on the transcript in the profile plot. The fold change between samples Day 0 and Day 4 of each haematopoietic transcript were calculated using averaged data values (from two sample replicates).

The results of the fold changes of the haematopoietic genes displayed in Figure 57 shows a spectrum of fold change levels ranging from a low of 1.3 (*CD117*) to a high of 342 (*CD144*) fold change. With the exception of *CD117*, the rest of the haematopoietic genes all have fold change level which are greater than 8, and fold change implementation of this magnitude on the microarray data results in a gene list of just 323 genes (Figure 56) which is a more manageable size for further analysis. However the fold change result of the *CD117* cannot simply be ignored and the implementation of an 8 fold change threshold will exclude potential haematopoietic candidates such as that of *CD117* which plays an important role in haematopoietic development as previously described. A possible alternative would be to set the fold change threshold to match that of *CD117* at 1.3, but at this threshold the gene list contains over 8,000 transcripts which would make downstream analysis more difficult as the large number of transcripts would make it harder to identify the transcripts which play a role in early haematopoiesis. A more manageable number of genes would be found at the 2 fold change threshold which ensures that any potential haematopoietic genes that increase between 2 and 8 fold from Day 0 to Day 4 would be included in this gene list containing a more modest 2,391 genes (**Day0toDay4Up2fold**). The 2 fold level was applied to the gene list **Day0toDay4Up1fold** using the 'Fold Change' tool within the 'Analysis' workflow menu. The **Day0toDay4Up1fold** gene list was selected as the 'Entity List' and the 'Interpretation' was set as 'Day 0 vs. Day 4', and the fold change value was modified to 2 within the 'Change cut-off' option. By clicking on the 'Fold Change' tab the fold change was sorted by direction (up /

down) using the column header 'Regulation([0] vs [4])'. All probe set IDs which displayed an 'up' in the 'Regulation([0] vs [4])' column were selected and saved as '**Day0toDay4Up2fold**'.

However within this gene list, the haematopoietic gene *CD117* was still missing, a thus the solution was to include all genes which have a similar expression pattern to *CD117* using a spearman similarity metric correlation (p-value >0.95 – 138 transcripts) (Figure 55) within GeneSpring and combine these transcripts with **Day0toDay4Up2fold** (2,931). For consistency, all transcripts with a similar expression pattern to all haematopoietic genes (CD31, CD34, KDR, and CD144; Figure 55) were also selected using the same similarity metric correlation (**HaemSimList;** 1,186). The **Day0toDay4Up2fold** gene list was combined with the gene list **HaemSimList** (which was created by 'Haematopoietic Expression Profiles' protocol) using the 'Venn Diagram' tool and this new gene list was named '**Day0toDay4Up2fold+HaemSimList**' which contain a total of 3,162 transcripts.

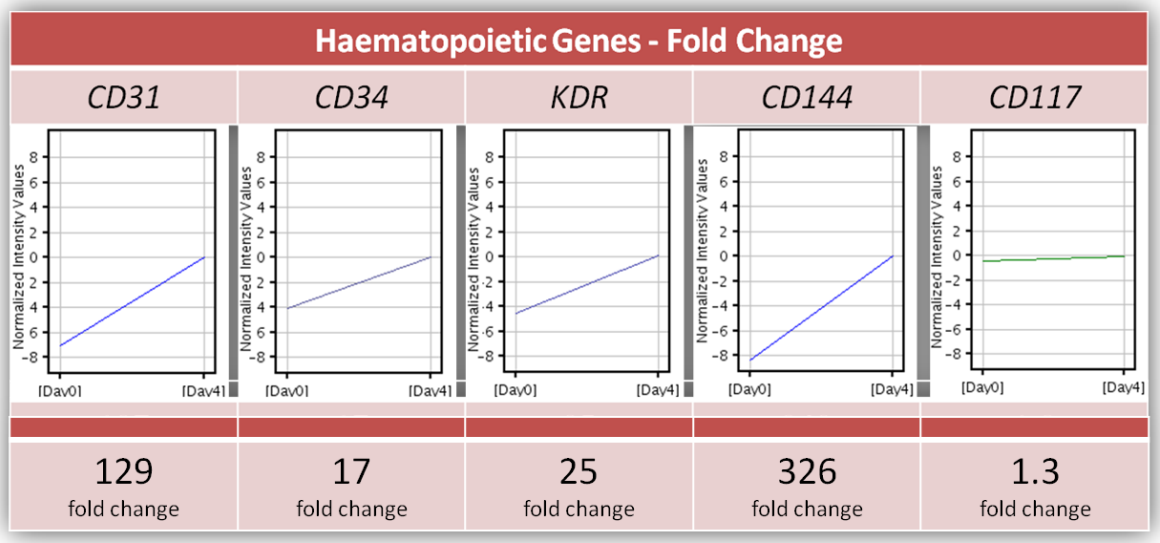


Figure 57: Fold Changes of Haematopoietic Genes from Day 0 to Day 4

The fold changes of the common haematopoietic genes between Day 0 and Day 4 samples were analysed and a wide range of fold changes were calculated. There doesn't appear to be any consensual fold change values and the only conclusion which can be deduced from these results is that any novel haematopoietic transcripts will show an increase in expression from Day 0 to Day 4 but the size of this increase can be anywhere from as low as 1.3 fold (CD117) to as high as 326 fold (CD144).

In the selection of a p-value, the same trial and error approach was adopted for the transcripts within this experiment, with the Student's t-test used in the selection of genes that are statistically different from one condition to another (for example Day 0 against Day 4) by p-value. An asymptotic paired t-test with Benjamini Hochberg False Discovery Rate (FDR) multiple testing corrections was performed on the genes. This was carried out within Gene Spring by clicking on 'Statistical Analysis' within the 'Analysis' workflow, the fold change

limited gene list was selected in 'Entity List', and the 'Interpretation' was set as 'Day 0 vs. Day 4'. Day 0 sample was placed into the 'Control' box and Day 4 was placed into the 'Treatment' box, and the 'TTest paired' statistical test was selected in the 'Select test' box. 'Asymptotic' was selected in the 'p-value Computation' setting and 'Benjamin Hochberg FDR' was ticked for the 'Multiple Testing Correction' box. The calculated p-values of each transcript were listed in chronological order from low to high, and a number of p-value cut-off thresholds (0.05, 0.1, 0.2, 0.5, 0.8) were applied using the 'Change cut-off' button, with the resulting number of transcripts that qualified under each threshold were analysed to determine the optimal p-value threshold (Figure 58).

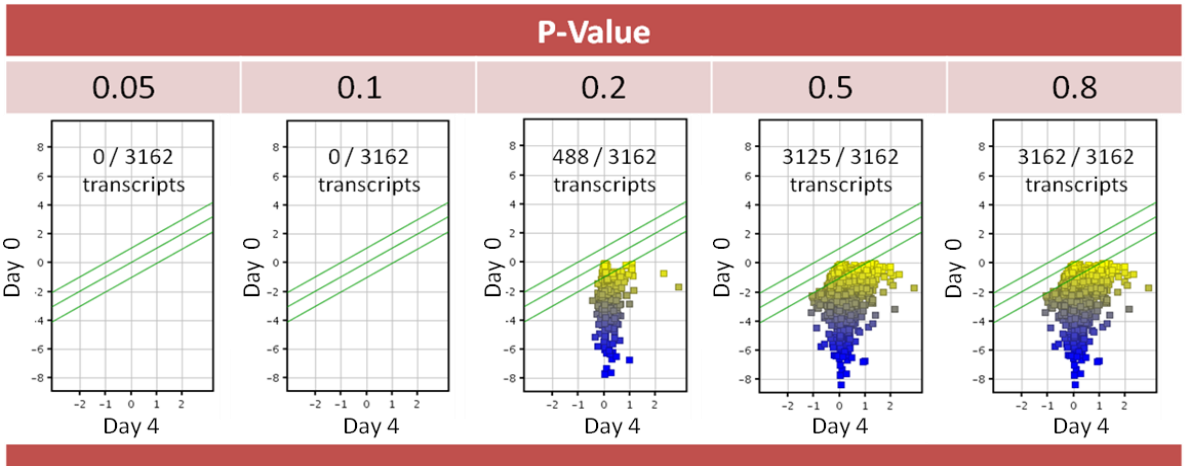


Figure 58: Number of Up-Regulated Genes at Different T-Test P-Values between Day 0 and Day 4

The p-values within this experiment are higher than expected with zero transcripts meeting the 0.1 p-value threshold and only 15% pass the 0.2 threshold, but however at the 0.5 threshold nearly all of the transcripts (99%) are present.

In the selection of the optimal p-value for threshold, a number of values from 0.05 to 0.8 were applied to the **Day0toDay4Up2fold+HaemSim** gene list, and there appeared to be no transcripts with a p-value of 0.1 or lower (Figure 58). At the 0.2 threshold, only 15% of the gene list fulfilled this criterion and more importantly the absence of several important haematopoietic genes including CD144, CD31, and CD117 meant that the 0.5 threshold (which contains 99% of transcripts in the gene list) should be used. Although in this case the inclusion of a p-value cut-off threshold has little effect on the size of the gene list, it should never the less be applied regardless in all experiment to remove transcripts that contain replication inconsistencies whilst maintaining transcripts of biological importance. As previously described, within Gene Spring an asymptotic paired t-test (Day 0 against Day 4) with Benjamini Hochberg False Discovery Rate (FDR) multiple testing corrections was performed on the **Day0toDay4Up2fold+HaemSimList** gene list. The p-value cut-off thresholds of 0.5 was applied

using the 'Change cut-off' button and a new gene list of all transcript which met this p-value criterion were created and saved as '**Day0toDay4Up2fold+HaemSimList+p0.5**' (APPENDIX A: Table 39).

The flow chart within Figure 59 illustrates the process through which the transcripts within the gene list '**Day0toDay4Up2fold+HaemSimList+p0.5**' were selected, starting from 54,675 transcripts and finishing with 3,125 transcripts. These selected genes may be important in early haematopoietic development and their known functions and interactions (sought from publications) were analysed by gene ontology, gene maps and process networks within the web-based pathway analysis tool Metacore (using the protocol as previously described).

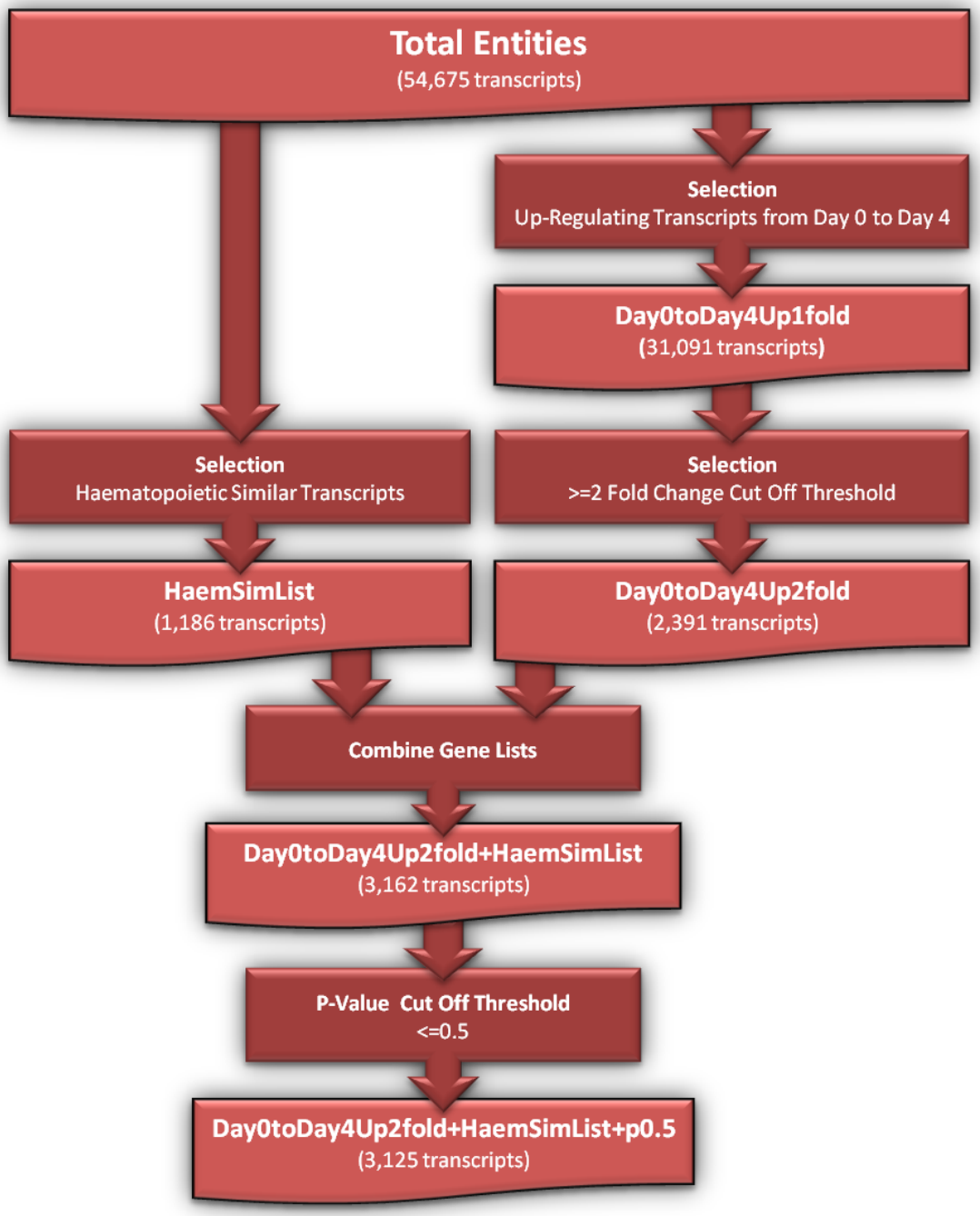


Figure 59: Day 0 to Day 4 Haematopoietic Transcript Gene List Creation Flow Diagram

This flow chart shows the process of selection of the final gene list which contains transcripts that are up-regulated by 2 fold greater in gene expression between Day 0 and Day 4 and combined with haematopoietically similar transcripts and with a p-value of 0.5 or lower. The gene lists which are created for each step are shown in pink and the processes by which the various changes (fold change and p-value) were applied are shown in red.

3.3.3.4 Day 0 to Day 4 analysis reveals SOST as a potential haematopoietic candidate

A simple ranking analysis of the **Day0toDay4Up2fold+HaemSim+p0.5** gene list (APPENDIX: Table 39) reveals that the top ten gene expression fold changing genes between Day 0 and Day 4 are *CD144 (CDH5), ERG, MMRN2, SAMSN1, CD93, SOST, RASGRP3, PECAM1, ECSCR*, and *EMCN*. The fold change of these genes range from as low as 120 fold (*EMCN*) to as high as 326 fold (*CD144*) and the majority of these genes within this list have already been described within publish literature with an association with haematopoietic and endothelial cells, including *CD144* (haematopoietic and endothelial cells [145, 216]), *ERG* (definitive haematopoiesis [289]), *MMRN2* (pan-endothelial cells [290]), *SAMSN1* (haematopoietic cells [291]), *CD93* (endothelial, platelets, myeloid cells), *RASGRP3* (endothelial and B-cells [292, 293]), *PECAM1* (haematopoietic and endothelial cells [216, 294]), *EMCN* (haematopoietic stem cells [295]), and *ECSCR* (endothelial cells [296]). The gene *Sclerostin* (*SOST*) was the only transcript within the top ten highest ranked gene list that has not been shown to play a role in haematopoiesis or vasculogenesis. *Sclerostin* encodes a cystine knot-containing glycoprotein, which was identified in sclerosteosis patients, where a loss-of-function mutation in this gene resulted in progressive bone overgrowth [297]. A search for known networks and pathways involving *SOST* within the Metacore database reveal no links, which suggests that this gene may not have been very well investigated. It is therefore of great interest to further investigate the role of this gene (if any) in haematopoietic differentiation, and the reason why its expression is so high (197 fold up-regulation from day 0 to day 4) during this early stage of differentiation.

3.3.3.5 Gene Ontology Analysis reveals Endothelial Beginnings

A Metacore analysis was performed on the **Day0toDay4Up2fold+HaemSim+p0.5** gene list whereby the gene ontology, gene maps, and process networks were identified and ranked in terms of p-value significance. The Metacore results of the gene ontology analysis, as shown in Figure 60, was divided into three primary sections (molecular function, cellular component, and biological process) and each of these sections contain more details which are level specific.

The molecular function contains information regarding 1839 transcripts from the gene list and these are generally associated with binding (protein, calcium ion, cytoskeletal protein,

and growth factor) and other common functions such as enzyme regulation and a number of GTPase-related activities. Within the cellular components section, there appears to be a large number of membrane related transcripts (1032), which may suggest that a proportion of these transcripts may be associated with haematopoiesis, and may include novel cell surface antigen markers. An equally large number of cellular components appear to play a role within the cytoplasm (1016) and specifically contributing to the formation of various vesicle / vacuole / lysosome components suggesting the development of the immune system and haematopoiesis. The biological processes shows a variety of general activity, with cell communication and signal transduction being the two main processes, however there is also a clear link with the onset of haematopoiesis. A number of transcripts are involved in blood vessel and vasculature development (58) and cell-cell adhesion (84) which strongly signifies the beginnings of haematopoiesis, and the inclusion of many immunology processes such as endocytosis (55) and membrane invagination (55) adds support to this hypothesis. These results coincide and support the theory of the haemogenic endothelium as discussed in Chapter 1, whereby haematopoietic cells are formed from this intermediate endothelial progenitor (haemogenic endothelial) [192, 193] which first generate the blood vessels (endothelium) and during development, these bi-potential progenitors begin budding off to become haematopoietic cells.

Within this gene list, a number of gene networks (30) have been characterised that are responsible for blood vessel development, with the top three networks being, ‘MMP-2, TGF-beta receptor type II, TGF-beta receptor type I, TGF-beta 2, Biglycan’; ‘Beta-catenin, c-Jun, MMP-2, FAK1, TCF7L2 (TCF4)’; and ‘FAK1, AKT(PKB), Shc, ERK2 (MAPK1), p120GAP’. The presence of transcripts involved in as much as 30 networks relating to blood development adds confidence to the theory of haemogenic endothelium. It would be of great interest to examine the up-regulating (> 2 fold change and p-value of <0.5) transcripts belonging to the later days of differentiation, for example Day 6 and Day 8, and observe if these transcripts are involved in the differentiation of primitive haematopoietic cell types (for example erythrocytes and megakaryocytes) from the haemogenic endothelium intermediate.

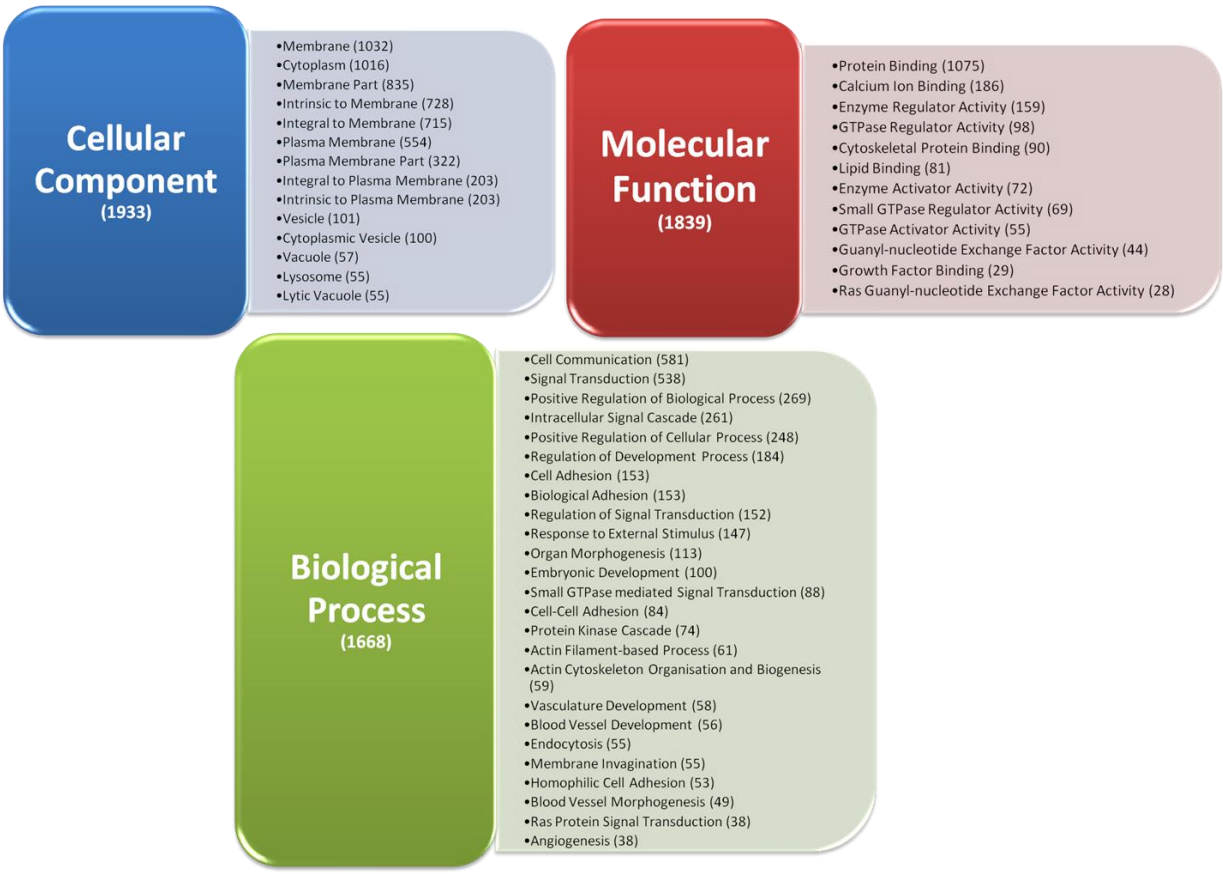


Figure 60: Gene Ontology - Day 0 to Day 4 Up-regulation and Haematopoietic Similarity Gene List

This figure shows the most significant (lowest p-value) hits for the three primary sectors of the gene ontology (molecular function, biological process, cellular component) from the gene list Day0toDay4Up2fold+HaemSim+p0.5 and these have been ranked in order of number of genes (largest number at the top) contained in each. The molecular function suggests that the majority of transcripts in the list are associated with binding (protein/calcium) and other general regulatory functions. The cellular component shows that these transcripts are primarily membrane/cytoplasm based. The biological function suggests a link with haematopoiesis and other related processes including blood vessel development, cell adhesion and angiogenesis.

3.3.3.6 Gene pathways indicates the onset of Erythropoiesis

The vascular developmental pathway and the haematopoiesis pathways (**Figure 61**) were both highly ranked in the analysis and in the discovery of novel haematopoietic transcripts through the production of the gene list (**Day0toDay4Up2fold+HaemSim+p0.5**), it brings confidence to the validity of the experiment when vascular and haematopoietic pathways are identified within this list of genes. Upon closer examination of the individual specific pathways, the vascular development pathway breaks down into the development pathways which include **'**Platelet-Derived Growth Factor (PDGF)**'**, **'**Notch, Endothelin Receptor B (EDNRB)**'**, and **'**Thrombospondin signalling**'**, which are all well established and documented in literature and play important roles in early vasculogenesis [298-300]. The haematopoiesis gene pathway comprises of developmental and signalling pathways that largely comprise of erythropoietin (EPO) as the key factor in the majority of the pathways displayed. EPO is a glycoporotein hormone which is the principal factor in the regulation of erythrocyte formation [301] and its role extends to the safeguarding of circulatory erythrocyte numbers by regulating erythropoiesis during adult [302]. These results strongly imply the onset of erythropoiesis with evidence showing the activation of a large number of EPO-related pathways. These results support studies that have demonstrated that erythrocytes together with megakaryocytes are one of the very first blood lineages to emerge from the developing yolk sac, and that these emerge from a primitive megakaryocyte-erythroid precursor (MEP) [205]. However, other important components of erythropoiesis must also be investigated prior to concluding judgement on this theory, with one important component being the expression level of globin genes over the 8 days of differentiation. Globin genes have been well studied and their role well established, with their expression expected to be selectively up-regulated during erythropoiesis through the generation and formation of heterotetramers with haeme molecules (haemoglobin) which constitute a substantial portion of the functioning erythrocyte [303].

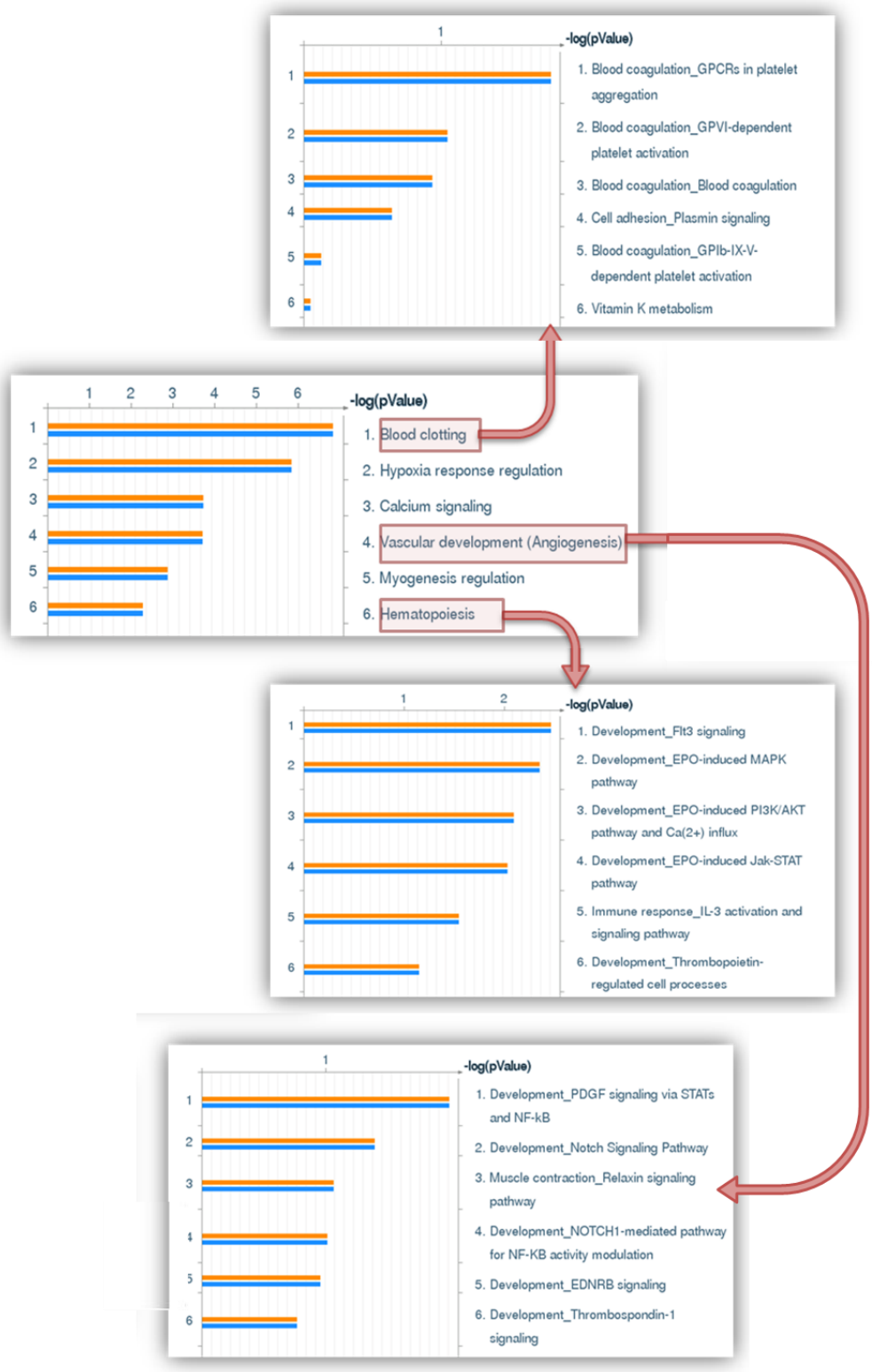


Figure 61: Metacore Gene Maps – Top Ranking Gene Maps from Day 0 to Day 4 Analysis

The gene maps are ranked in terms of p-value with the most significant gene maps at the top and the top six include blood clotting, hypoxia response regulation, calcium signalling, vascular development, myogenesis regulation, and hematopoiesis. The highlighted maps (red boxes) are of particular interest and expansion of these categories reveals the corresponding maps in order of significance. Within blood clotting, the most significant gene map appears to be associated with blood coagulation and GPCRs in platelet aggregation. The top gene map in vascular development is associated with development and in particular PDGF signalling via STATs and Nf-kB. Flt3 (FMS-like Tyrosine Kinase-3) signalling is the most significant gene map within hematopoiesis and plays an important role in cell survival [304], acute myeloid leukemia [305], and haematopoietic differentiation [306].

3.3.3.7 Gene pathway indicates a role for Hypoxia in erythropoiesis

The remaining pathways outlined in Figure 61 of which transcripts within this gene list have roles within are calcium signalling, myogenesis regulation, and hypoxia response regulation. The well-established role in which calcium signalling plays in immune function [307, 308] may offer an explanation as to why genes of this pathway were over-expressed, but again it is the expression of these transcript at such an early stage of differentiation which makes this finding very interesting. Perhaps this is connected with the emergence of megakaryocytes (and mast cells) from megakaryocyte / erythroid progenitors (MEP) during erythropoiesis [309], however this remains to be confirmed with a more thorough investigation of the later days of differentiation (day 6 and day 8). The presence of myogenesis regulation genes again shows that this HESC differentiation combined with the use of markers KDR and CD31 alone are not specific enough in the selection for solely haematopoietic progenitors. However, at this stage of differentiation (day 0 to day4) where BMP4 actively promotes differentiation of HESCs to mesoderm (which in turn gives rise to bone, muscle, connective tissue, cartilage, circulatory system, reproductive system, spleen, urinary system, and haematopoietic system), the presence of genes relating to myogenesis may therefore not be unexpected. The isolation of cells containing a greater number of haematopoietic-specific markers may result in fewer up-regulated genes being linked to myogenesis, and result in a more haematopoietic-specific pathway analysis. In regards to hypoxia response regulation, studies have commented on the importance of low oxygen in the haematopoietic niche during development [310, 311] and this result appears to add support to this concept. Hypoxia genes appears to be up-regulated at this stage and studies have linked this with EPO expression, which is hypoxia-inducible, and is responsible for a number of important roles in a number of roles during development including survival, proliferation, and differentiation of HSCs into MEPs [312].

3.3.4 Microarray Analysis of Day 4 to Day 6 to Day 8

3.3.4.1 Day 4 to Day 6 – Up-Regulation – 2 Fold Change and 0.5 P-Value Gene List

A gene list containing transcripts which show an up-regulation in gene expression from Day 4 to Day 6 by a fold change of 2 or greater with a p-value of 0.5 or smaller were created using the same protocol as previously described in the creation of **Day0toDay4Up2fold+p0.5**. This gene list was named **Day4toDay6Up2fold+p0.5** (APPENDIX: Table 40) and contained just 339 genes (Figure 62), which is approximately 10 fold less than the number of up-regulated genes observed from the first gene list between Day 0 and Day 4 (even without the additional haematopoietic similarities genes). The reason for this large variation in transcript number between these two gene lists may be attributed to a number of reasons, and with the obvious reason being, day 0 contains pluripotent cells that have not undergone differentiation. This difference was discussed earlier in the chapter within the section on the quality control of samples (Figure 42), where undifferentiated day 0 samples was seen to correlate very poorly with the differentiated samples. Day 4, day 6, and day 8 samples have all been exposed to haematopoietic differentiation conditions, and thus one would expect greater gene expression changes between undifferentiated and differentiated than between the differentiated samples. Based on these reasons, assumptions may be inferred between differentiated samples day 4, day 6 and day 8, with gene expression differences between day 4 and day 6 predicted to be greater than differences between day 6 and day 8. This assumption is based on differences in differentiation media composition, where VEGF was added to day 6 and day 8, but not to day 4. This assumption was verified after an up-regulating gene list was produced of day 6 to day 8 with the comparative differences in gene numbers shown in Table 41.

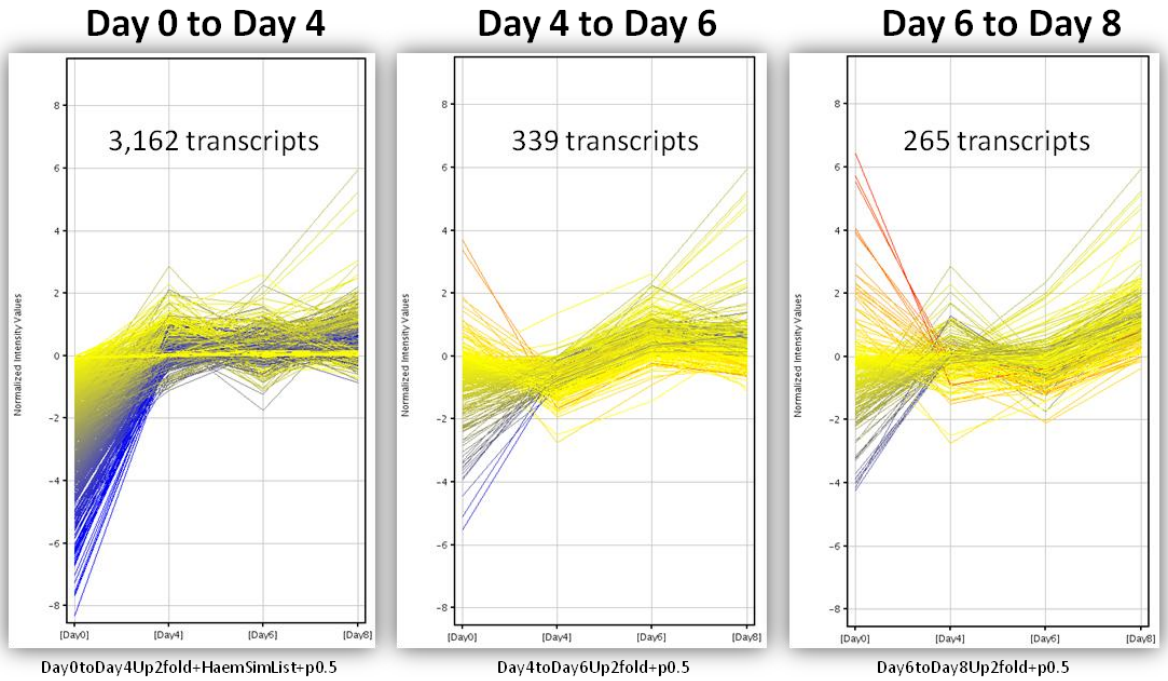


Figure 62: Gene Expression Profiles – Gene Lists of Up-regulating Transcripts

A visual gene expression profile comparison between the three lists up-regulating transcripts from Day 0 to Day 4 (Day0toDay4Up2fold-HaemSimList+p0.5), from Day 4 to Day 6 (Day4toDay6Up2fold+p0.5), and from Day 6 to Day 8 (Day6toDay8Up2fold+p0.5) shows significantly (>10 fold) fewer transcripts that are up-regulated between the later time points (Day 4 to Day 6 to Day 8) than between the first (Day 0 to Day 4). Relative expression is shown on the y-axis and Days (Day 0, Day 4, Day 6, and Day 8) are shown on the x-axis.

3.3.4.2 Day 6 to Day 8 – Up-Regulation – 2 Fold Change and 0.5 P-Value Gene List

A gene list named Day6toDay8Up2fold+p0.5 (APPENDIX: Table 41) was created containing transcripts which show an up-regulation in gene expression from Day 6 to Day 8 by a fold change of 2 or greater with a p-value of 0.5 or smaller using the same protocol as previously described in the creation of **Day0toDay4Up2fold+p0.5**.. This gene list (**Day6toDay8Up2fold+p0.5**) contained just 265 genes (Table 41) which is again smaller than the comparison between Day 4 and Day 6 (339 genes) as previously discussed.

3.3.4.3 Gene map exposes the sequential up-regulation of Erythropoietin in erythropoiesis

A Metacore analysis was performed on the gene lists **Day4toDay6Up2fold+p0.5** and **Day4toDay6Up2fold+p0.5**, these were selected and imported into Metacore as previously described, and the most significant resulting gene pathways were investigated with the results

shown in Figure 63. The main categories of gene processes include cell differentiation, hypoxia response regulation, haematopoiesis, vasodilation, and protein synthesis, and closer inspection revealed a bias towards EPO pathways (Figure 63). These results support those from the analysis of the **Day0toDay4Up2fold+HaemSim+p0.5** gene list, and it appears from these results (Figure 63) that differentiation is more haematopoietic directed between day 4 and day 6.

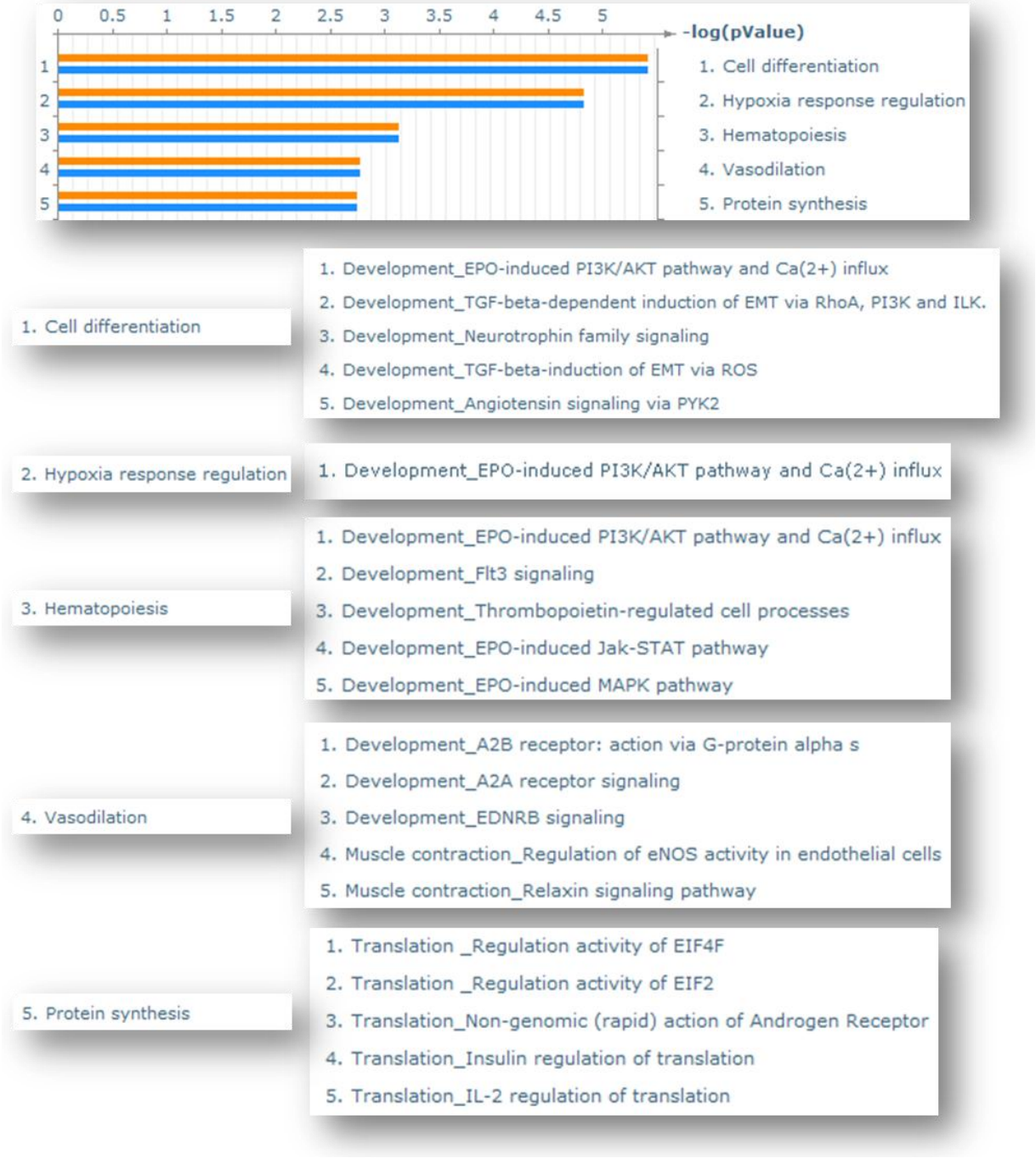


Figure 63**: Metacore Gene Maps – Top Ranking Gene Maps from Day 4 to Day 6 Analysis**

The most significant gene maps are those associated with cell differentiation, hypoxia response regulation, hematopoiesis, vasodilation, and protein synthesis. Within the top three maps (cell differentiation, hypoxia response regulation, hematopoiesis) the top pathway in each of these is 'Developent_EPO-induced PI3K/AKT pathway and Ca2+ influx'.

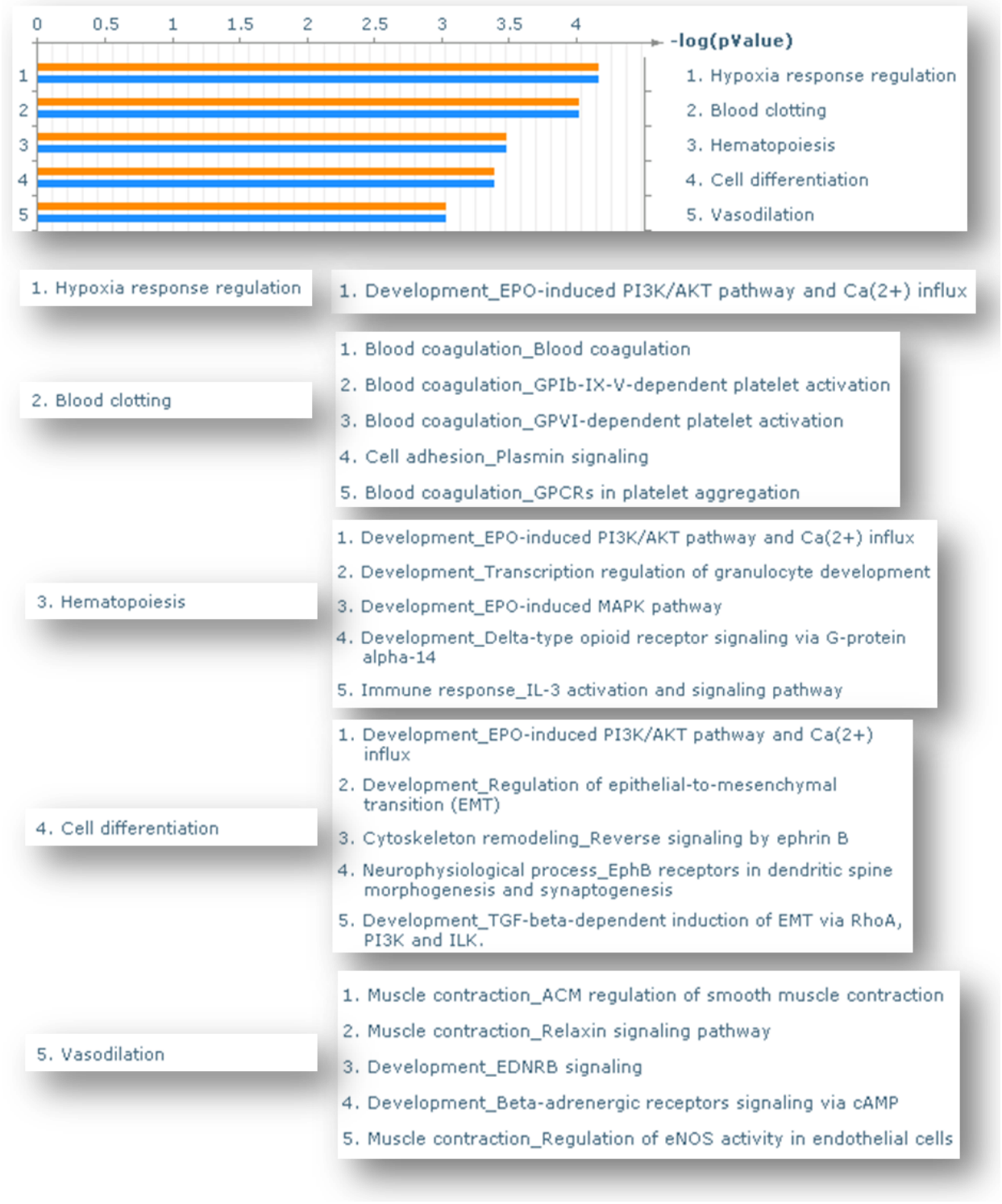


Figure 64: Metacore Gene Maps – Top Ranking Gene Maps from Day 6 to Day 8 Analysis

The most significant gene maps are those associated with hypoxia response regulation, blood clotting, hematopoiesis, cell differentiation, and vasodilation. Within three of the four top maps (hypoxia response regulation, hematopoiesis, and cell differentiation) the top pathway in each of these is 'Developent_EPO-induced PI3K/AKT pathway and Ca2+ influx', with blood clotting map being the exception where blood coagulation pathways are highly upregulated from Day 6 to Day 8.

A detailed analysis of the principle EPO gene pathway 'Development_EPO-induced PI3K/AKT pathway and Ca(2+) influx' (Figure 65) from the Metacore pathway analysis of the gene lists (**Day0toDay4Up2fold+HaemSim+p0.5**, **Day4toDay6Up2fold+p0.5**, and **Day6toDay8Up2fold+p0.5**) shows the sequential up-regulation of genes from start to finish within this pathway.

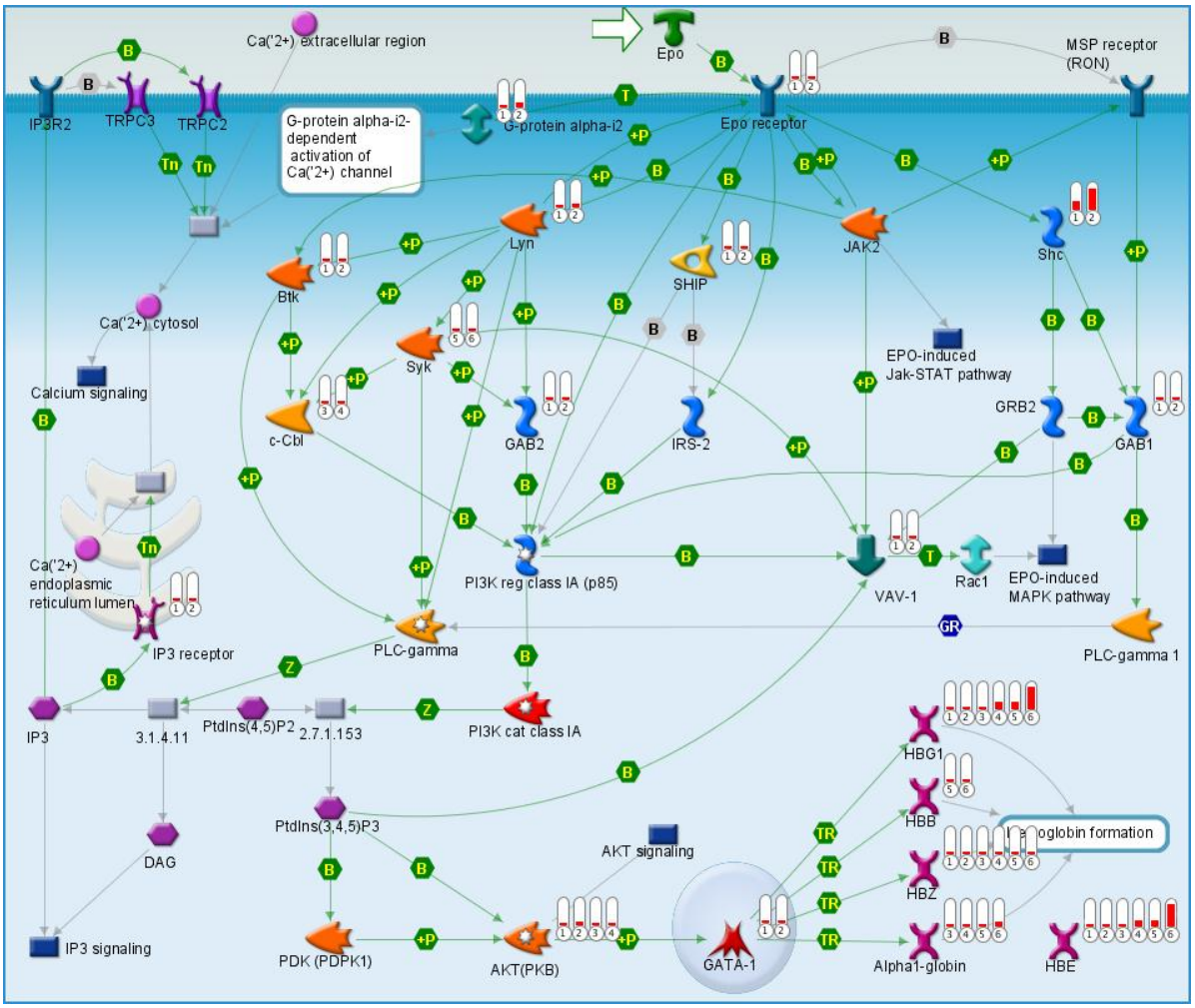


Figure 65: Metacore Gene Map - Development_EPO-induced PI3K/AKT pathway and Ca(2+) influx

This pathway shows the interactions of the Erythropoietin (EPO) pathway beginning with EPO ligand which gives rise to erythropoiesis and haemoglobin formation. The genes containing bars next to them means that they are within the selected gene lists Day0toDay4Up2fold+HaemSim+p0.5, Day4toDay6Up2fold+p0.5, and Day6toDay8Up2fold+p0.5 and the size of the red within each bar describes the intensity of expression. Each gene is connected to another gene through an interaction that is either activation (green line) or inhibition (red line) or unknown (grey line), and the mechanism of interaction could be binding (B), phosphorylation (+P), transformation (T), transport (Tn), catalysis (Z), transport regulation (TR), or group relation (GR). From Day 0 to Day 4, Epo Receptor becomes activated and further down the pathway genes G-protein alpha, Lyn, SHIP, Shc, GAB1, GAB2, Btk, VAV-1, IP3, AKT, GATA-1, HBG1, HBE and HBZ were also up-regulated by greater than 2 fold between the two days. From Day 4 to Day 6, c-Cbl, AKT, HBG1, HBE, HBZ, and Alpha-1-globin were unregulated in expression by 2 fold or greater. The comparison between Day 6 and Day 8 shows a 2 fold up-regulation of Syk, HBG1, HBB, HBE, HBZ, and Alpha-1-globin.

The up-regulation of the EPO receptor between Day 0 and Day 4 of differentiation appears to activate a number of immediate downstream genes / proteins, including guanine

nucleotide binding protein alpha inhibiting activity polypeptide 2 (*G-protein alpha-i2*; a calcium signalling activator [313]), v-yes-1 Yamaguchi sarcoma viral related oncogene homolog (*Lyn*; a phospholipase C-gamma 2 phosphorylator and phophatidylinositol 3-kinase activator [314]), Src homology 2 domain-containing inositol 5'-phosphatase (*SHIP*; a modulator of Ras and inositol signaling [315]), and Src homology 2 domain containing transforming protein 1 (*Shc*; a signaling effector for type I cytokine receptors [316]). During the same stage, (day 0 to day 4), Lyn interacts and phosphorylates Bruton agammaglobulinemia tyrosine kinase (Btk; erythroid modulator essential for efficient phosphorylation of Epo-induced targets [317]) and GRB2-associated binding protein 2 (*GAB2*; EPO-induced signal transducer [318]). Further down the pathway v-akt murine thymoma viral oncogene (*AKT*; Epo regulator of erythroid-cell maturation [319]) phosphorylates GATA binding protein 1 (*GATA-1*; foetal / adult haemoglobin regulator) expression, which in turn interacts with haemoglobin gamma 1 (*HBG1*) and haemoglobin zeta (*HBZ*) at this early stage (Day 0 to Day 4). Microarray data shows that haemoglobin epsilon (*HBE;* inserted into Figure 27) is also expressed at this early stage along with *HBG* and *HBZ*. Haemoglobin studies have shown that the first haemoglobin tetramer complex which appears at the early embryonic stage is composed of HBZ and HBE globins, forming HbE Gower 1 complex [320] (Figure 66). The proceeding haemoglobin tetramer complex which is formed at the overlap of embryonic and foetal stages of development is HbE Gower 2 and is composed of HBA (haemoglobin alpha) and HBE globins (Figure 66). At the foetal stage of development, expression of haemoglobin gamma (*HBG)* in combination with HBA forms the foetal haemoglobin, HbF (Figure 66). Shortly after birth, adult haemoglobin is produced in two forms, HbA2 (which is composed of HBA and HBD (haemoglobin delta) and accounts for 10% of the total haemoglobin in the circulatory system) and HbA (which is composed of HBA and HBB (haemoglobin beta) and it is this form that constitutes the majority of the haemoglobin within the adult.

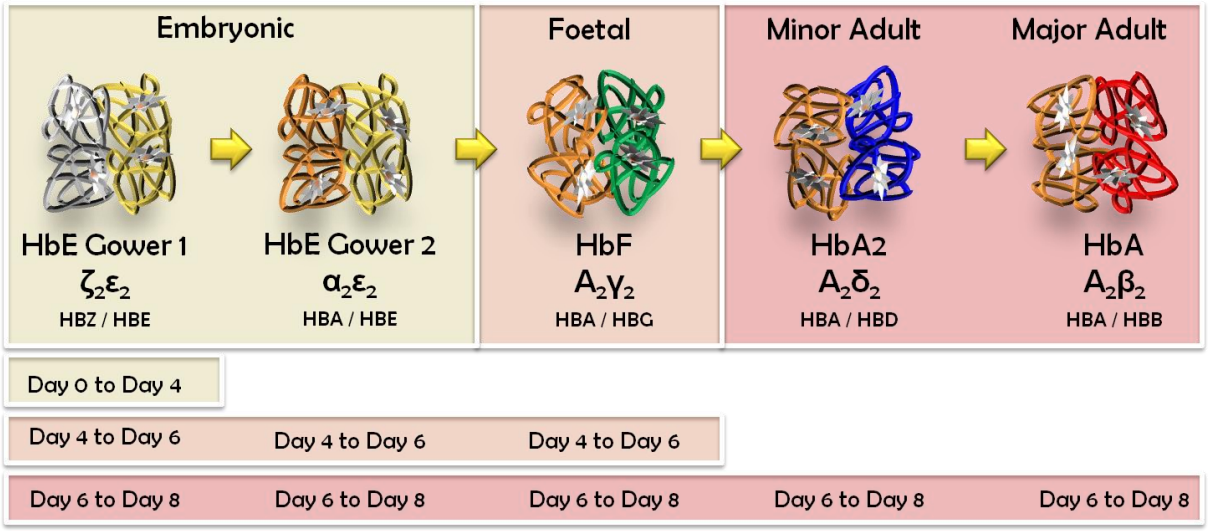


Figure 66: Developmental Stages of the Haemoglobin Complex

The schematic diagram of the human haemoglobin shows the transition of the composition of the complex at various stages of development from HbE Gower 1 and HbE Gower 2 representative of the embryonic form, HbF of the foetal form, HbA2 of the minor adult, and HbA of the major adult. The results from the microarray data are also shown in the diagram indicating the stage of development where each complex may be formed from the gene expression information. HBZ and HBE genes (HbE Gower 1) were expressed by two fold or greater between Day 0 and Day 4, and this expression increased by a further 2 fold from Day 4 to Day 6, and increased again by a further 2 fold from Day 6 to Day 8. HBA and HBG (HbE Gower 2 and HbF) were expressed between Day 0 and Day 4 but by less than 2 fold change, but their expression increased by more than 2 fold from Day 4 to Day 6, and increased again by over 2 fold from Day 6 to Day 8. HBD and HBB (HbA2 and HbA) were expressed between Day 0 and Day 6 albeit at a very low level and by less than 2 fold change, but their expression increased dramatically from Day 6 to Day 8 by greater than 2 fold.

3.4 Discussion

3.4.1 Discovery of Novel Haematopoietic Reference Genes

Housekeeping genes are constitutively expressed and are required for the maintenance of the basal cellular function in all human cells [321]. Many housekeeping genes are commonly used in research in the normalisation of quantitative gene expression data (for example, mRNA levels) between samples, and although housekeeping genes are constitutively expressed in every cell of the human body, their expression level may vary amongst different tissue types and different stages of development [322]. The commonly used housekeeping genes in PCR/qPCR studies, *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) and *beta-actin* (*ACTB*) are not expressed at the same level in all cell type at all stages of development, with a study by Barber *et al.* (2005) [323] demonstrating a 15-fold difference in *GAPDH* mRNA copy number between screens of 72 different normal human tissue types. The search for stably expressed genes for use in the normalisation of gene expression data for HESC differentiation remains a challenge with these reference genes required to be expressed at the same level in all cells types that are produced from HESCs.

Through the analysis of the microarray data, the identification and separation of transcript expression which change significantly and those which does not, have uncovered a set of transcripts which appear to be stably expressed at a high level throughout this haematopoietic differentiation from HESCs. This list has been included in the APPENDIX: Table 38 and these transcripts have been rank in order of expression intensity with the highest at the top. These transcripts may not be stably expressed in all cell types at all stages of differentiation, but from the data presented from this experiment, these genes appear to be excellent candidates for reference genes in the haematopoietic differentiation of HESCs for use in future studies. A closer examination of this list of stably expressed haematopoietic reference genes reveal a bias towards ribosomal proteins (RP) which, accounts for 80% of the transcripts within the top 20 highest ranked candidates. It is known that ribosomal proteins (RP) in conjuction with ribosomal RNA (rRNA) make up the ribosomal subunits of ribosomes and function within the cellular translation system in protein synthesis [324]. A search of the list for RPs which have been identified as stably expressed housekeeping genes in literature [321] reveals a number of matches including, *RPS18, RPL36A, RPS19, RPLP2, RPL3, RPL15, RPS2, RPS16, RPS12, RPS10,* and *RPL35*. However the transcripts within this list requires further

analysis and for their expression levels to be re-evaluated during different method and conditions of haematopoietic development, but the inclusion of a number of known reference genes including *GAPDH* and *ACTB* adds confidence to the validity of novel reference genes contained within this list.

3.4.2 Formation of the Haemogenic Endothelium

Gene ontology and pathway analysis has shown the up-regulation of a large proportion of transcripts involved in blood vessel and vasculature development including cell-to-cell adhesion factors, suggesting the onset of haemogenic endothelium. As previously described, the haemogenic endothelium acts as an intermediate between the formation of haemangioblasts, and the generation of haematopoietic cells [192, 193]. The highest-ranking vascular development gene map revealed platelet derived growth factor (PDGF) signalling as the most significant pathway from the Meta analysis of the imported gene list, due primarily to the large number of transcripts that are comprised within this map (Figure 68). PDGF-B was up-regulated between day 0 and day 4 of differentiation, and studies suggest that this may act to modulate endothelial proliferation and angiogenesis through PDGF-B-receptors present on endothelial cells [325, 326]. Studies have shown that during vascular injury, endothelial cells produce and secret the coagulation protein, thrombin (Figure 67), of which one of its roles is a mitogenic factor in the induction of DNA synthesis and proliferation of the surrounding endothelial cells [327, 328]. This mitogenic activity of thrombin may act either directly through the activation of a number of protein kinases [329], or indirectly through the release of a number of growth factors which include PDGF-B [330]. From these results of these studies, it may be hypothesised that PDGF-B plays a role in vasculogenesis and angiogenesis during development in the formation of blood vessels from the haemangioblast. Studies have shown that the PDGF-B growth factor regulates biological functions through the binding to its receptor, PDGF-R-B and is able to regulate survival by inhibiting apoptosis and promoting proliferation though the signal transducer and activator of transcription (STAT) pathway and through nuclear factors of kappa light polypeptide in B-cells (NF-KB). Janmatt *et al.* (2010) [331] showed in mice studies, endothelial progenitor cell numbers within peripheral blood increased in correlation with increasing EPO dosage. Ashley *et al.* (2002) [332] demonstrated that EPO stimulates vasculogenesis and angiogenesis through mitogenic-induction within neonatal rat microvascular endothelial cells. Experiments also showed that EPO robustly induced signal transducer and activator of transcription 5 (STAT5) phosphorylation within

human umbilical vein endothelial cells (HUVECs), with EPO inducing PDGF-B, which in turn also played some role in the induction of STAT5 [331].

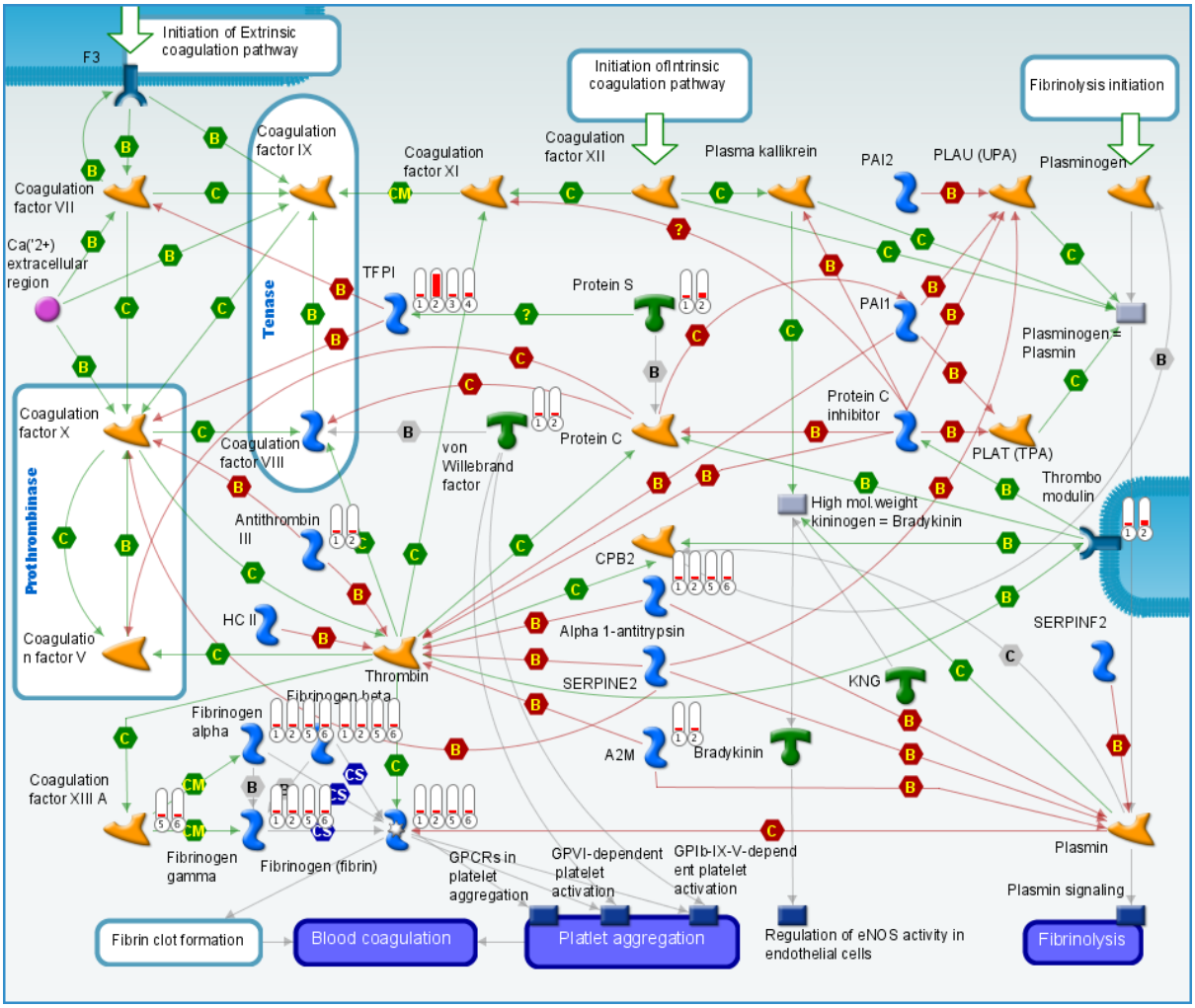


Figure 67: Metacore Gene Map - Blood coagulation and Platelet Activation

This pathway shows the interactions of the blood coagulation and platelets pathway. The genes containing bars next to them means that they are within the selected gene lists Day0toDay4Up2fold+HaemSim+p0.5, Day4toDay6Up2fold+p0.5, and Day6toDay8Up2fold+p0.5 and the size of the red within each bar describes the intensity of expression. Each gene is connected to another gene through an interaction that is either activation (green line) or inhibition (red line) or unknown (grey line), and the mechanism of interaction could be binding (B), phosphorylation (+P), transformation (T), transport (Tn), catalysis (Z), transport regulation (TR), or group relation (GR).

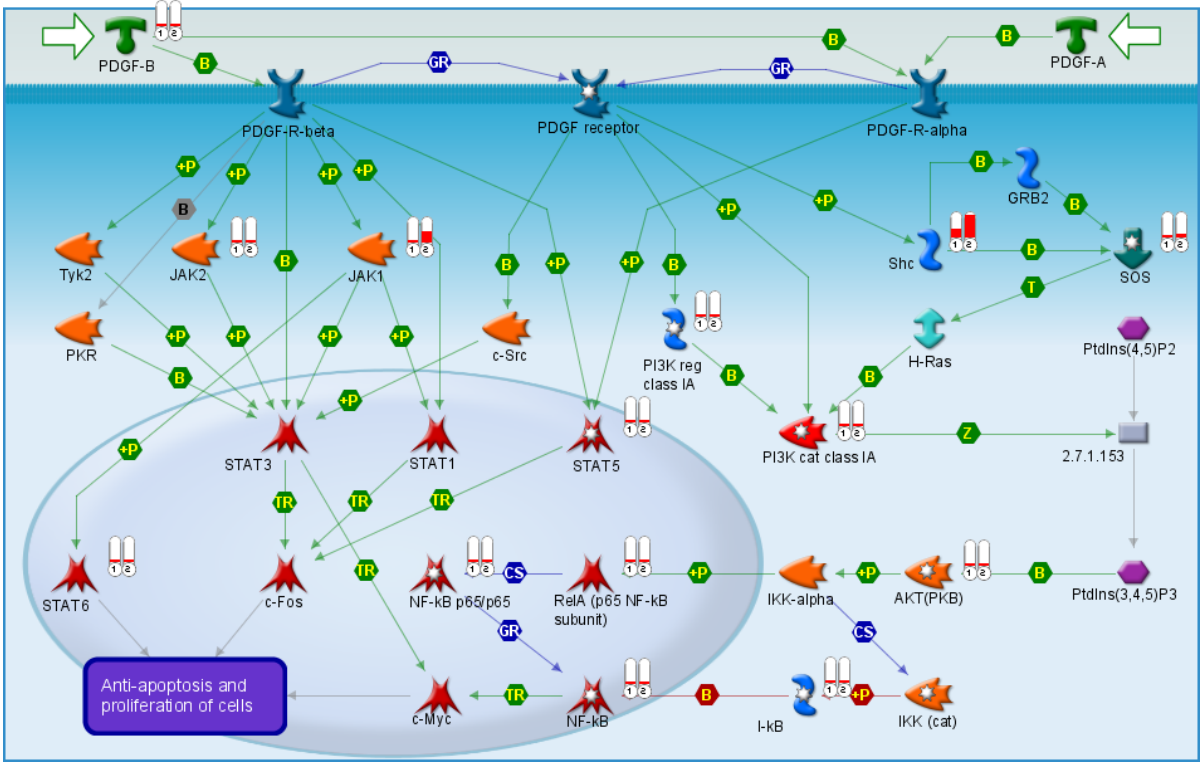


Figure 68: Metacore Gene Map: PDGF signalling via STATs and NF-kB

This pathway shows the interactions of Platelet Derived Growth Factor (PDGF) signalling through signal transducers and activators of transcription (STAT) and nuclear factor kappa-light-chain-enhancer of activated B (NF-κB). The genes containing bars labelled 1 and 2 show that they are within the selected gene lists Day0toDay4Up2fold+HaemSim+p0.5, and the size of the red within each bar describes the intensity of expression. Each of the genes are interconnected with interactions in the form of activation (green line) or inhibition (red line) or unknown (grey line), and the mechanism of interaction could be binding (B), phosphorylation (+P), transformation (T), transport (Tn), catalysis (Z), transport regulation (TR), or group relation (GR).

Kyba et al. (2003) [333] has shown that although STAT5 plays a positive role in the regulation and promotion of primitive haematopoietic blast cells, STAT5-knockout experiments have however still resulted in haematopoiesis which questions the importance of STAT5 in haematopoietic development. Shelbourne et al. (2003) [334] demonstrated that the loss of STAT5 expression results in the complete absence of mast cell development in vivo, which suggests a specific role in mast cell development. Studies have also shown an essential role for STAT5 in the induction of survival through the regulation of a number of survival (anti-apoptosis) proteins including, X-linked Inhibitor of Apoptosis Protein (XIAP), B-cell lymphoma-extra large (Bcl-XL), and Proto-oncogene serine/threonine-protein kinase (Pim-1), and thus demonstrate a pro-survival role. It was hypothesised that the up-regulation of platelet-related factors may result in the production of PDGF, and lead to the activation of JAK-STAT pathway, but studies show that this pathway may also be directly induced by EPO via its receptor EPOR [335]. It is known that a number of cytokines are mediated though the STAT5 pathway (via JAK2) including interleukin 2 (IL-2; T-Cell-derived cytokine in the regulation of growth /

differentiation monocyte cells), interleukin 3 (IL-3; multi-lineage-colony stimulating factor), interleukin 12 (IL-12; immunoregulatory cytokine), interleukin 22 (IL-22), and stem cell factor (SCF), however, Bunting (2007) [336] stated that the direct activation of STAT5 may result in cytokine-independent cell expansion. Studies have insisted that erythropoiesis is dependent on the JAK2 / STAT5 signal transduction pathway, with Kerenyi *et al.* (2008) [337] suggesting the reason for this may be attributed to the regulation of *transferring receptor-1* (*TfR-1*) and *iron regulatory protein 2* (*IRP-2*), which play important roles in iron metabolism.

These findings suggest that up-regulation of a number of transcripts in the EPO pathways (Figure 65) and blood coagulation and platelet activation pathway (Figure 67) may result in the direct and indirect (via PDGF (Figure 68)) activation of the JAK2 / STAT pathway, leading to the enhancement of endothelial proliferation, possibly as part of the haemogenic endothelium.

3.4.3 Activation of Globins marks the onset of Erythropoiesis

The human haemoglobin is formed by the symmetrical pairing of globular proteins, globin subunits, with each carrying a haem group consisting of protoporphyrin IX and a ferrous iron atom [338]. The globin composition within the haemoglobin changes over the course of development from embryonic, to foetal, and to adult, with these five sequential forms shown in Figure 66. The development of the human haematopoietic system, in particular that of erythropoiesis, may be tracked through the sequential expression of the different forms of the globin subunits. Within this microarray experiment, the globin expression has revealed the emergence of HBG, HBZ, and HBE from day 0 to day 4 (Figure 65), which identifies the possible development of the first embryonic haemoglobin, HbE Gower 1 (composed of HBZ and HBE globin subunits; Figure 66). The expression of HBG, HBZ, HBA, HBE globins from day 4 to day 6, suggests the possible formation of embryonic haemoglobin, HbE Gower 2 (HBA / HBE), and foetal haemoglobin, HbF (HBA / HBG). From day 6 to day 8, the up-regulation of all globin transcripts of HBG, HBB, HBZ, HBA, HBD, and HBE, includes the possible formation of adult haemoglobins. This sequential activation of globins may mark the onset of erythropoiesis, thus adding confidence to this differentiation study in the recapitulation of *in vivo* haematopoietic development.

The EPO-induced PI3K/AKT pathway (Figure 65) links the up-regulation of the EPO receptor, EPOR, to the up-regulation of the various other components of the pathway, including the globin transcripts. In this study, the level of expression of EPO activity does not appear to change through this differentiation experiment, remaining at a basal level comparable to that of day 0. It may therefore be hypothesised that the addition of EPO cytokine into the differentiation media may enhance erythropoiesis through the positive activation of the EPO-induced PI3K/AKT pathway, and may possibly positively affect haematopoietic development in general. EPO has been routinely used within Methocult methylcellulose media in the formation of colony-forming units – erythroids, but there has not been a published differentiation protocol that utilises EPO in the enhancement of the haematopoietic differentiation from HESCs.

3.4.4 A Hypoxic role in EPO regulation

In the human adult, EPO release from the kidney / liver is triggered by the detection of low oxygen levels via hypoxia-inducible factors, which in turn stimulates the overproduction of red blood cells within the bone marrow [339]. It may be hypothesised that the up-regulation and activation of transcripts associated with the erythropoietin (EPO) induced phosphoinositide 3-kinase (PI3K) / AKT pathway (Figure 65) may also be regulated by hypoxia. It would be of interest to attempt to repeat the differentiation of hESC using the same protocol as previously described in this project, but under hypoxic conditions (5% oxygen) and during each day of differentiation to sample and test the cells using qPCR in monitoring EPO / EPOR expression levels to verify the proposed link. Recent findings by Prado-Lopez *et al.* (2009) [340] demonstrated that under hypoxic conditions, HESCs would readily differentiate into functional endothelial cells with over 50% of cells displaying CD34+ after just 24 hours under hypoxia. The incorporation of this discovery into the described differentiation protocol may dramatically enhance haematopoietic differentiation, but the same study also showed that a much larger percentage of apoptosis was also seen and increased exponentially with increasing hypoxia. Satoh *et al.* (2006) [341] observed that hypoxia enhances the expression of erythropoietin receptor (*EpoR*) on mouse endothelial cells during hypoxia-induced pulmonary hypertension. The same study also showed that hypoxia activated nitric oxide (NO) synthase in wild-type mice but not in EpoR(-/-) rescued mice, which suggests that hypoxia acts to activate nitric oxide signalling essentially through EPO / EPOR activation. Nitric oxide (NO) production appears to be enhanced by EPO through modulation of endothelial nitric oxide (NO) synthase [342], and recent studies have suggested a close regulation of this pathway by shear stress and other biomechanical forces [343]. Biomechanical forces including shear stress have been shown to promote embryonic haematopoiesis and increase haematopoietic colony forming potential through nitric oxide (NO) signalling, and abrogation of NO severely compromises haematopoiesis [344]. The microarray results in conjunction with the described published studies appear to suggest that both hypoxia (via EPO) and shear stress both may act on the NO signalling pathway in the promotion of haematopoiesis.

NO signalling appears to play an important role in cellular signalling and studies have suggested that it has an important role in many biological processes, which include cell communications (neurons) [345], immune defence against pathogens [346], and vasodilation [347]. NO signalling is controlled by a family of Nitric Oxide Synthases (NOS) of which there are 3 isoforms (neuronal NOS, inducible NOS, and endothelial NOS) which catalyse the production of NO in different tissues / organs of the body. *In vitro* and *in vivo* mouse studies of knockout

NOS isoforms by Krasnov *et al.,* 2008 [348] has suggested that the key NOS in haematopoietic regulation is neuronal Nitric Oxide Synthase (nNOS), however little is known with regards the mechanism behind these findings. In contrary to these findings, a search through the microarray gene lists reveals the presence of the endothelial NOS transcript (eNOS) but not neuronal NOS (nNOS) or inducible NOS (iNOS) which suggests that eNOS may play a role in early human haematopoiesis but perhaps less so in mouse. Future work including siRNA repression or overexpression differentiation studies of the NOS family of genes may reveal more information on the isoform(s) that may be important in haematopoietic development.

3.4.5 A role for SOST in haematopoietic development

Sclerostin (*SOST*) was first reported in sclerosteosis patients where two nonsense mutations and a single splice site mutation were identified, and it was observed that this loss of function resulted in the dramatic over production of bone within these patients [297]. Sclerosteosis is an inherited autosomal recessive bone dysplasia which causes hyperostosis and sclerosis, leading to characteristically thickening of bones throughout the body of which the skull, jaw, and ribs are most prominent in these patients. The gross hyperostosis can result in the narrowing of the foramina of the skull through which nerves, arteries, veins, and other structures pass, and can induce to a number of disorders including hearing loss, atrophy of optic nerves, and facial nerve palsy. The *SOST* gene was first implicated in sclerosteosis when it was noted that the phenotypic representation of sclerosteosis shares close similarities to van Buchem's disease which is also responsible for 'high bone mass'. Linkage analysis had tracked both disease genes to being on the same chromosomal region of 17q12-q21, and further positional cloning strategies identified that the both diseases were caused by mutations in the *SOST* gene. In sclerosteosis patients, the *SOST* gene is inactivated by mutations leading to complete loss-of-function of the SOST protein, whereas in van Buchem patients a 52kb homozygous deletion 35kb downstream of the *SOST* gene affects transcription and results in a very restricted expression [349].

A gene map and process network search within Metacore revealed no pathway or network hits, which suggest that the *SOST* gene's functional mechanisms and interactions may not have been thoroughly investigated. However studies to this date, have implicated Bone Morphogenetic Protein (BMP) 2, 4, and 6 in the regulation of *SOST* expression in mouse and human osteoblasts,[350] and that bone mass regulation was regulated by BMP through SOST

by the inhibition of the canonical Wnt signalling pathway [351]. It has however been noted that sclerosteosis shares close similarities van Buchem's disease which is responsible for 'high bone mass' resulting from gain-of-function mutations in the Low-density lipoprotein receptor-related protein 5 (LRP5) gene which is a wingless-type MMTV integration site family (Wnt) co-receptor [352]. Studies have established that SOST regulate and function as a bone mass regulator on the Wnt signalling cascade (as a antagonist of canonical Wnt signalling) [353] through LRP5 and LRP6 receptors [354] by the disruption of Wnt-induced Frizzled-LRP complex formation (Chapter 1: Figure 6) [124, 352]. Wnt proteins (as well as bone mass regulation) have been implicated in a diverse number of developmental roles including cell proliferation, stem cell maintenance, cell fate, cell migration, and tissue polarity establishment in all animal species studied to date [355]. The Wnt-induced Frizzled-LRP complex formation (Figure 6) is required for the regulation of the beta-catenin signaling pathway [356] which in turn leads to the inductions of a series of osteoblastogenic transcription factors including runt-related transcription factor 2 (RUNX2), distal-less homeobox 5 (DLX5), and Sp7 transcription factor (Osterix; SP7) [357].

To this date, there has been no direct links between SOST with regulation of haematopoiesis, but past studies have uncovered that bone morphogenetic protein-4 (BMP-4) is crucial for early development for mesoderm induction [111], and this gene is one of many that is activated / enhanced by lymphoid enhancer factor (LEF) / T cell factor (TCF) (Chapter 1: Figure 6). BMP-4 is an essential component of the haematopoietic differentiation media and the introduction of SOST into the media may assist to inhibit non-haematopoietic gene expression through the canonical Wnt pathway. The expression of SOST was highly up-regulated (increased by 197 fold) from Day 0 to Day 4, but its expression plateaued from Day 4 to Day 6 and decreased from Day 6 to Day 8, reinforcing a possible new role in mesoderm lineage-specification.

CHAPTER 4 – SCL

4.1 Introduction

4.1.1 Microarray analysis supports role of SCL in Haematopoiesis

Within the microarray results in Chapter 3, the gene expression change of *Stem Cell Leukaemia* / *T-cell Acute Lymphocytic Leukaemia* (*Scl*/*Tal1*) gene was ranked 11th highest of the 3,162 transcripts that were up-regulated between day 0 and day 4 (APPENDIX: Table 39). This finding correlates perfectly with what has already been discovered within the published literature, however the exact role and function of SCL within HESC differentiation remains unclear. Studies have shown that Scl plays a crucial role in the specification of the mesoderm and the formation of the haemangioblast [358], in the formation of blood vessels [359], and in erythropoiesis [360] in a number of species. It was of great interest to investigate if all or any of these roles of SCL are concordant within HESC differentiation.

4.1.2 The Discovery of Scl and its Importance in Haematopoiesis

The *Stem Cell Leukaemia* / *T-cell Acute Lymphocytic Leukaemia* (*Scl*/*Tal1*) gene was first reported 21 years ago, where it was discovered at the breakpoint of chromosomal translocations within T-cell acute lymphoblastic leukaemia (T-ALL) patients [361]. Clinical studies have shown that this *Scl* rearrangement was associated solely with leukaemic T-lymphocytes, but this was not detected in the peripheral blood lymphocytes of disease-free individuals [362].

The detection of *Scl* mRNA in early haematopoietic tissue (of the foetal liver) by Begley *et al.* (1989) [363], had led to speculation that it may play a role in differentiation and commitment of haematopoiesis. In the assessment of the importance of Scl in haematopoiesis, Dooley et al. (2005) utilised site-directed, anti-sense morpholinos in the inhibition of Scl mRNA within Zebrafish, which resulted in the completely loss of primitive and definitive haematopoiesis. The same results were also observed within the mouse model, but due to the early lethality of Scl-knockout mice (from E9.5-10.5), effects of Scl during later stages of development, and that of the adult stage could not be analysed. In overcoming this, Hall et al. (2003) [364] generated

conditional Scl-knockout mice using IFN-inducible Cre / LoxP-mediated mutagenesis, and their results (showed low haematocrit and platelet counts) suggested that *Scl* may be a critical regulator of megakaryopoiesis and erythropoiesis during adult haematopoiesis. A different approach was taken by Elefanty et al. (1998) [365] whereby the *Escherichia coli lacZ* reporter gene was knocked-in to the *Scl* locus of mice, and thereby linking *β-galactosidase* expression to *Scl* expression. It was discovered through analysis of varying levels of flow cytometry sorted Scl fractions, that Scl expression was transient and confined to haematopoietic progenitor cells but not their differentiated progeny. The role of Scl within these haematopoietic progenitor cells was examined by Lacombe et al. (2010), where is was discovered that Scl was highly up-regulated within long-term reconstituting haematopoietic stem cells (LT-HSCs) compared with short-term reconstituting haematopoietic stem cells (ST-HSCs). Scl appears to regulate cell cycle within LT-HSCs, resulting in the majority of these arresting in the G0 phase in a state of quiescence, but unlike quiescence associated with senescence this state is reversible. This requirement of inactivity may be crucial in the contribution to long-term stem cell function, and in balancing the regenerative needs with the potential for cell overgrowth (as is the case in cancers where growth is deregulated).

4.1.3 A possible role for Scl in Blood Vessel Development

Within Scl-knockout mice, Robb *et al.* (1995) [169] discovered although endothelial capillaries were present within the yolk sac, the fusion of the associated capillary spaces in the formation of large vitelline vessels did not occur. These results were further supported by Visvader *et al.* (1998) [359] who, through the rescue of Scl-/- mouse embryos with the introduction of a *Scl* cDNA-transgene, demonstrated that Scl may be dispensable for the initial formation of endothelial cells (vaculogenesis), but it is crucial for the latter formation of vitelline vessels from the vascular plexus (angiogenesis). However in a separate study by Gering *et al.* (1998) [177], it was suggested that Scl may play an important role in the specification and formation of endothelial progenitor cells (haemangioblast). It was discovered through the ectopic expression of Scl mRNA in zebrafish embryos, this correlated with the over-production of haematopoietic and endothelial progenitors at the expense of other mesoderm cell types. The importance of Scl in the mesodermal specification of the haemangioblast has been discussed and described previously in chapter 3, however the

mechanisms and interactions of Scl within this role will be analysed in detail later in this chapter.

4.1.4 The Structure and Function of Scl

Although it has been well established that Scl regulates transcripts of a multitude of developmental pathways including those of haematopoietic and angiogenic, its direct targets are largely unknown. However, an examination of the structure of *Scl* may provide clues with regards its function and interactions, with studies showing that *Scl* is a member of the basis-helix-loop-helix (bHLH) family of transcription factors [167]. The bHLH family of proteins are known to be critically involved in the transcriptional regulation of a number of developmental pathways in many organisms, with examples including worms (*CeMyoD / HLH-1* – body wall muscle regulation [366]), fruit flies (*MyoD* – cell patterning and embryonic muscle formation [367]), zebrafish (*HAND2* – heart and fin development [368]), mice (*dHAND* / *eHAND* – cardiac mesodermal and neural crest development [369]), and humans (*TWIST* - cell type determination and differentiation [370]). The functional distinct regions of the family of bHLH of nuclear proteins may be associated with a conserved ~60 amino acid motif (Figure 69) of which studies have found to be essential for DNA binding [371]. This bHLH motif is composed of a basic region that is involved in DNA binding and a HLH region that functions as a dimerisation domain, constituting of hydrophobic residues that form two helices separated by a loop region (Figure 70). An amino acid (a.a.) sequence comparison between Scl homologs from a number of species (Figure 69) revealed striking similarities, with an average similarity score of approximately 58% between all species. The two species with the greatest difference in a.a. sequence are mouse and zebrafish (a score of 48%), and the two most similar species are mouse and human (a striking 93%). However across all species, there is a string of 57 a.a. which is identical in every sequence (Figure 69), and this string encodes the highly conserved bHLH domain. It is largely due to this conserved domain, which governs the function and overall structure of the Scl protein in humans and its homologs, and a protein structure comparison in Figure 70 neatly demonstrates this similarity with the consensus HLH formation.

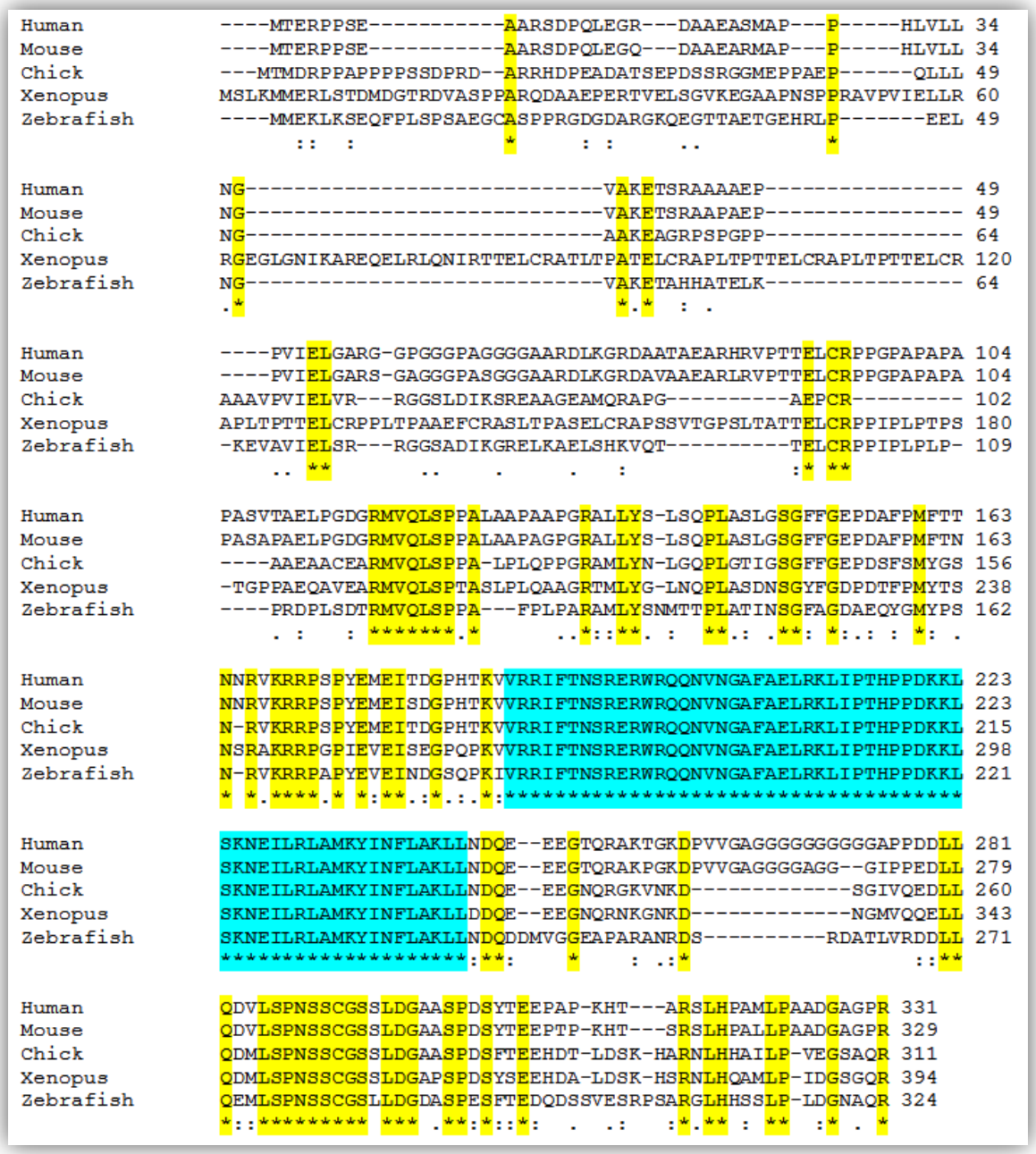


Figure 69: Amino acid sequence of SCL from Human, Mouse, Chick, Xenopus, and Zebrafish

The full amino acid (a.a.) sequences of human (331 a.a.) , mouse (329 a.a.), chick (311 a.a.), xenopus (394 a.a.), and zebrafish (324 a.a.) SCL homologs were aligned with using ClustalW2 from EMBL-EBI [372]. The results show a homology of approximately 50% across all 5 species, with mouse and xenopus homologs being the most different (48%) and human and mouse homologs being the most similar (93%). Nucleotides that are identical across all species are coloured in yellow, and the conserved bHLH domain is marked in blue (57a.a.).

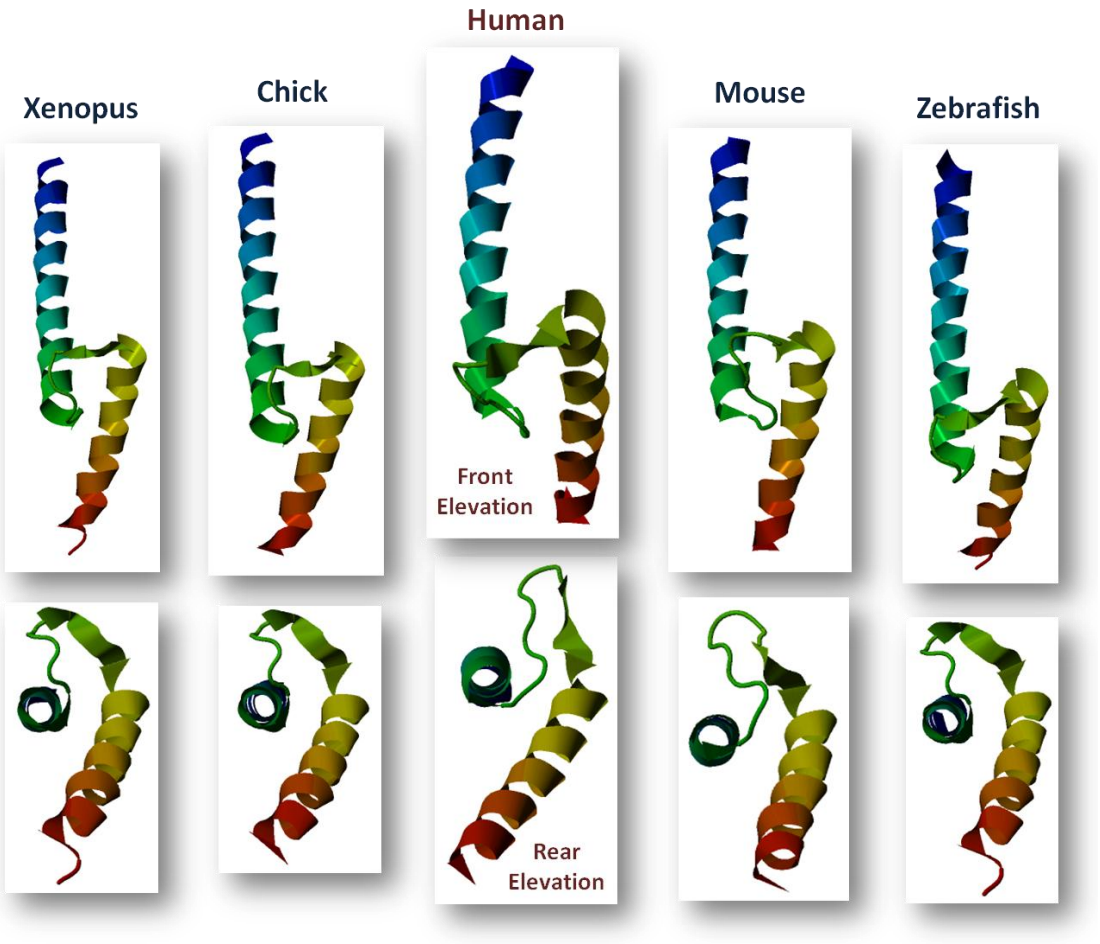


Figure 70: 3D Protein Structure of SCL / TAL1

Computational prediction of protein structure models of SCL / TAL1 showing two elevations (views) of the basic helix-loop-helix (bHLH) structure of human, mouse, chick, zebrafish, and xenopus homologs. These computational structures were produced from The Protein Model Portal [373, 374] based on the individual amino acid sequences and a model template (2ql2B) from the x-ray diffraction study of the basic-helix-loop-helix domains of the heterodimer E47 / NeuroD1 [375].

4.1.5 Scl Interactions

With all bHLH proteins, dimerisation is required for DNA recognition, and studies have shown in the case of Scl, dimerisation occurs with E12 and E47 (bHLH proteins encoded by *transcription factor 3* (*TCF3* / *E2A*)), to produce heterodimers Scl/E12 and Scl/E47 [376]. In fact, Hsu *et al.* (1994) [376] reported that Scl (a class II bHLH) may readily form heterodimers with any of the known class I bHLH proteins (E12, E47, E2-2, and HEB) which are ubiquitously expressed as general co-factors (as apposed to class II bHLH proteins which are expressed in tissue-specific patterns). As a heterodimer, Scl/class I bHLH co-factor will bind to DNA in a sequence-specific manner, recognising and binding with high affinity to the specific nucleotide-binding site, CANNTG (E-box motif) [376].

There has been increasing evidence that the Scl heterodimers do not directly bind to DNA alone, but rather function as part of a larger complex in the regulation of transcription. A search through the published literature using the metacore database has revealed a number of Scl interactions including, ubiquitin (UB), lim domain only 2 (LMO2), v-akt murine thymoma viral oncogene homolog 1 (AKT1), lim domain only 1 (LMO1), nuclear factor erythroid-derived 2 (NF-E2), Friend leukemia virus integration 1 (FLI1), transcription factor 3 (TCF3/E2A), v-myb myeloblastosis viral oncogene (C-MYB), v-ets erythroblastosis virus E26 oncogene homolog 1 (ETS1), Sp1 transcription factor (SP1), histone deacetylase 1 (HDAC1), Sp3 transcription factor (SP3), v-ets erythroblastosis virus E26 oncogene homolog 2 (ETS2), transcription factor 4 (TCF4/ITF2), forkhead box C2 (FOXC2), E74-like factor 2 (ELF2), GATA binding protein 3 (GATA3), GATA binding protein 1 (GATA1), runt related transcription factor 1 (RUNX1), runt related transcription factor 2 (RUNX2), runt related transcription factor 3 (RUNX3), inhibitor of DNA binding 2 (ID2), nuclear factor kappa-B (NF-KB), octamer-binding protein 4 (OCT4), sex determining region Y (SRY), and human immunodeficiency virus type I enhancer binding protein 2 (HIVEP2) (Figure 71). However, not all of these interactions are haematopoietic-related, and an example of this is the detection of Scl mRNA in the brains and spinal cord of zebrafish, which suggests that the protein interactions within this location may not correspond to interactions within the bone marrow [377]. Not all interactions are positive, for example, Terme et al. (2009) [378] reported that Scl is degraded by the actions of UB through AKT-mediated phosphorylation regulated by the transforming growth factor beta (TGF-beta). Studies have shown that FLI1 sits at the top of the haematopoietic transcriptional network and can induce the expression of a number of key haemangioblast genes including *GATA2, KDR, LMO2,* and *Scl* [379]. With reference to the microarray gene list produced within chapter 3

(**Day0toDay4Up2fold+HaemSim+p0.5)** which displayed genes with significant gene expression differences between day 0 and day 4 of differentiation, *FLI1* showed a fold change difference of 89 (ranked 14th in the gene list). Perhaps it is the high expression of this master regulator which has induced the expression of the downstream haematopoietic targets, with *GATA2* showing 6 fold change difference between day 0 and day 4, and *KDR* showing a fold change of 25, with *LMO2* displaying 41 fold change, and *Scl*, a 92 fold change. As well as FLI1, the published literature describes a number of other proteins that regulate the Scl promoter activity and these include PU.1, ELF1, SP1, SP3, and GATA1 [380].

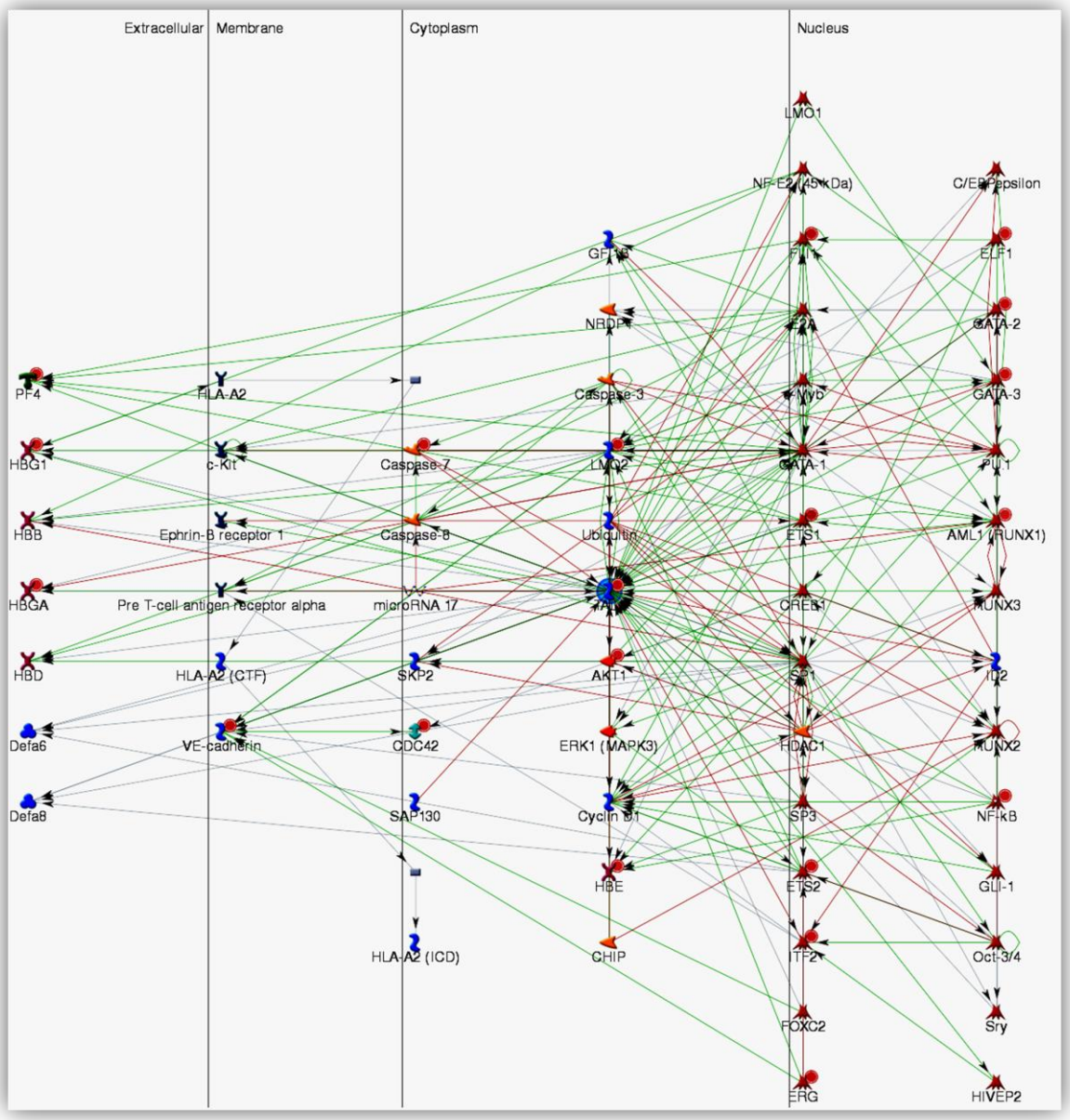


Figure 71: Network of SCL / TAL1 interactions

This network map of SCL / TAL1 interactions was produced from the Metacore database [282] of gene and protein interactions, human curated from the published literature. The green lines indicated a positive (enhancing) interaction and the red lines suggest a negative (inhibiting) interaction.

With regards to Scl function, Gering *et al.* (2003) [358] implicated the LIM domain protein Lmo2 in a complex with Scl/E2A heterodimer, in the specification of the haemangioblast giving rise to endothelial cells in the absence of gata1 (globin transcription factor). This result is fascinating as it appears that these endothelial cells may possibly also have haematopoietic potential (haemogenic endothelium), and a simple test for this may be the addition of haematopoietic inducers and testing for presence of haematopoietic cell surface markers. Using yeast two-hybrid assays, Wadman *et al.* (1997) [381] the protein-protein and protein-DNA interactions of Scl were analysed, and it was shown within erythroid cells, Scl first forms a heterodimer with e2a, before binding to Lmo2, Ldb1, and Gata1 to form a pentameric complex that binds to a DNA sequence consisting of an E-box domain upstream of a GATA site. These E-box/Gata1 sites were first identified within erythroid genes, however these have since been found in the regulatory elements of a number of haematopoietic genes including *Scl* and *Gata1* themselves [360]. The important of Lmo2 within this pentameric complex was supported by the experiments of Warren *et al.* (1994) [382], who showed that the Lmo2-knockout phenotype within mice were identical to that of the Scl-knockout (failure of both primitive and definitive haematopoiesis). Mead *et al.* (2001) [383] over-expressed *Scl, Lmo2,* and *Gata1* in Xenopus and discovered a synergistic capacity to ventralise the embryo and observed blood throughout the dorsal-ventral axis, greater than that seen with just *Scl* over-expression alone. It is clear that although *Scl* is a master regulator of primitive and definitive haematopoiesis, it does not achieve this regulation alone, and largely relies on other members of its functional complex as well as its own regulator(s) of expression.

4.2 Methods

4.2.1 *Scl* over-expressing HESC lines

Due to time constraints, the *Scl* over-expressing HESC lines H9 and H1 were kindly donated by a former colleague of our group, Dr. Nicholas Slater, who produced these stable cell lines using a lenti-viral vector containing *GFP* and *Scl* (Figure 72) and transduced into H9 and H1 HESC lines.

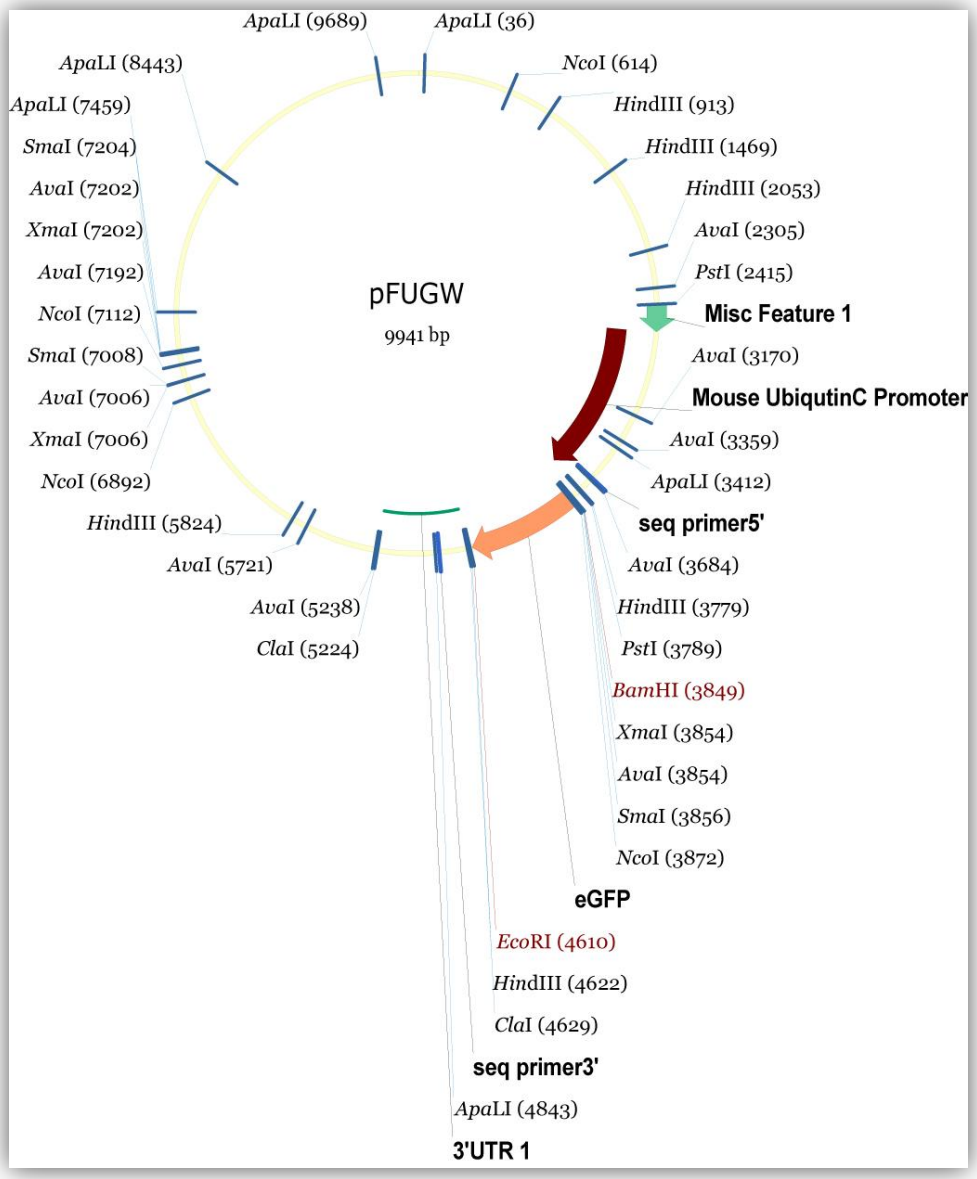


Figure 72: Lentiviral Vector Map of FUGW

This lentiviral vectorc containing the mouse ubiquitin promoter drives the expression of eGFP together with the transgene was used in the creation of H9 SCL, H9 FUG (control), H1 SCL, and H1 FUG (control) HESC lines. Unfortunately, it was not known precisely where *Scl* was inserted within this vector.

In addition, the precise location of the *scl* transcript insert within the pFUGW plasmid construct is unknown and therefore in order to verify that the *scl*-overexpressing cells (although shown by qPCR analysis that they do transcribe the *scl* transcript at high levels) are also capable of translating into the active Scl protein. As shown in Figure 72, eGFP is present in the same vector, and has been shown to be active at both the transcript and the protein level (verified through FACS analysis), this alone cannot be proof that the Scl protein is produced. Again, it comes back to the uncertainty of the cloning strategy used in the initial creation of the cell line, and the location of *scl* within the construct can have a dramatic effect on the translational capacity. For example, if *scl* was placed straight after eGFP in the EcoR1 site (Figure 72) and an internal ribosome entry site (IRES) was not placed immediately before the transcript, then translation initiation cannot occur but the mRNA transcript copy would still be available. Even though a simple solution would be to sequence the vector from the forward and reverse positions of *eGFP* to discover the location of the *Scl* gene, it is of greater concern that the Scl protein is produced and therefore may be accounted for the results shown in this chapter. A positive result from a Western blot which verifies the presence of eGFP and Scl protein is of paramount importance and must be completed prior to publication of any results from this study.

Unfortunately, the H9 SCL and H1 SCL cell lines contain the full *Scl* transcript of the mouse homolog and not that of the human. It is unknown how these different proteins will behave *in vitro* but a protein sequence comparison between the human and mouse *Scl* homologs reveal very little difference, showing a 93% similarity score after alignment in EMBL-EBI's ClustalW [372] (Figure 73). The crucial bHLH domain (highlighted in blue) shows a perfect match between the two homologs, with other species in literature (chick, zebrafish, xenopus) showing greater differences in other areas of the a.a. sequence (Figure 69) yet their effects during development appear to consistent. An examination of the protein structures in Figure 70 also suggest that small single amino acid differences may not have a significant overall effect on the protein structure and function, which instead appears to heavily dependent on the highly conserved bHLH domain.

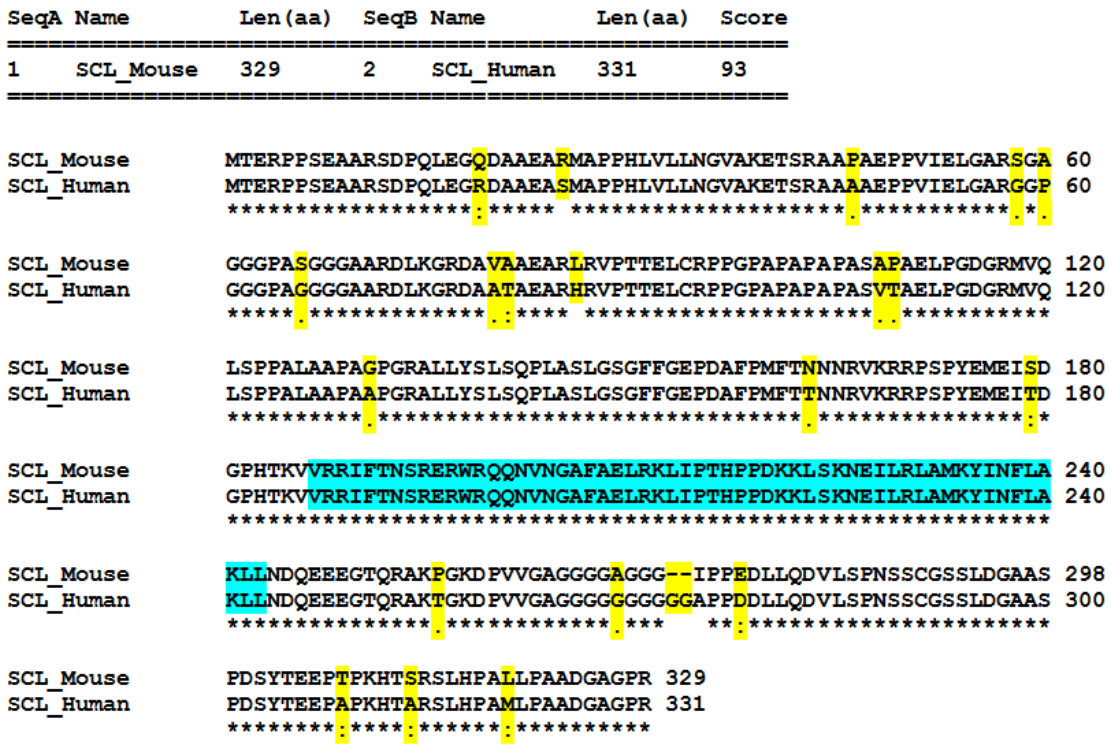


Figure 73: Amino acid sequence of SCL from Human and Mouse

The full amino acid (a.a.) sequences of human (331 a.a.) and mouse (329 a.a.) SCL were aligned with using ClustalW2 from EMBL-EBI [372]. The results show a 93% similarity rating with 21 a.a. differences (highlighted in yellow), however the bHLH motif (highlighted in blue) containing about 60 a.a. is 100% conserved between human, mouse, chick, xenopus, and zebrafish [384].

4.2.2 Protocol - SCL Differentiation and Analysis

Four cell lines were used in the investigation of the effects of Scl in the haematopoietic differentiation of HESCs, composed of two *Scl* over-expressing lines (H9 SCL and H1 SCL) and two controls (H9 FUG and H1 FUG) which contain an empty plasmid which expresses *GFP*. These were sub-cultured, passage and maintained using the same protocols as previous described in chapter 2 until the desire number of plates of cells has been reached. Differentiation was initiated with the formation of embryoid bodies (EBs), and these were incubated in the optimised differentiation media as outlined in chapter 2. Samples of EBs were taken at day 0, day 4, day 6 and day 8, and analysed by flow cytometry, quantitative polymerase chain reaction (qPCR), and colony forming assay (as previously described within methods of chapter 2).

4.3 Results

4.3.1 GFP expression correlates with *Scl* expression

The green fluorescent protein (GFP) reporter gene was included in the lentiviral vector to enable the *SCL* over-expressing cells to be tracked during differentiation and crucially for flow cytometry analysis (Figure 74A). Through visually observation under the microscope, it was observed that GFP expression within monolayer colonies during standard HESC sub-culture were not as bright when compared with EBs during differentiation. Flow cytometry analysis between undifferentiated and differentiated (H9 SCL and H9 FUG) cells confirm this hypothesis which shows the mean fluorescence intensity of GFP increase by over two fold between Day 0 to Day 8 of differentiation (APPENDIX: Figure 94). This phenomenon may be due to the intrinsic effect of silencing mechanisms within HESCs possibly brought on by DNA methylation of the promoter, histone modifications, and/or chromatin remodelling, which has been observed in a number of HESC studies [385-387].

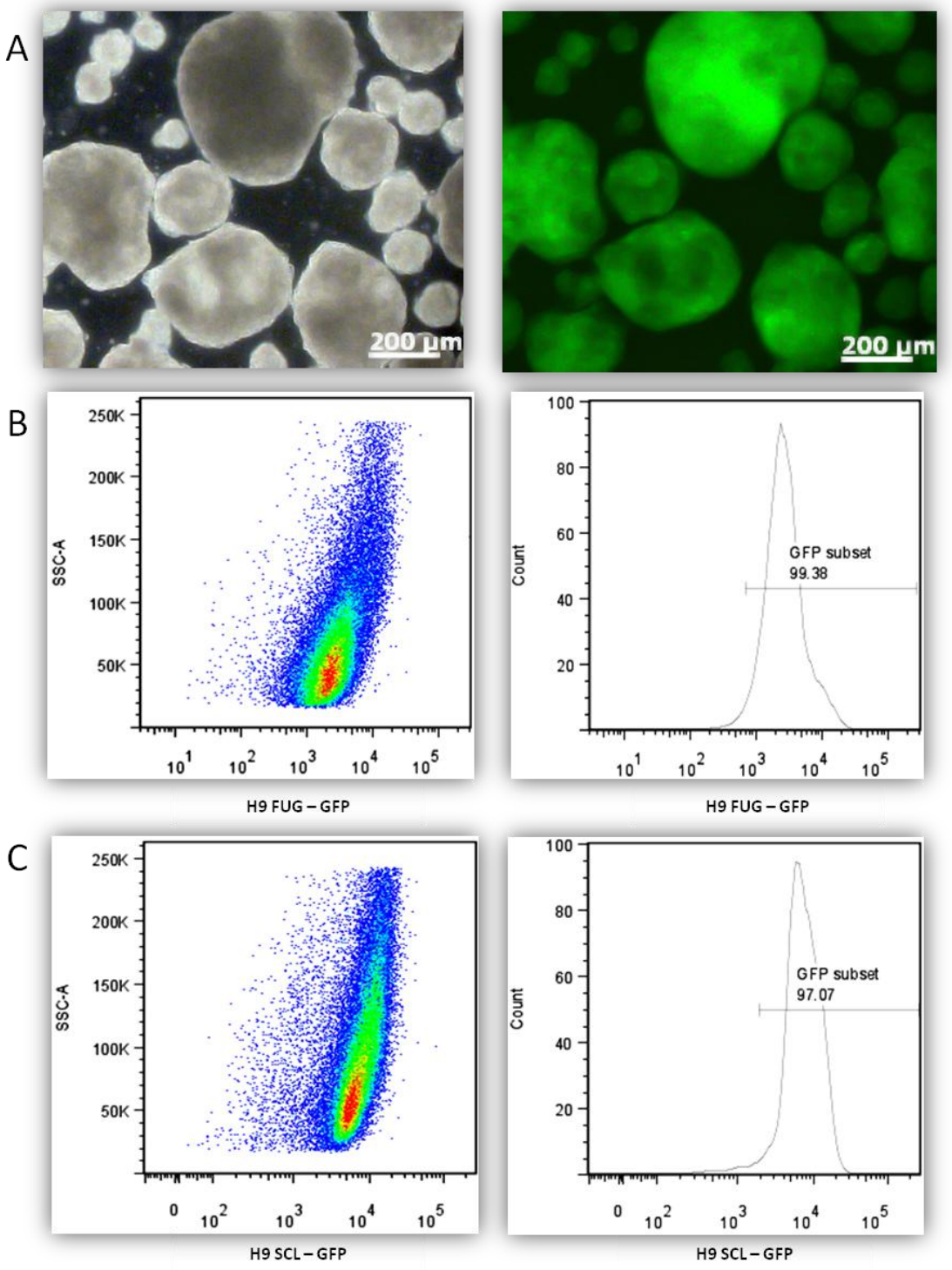


Figure 74: GFP expression from H9 FUG and H9 SCL

(A) Images of Day 8 EBs of H9 SCL under brightfield (left) and under blue light fluorescence fluoresces green (right). (B) Flow cytometry plot of Day 8 H9 FUG (empty lentiviral vector without *Scl* transgene) showing high % of GFP+ cells. (C) Flow cytometry plot of Day 8 H9 SCL (lentiviral vector with *SCL* transgene) showing high % of GFP+ cells.

These silencing mechanisms however appear to be relaxed during differentiation and higher levels of *GFP* and *SCL* expression emerge as shown in the flow cytometry (Figure 74) and qPCR analyses (Figure 75). The expression profile of the two *Scl* expressing HESC lines H9 and

H1 appear to portray similar trends, with both lines demonstrating significant over-expression of *Scl* during all days of differentiation compared to those of the non-*Scl* expressing control (FUG). The *GFP* expression profile between the SCL cell lines do not show any significant changes over the course of differentiation, however with regards to the relative expression levels, it appears that the level of *GFP* within the SCL over-expression lines are generally higher than that in the FUG control lines, with the only exception being at day 4 which perhaps may have been possibly due to pipetting errors.

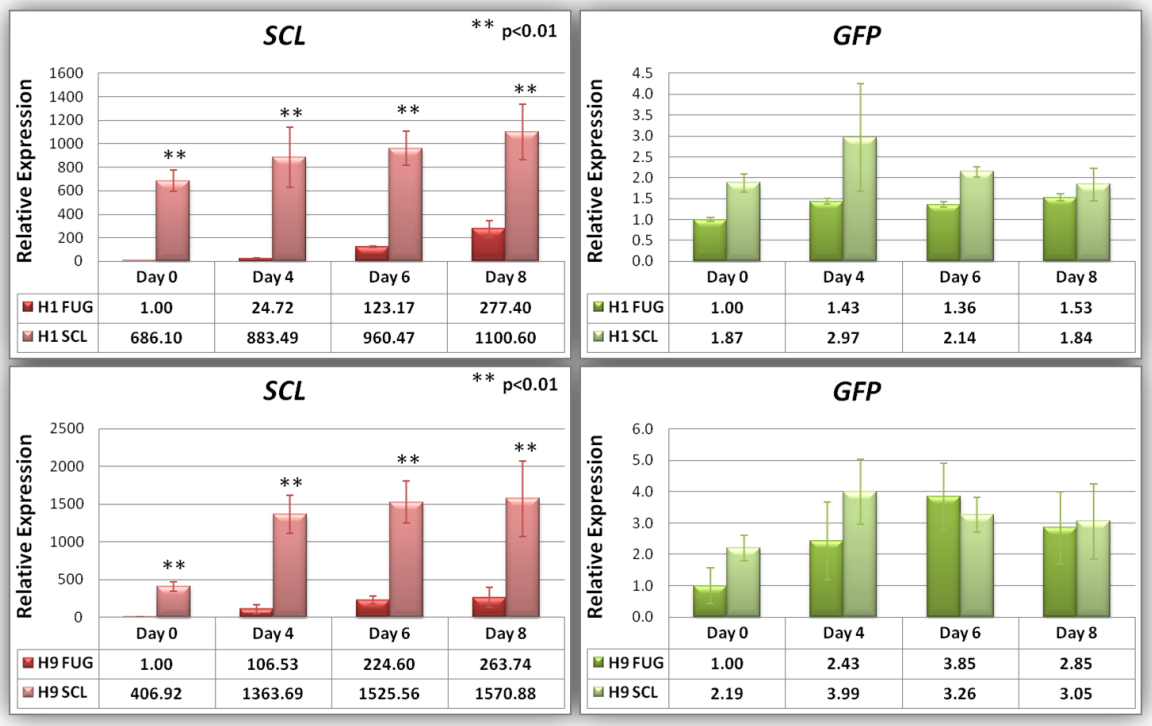


Figure 75: Real-time qPCR analysis of differentiated SCL and FUG cells – *GFP* and *SCL* genes

These qPCR plots show the profile of both GFP and SCL expression through haematopoietic differentiation within H9 FUG, H9 SCL, H1 FUG, and H9 SCL cell lines. Human/mouse compatible primers were used for the qPCR analysis and full primer sequences are shown in the APPENDIX. Differentiation samples were run in triplicate and the average expression values between the cell lines were calculated for the plots. Error bars were calculated using the standard error of the mean with three samples for each cell line, and p-value significance was calculated using t-test across three replicates for each cell line.

4.3.2 SCL Enhancement of Haemogenic-specification

The flow cytometry results (Figure 76; FACS plots in APPENDIX) suggest that the over-expression of *SCL* has a significant effect on the haemogenic differentiation capacity of HESCs as expected, however the degree of enrichment does not appear to be as great as hypothesised from the over-expression of such a widely published haematopoietic regulator.

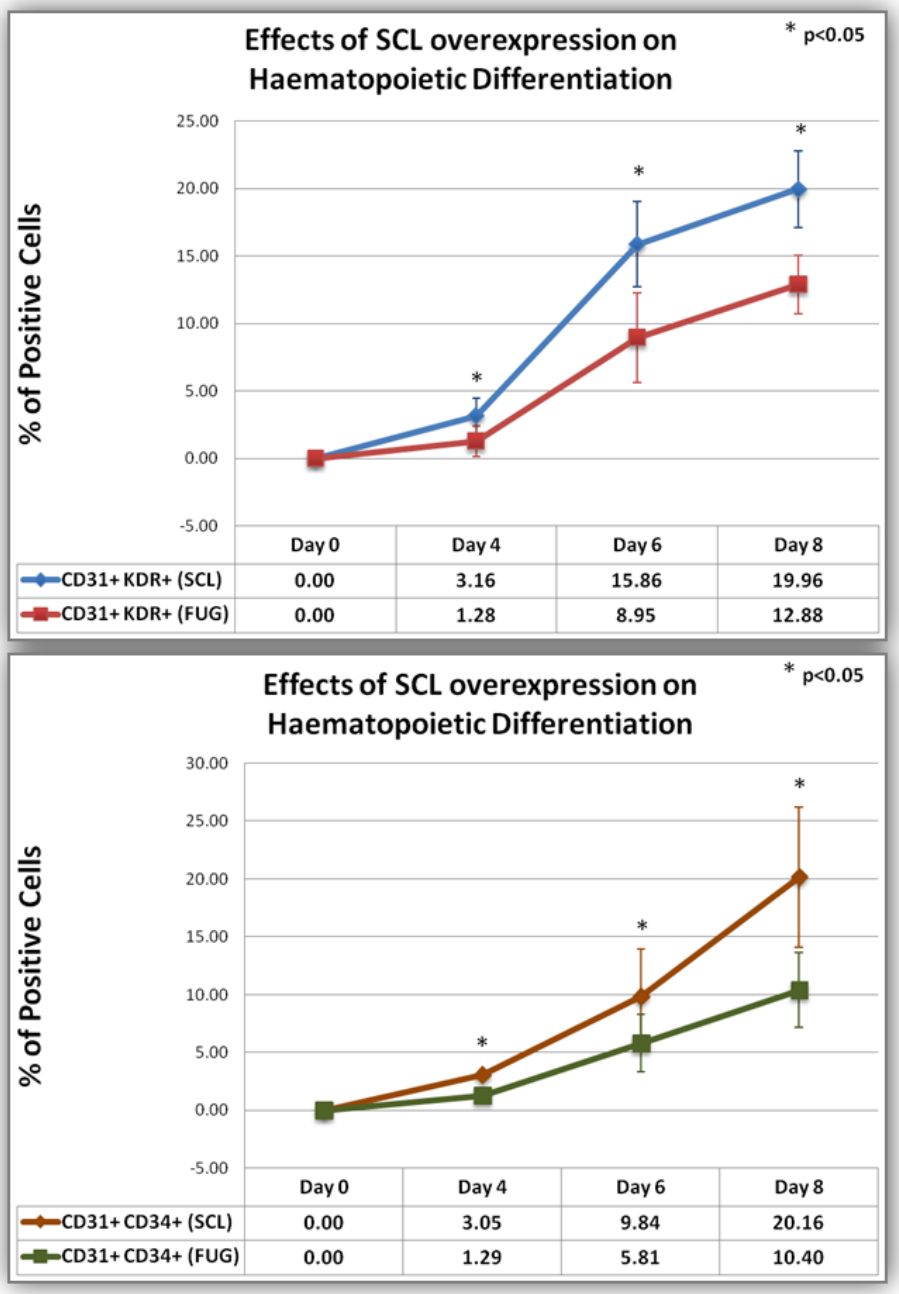


Figure 76: Flow Cytometry results of the effects of *Scl* over-expression during haemogenic progenitor specification

Flow cytometry results comparing the effects of *Scl* over-expression (SCL) with those of the control (FUG) shows significant enhancement in the percentage of haematopoietic cells identified by the markers CD31, CD34, and KDR. The combined averages of H9 SCL and H1 SCL flow cytometry results were used in the calculation of SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. The error bars were calculated using the standard error of the mean with a sample size of three. The generation of p-values in the assessment of statistical significance testing was performed using a student's t-test for three replicates of each cell line.

The effects of *Scl* over-expression on haematopoietic differentiation of HESCs was further assessed by the analysis of qPCR data of a number of pluripotent (Figure 77), lineage-specific (Figure 78), and haematopoietic genes (Figure 79). Analysis of the expression profile of pluripotent and self-renewal genes *OCT4* and *NANOG* (in Figure 77) revealed significant differences between SCL and FUG samples, however the general trend of down-regulation (of *OCT4* and *NANOG*) remained constant during differentiation. The rapid decline of both *OCT4* and *NANOG* suggest that, in addition to promoting haematopoiesis, *Scl* may promote differentiation of HESCs through the down-regulation of pluripotent / self-renewal genes. The level of *GFP* expression has been shown to increase from day 0 to day 4, correlating with the dimmer appearance of the GFP seen under the microscope when sub-cultured as undifferentiated colonies compared to the brighter fluorescence observed during differentiation. A comparison between FUG and SCL samples showed no significant difference in *GFP* expression intensity, however in contrast, the *SCL* gene is over-expressed by about five fold during the course of differentiation in the SCL sample compared to the FUG control.

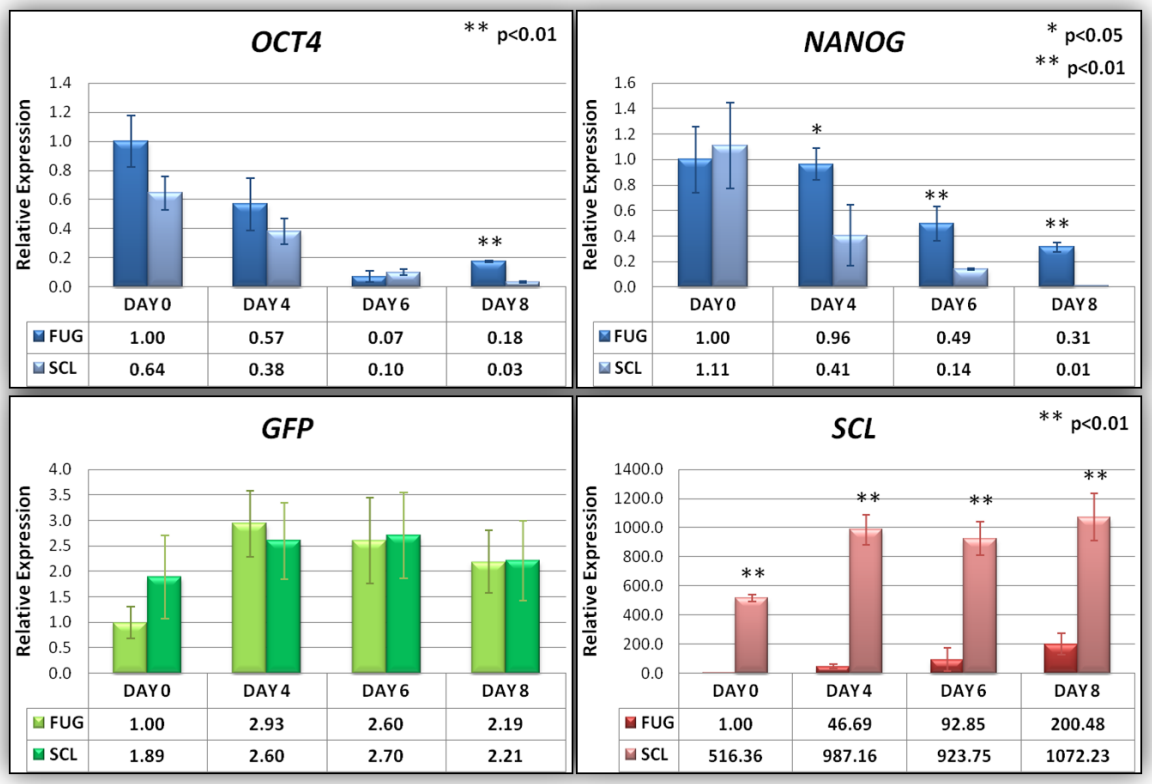


Figure 77: SCL / FUG qPCR Results of Gene Expression of Pluripotent markers and Transgenes (SCL and GFP)

These qPCR results show the effect of *Scl* over-expression on the gene expression pattern of a selection of pluripotency / self-renewal transcripts of which include *OCT4* and *NANOG*. The inclusion of *GFP* and *SCL* were used as controls. All samples were normalised by stably expressed genes *GAPDH*, *TBP*, *RPL13A*, and *SDHA* (expression plots shown in the APPENDIX). H9 SCL, H1 SCL. H9 FUG, and H1 FUG samples were run in triplicate and the averages of H9 SCL and H1 SCL qPCR results were used in the calculation of combined SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. Error bars were produced from the standard error of the mean with a sample size of three per cell line. The generation of p-values in the assessment of statistical significance testing was performed using a student's t-test (FUG compared with SCL) with p-values less than 0.05 deemed significant.

Inspection of the lineage-specific markers of mesoderm (*BRY* and *MIXL1*), endoderm (*AFP*), and ectoderm (*NESTIN*) by qPCR may provide clues to possible alternative roles (of SCL) in other lineages. To begin, the analysis of the mesoderm genes show a rapid increase of expression from day 0 to day 4, followed by a gradual decrease to day 8, with higher levels of expression of these genes in the SCL sample (on average) compared with FUG.

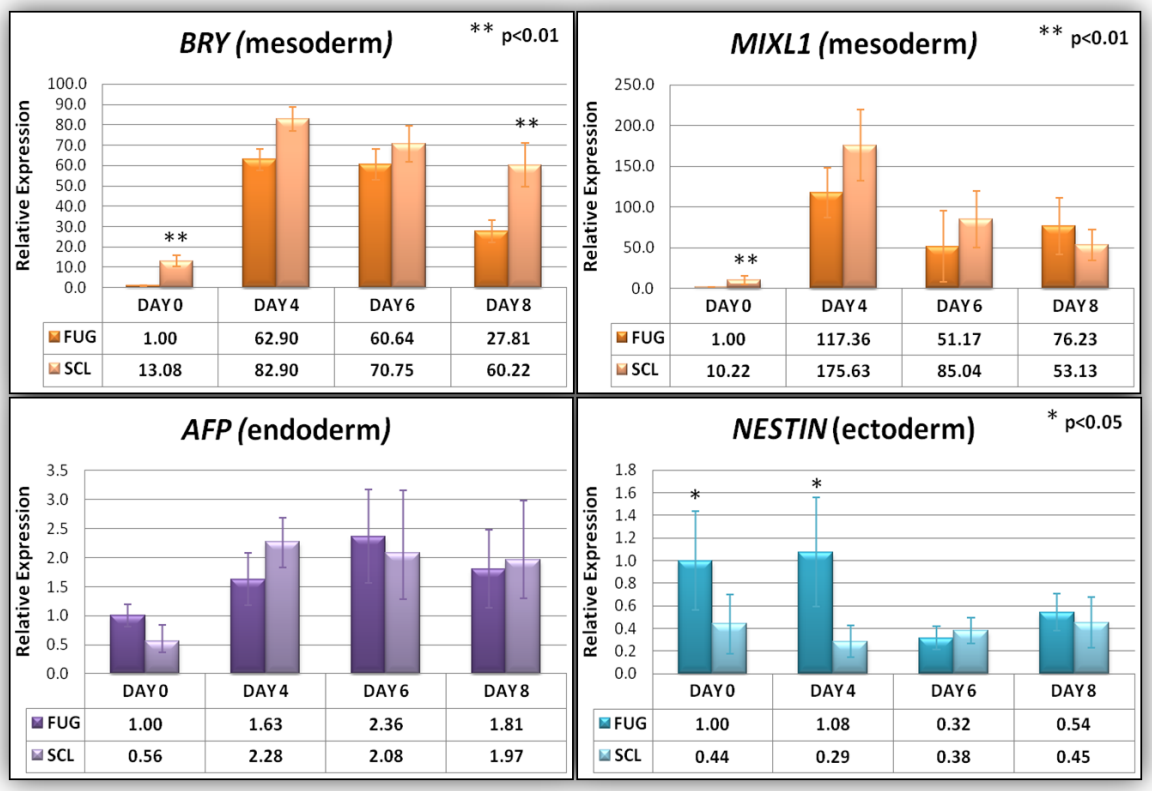


Figure 78: SCL / FUG qPCR Results of Gene Expression of Mesoderm, Endoderm, and Ectoderm markers

These qPCR results show the effect of *Scl* over-expression on the gene expression pattern of a selection of lineage-specific transcripts, including the mesoderm genes (*BRY* and *MIXL1*) ,endoderm gene (*AFP*), and ectoderm gene (*NESTIN*) through differentiation. All samples were normalised by stably expressed genes *GAPDH*, *TBP*, *RPL13A*, and *SDHA* (expression plots shown in the APPENDIX). H9 SCL, H1 SCL. H9 FUG, and H1 FUG samples were run in triplicate and the averages of H9 SCL and H1 SCL qPCR results were used in the calculation of combined SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. Error bars were produced from the standard error of the mean with a sample size of three per cell line. The generation of p-values in the assessment of statistical significance testing was performed using a student's t-test (FUG compared with SCL) with p-values less than 0.05 deemed significant.

It appears that the gene expression level of the endoderm gene, *AFP* (Figure 78) does not show any significant differences between FUG and SCL samples, with the overall expression profile displaying a slight increase in expression from day 0 to day 4, but this quickly plateaued over the remainder of differentiation. Analysis of the ectoderm gene, *NESTIN* (Figure 78), shows the basal level of expression in the SCL sample to be significantly lower than in the FUG control at day 0. The level of expression of the SCL sample across the subsequent days of

differentiation remains at a constant level, however the expression within FUG samples decreases rapidly from day 4 of differentiation before reaching similar levels to those in the SCL sample.

Analysis of the gene expression profile of the haematopoietic genes (Figure 79) clearly shows in general that SCL over-expression has a significant promotion effect (in the over half of sampled time points – 15/24) on haematopoietic differentiation efficiency. With these effects appearing to be greater in some haematopoietic genes than others, such as the gene expression change observed between FUG and SCL samples of *CD144* and *KDR*, which is significantly greater than the change seen within *CD34* and *CD31*. Due to the higher greater fold change difference (significance) observed between *CD144* and *KDR*, compared to the other haematopoietic genes, it seems logical to hypothesise that Scl may promote the pathways in the regulation of these two genes with greater effect than the others. Examination of *RUNX1* suggests that Scl does not significantly affect its expression during differentiation, but prior to differentiation the expression appears to be up-regulated at day 0 with the reasons for this unclear. With the exception of *RUNX1*, the results of these qPCR graphs overwhelmingly suggest that *Scl* over-expression does enhance the expression of haematopoietic genes and promote haematopoiesis albeit to varying degrees.

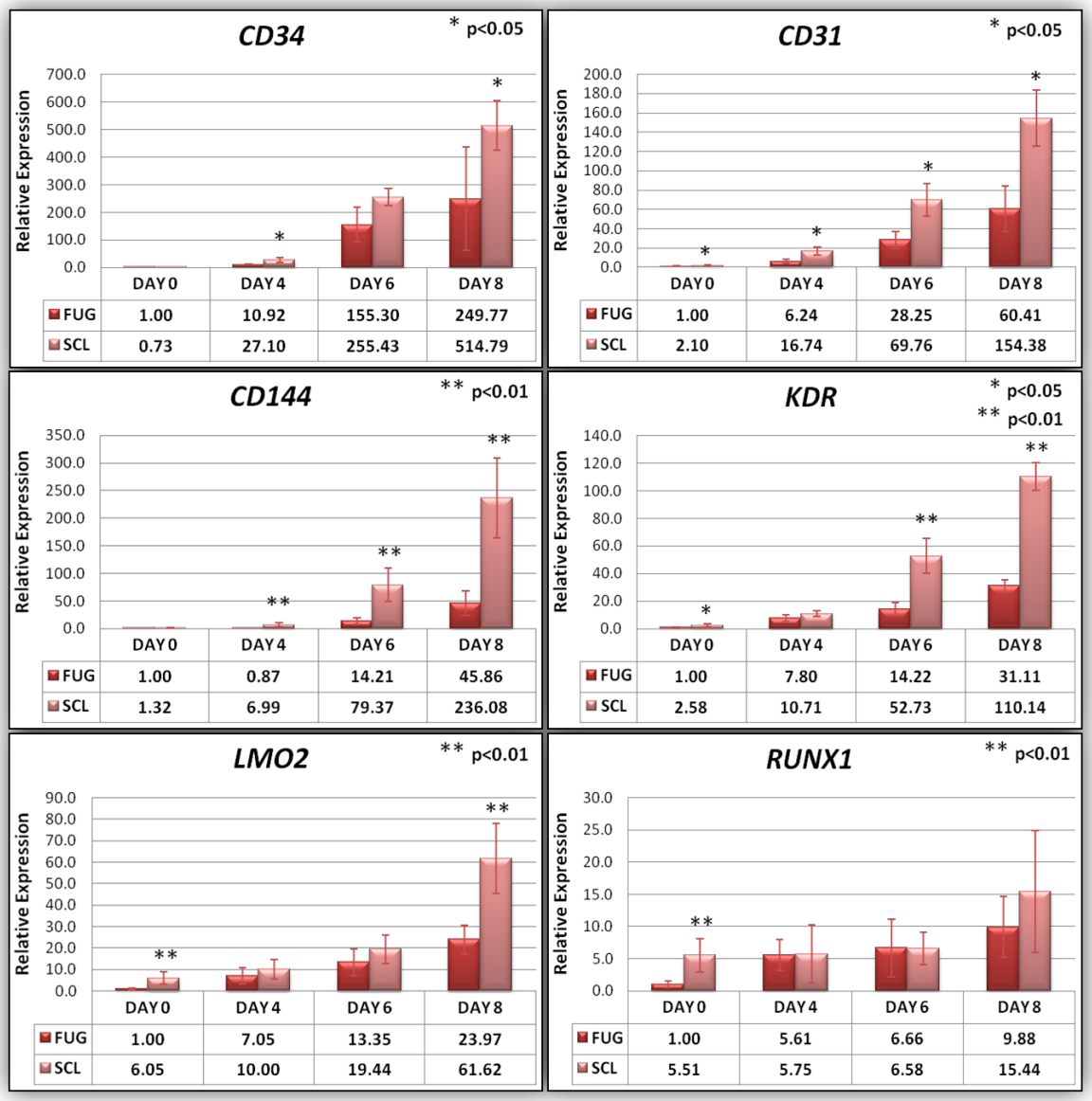


Figure 79: SCL / FUG qPCR Results of Gene Expression of Haematopoietic markers

The qPCR results of the haematopoietic genes all exhibit very similar gene expression profiles, which all begin gradually but as differentiation proceeds from day 6 on wards, the expression level increases between 100 - 400%. In all genes, it can be seen that the expression level in SCL (over-expressing) samples are greater than that in the FUG control samples. All samples were normalised by stably expressed genes *GAPDH*, *TBP*, *RPL13A*, and *SDHA* (expression plots shown in the APPENDIX). H9 SCL, H1 SCL. H9 FUG, and H1 FUG samples were run in triplicate and the averages of H9 SCL and H1 SCL qPCR results were used in the calculation of combined SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. Error bars were produced from the standard error of the mean with a sample size of three per cell line. The generation of p-values in the assessment of statistical significance testing was performed using a student's t-test (FUG compared with SCL) with p-values less than 0.05 deemed significant.

4.3.3 An essential role for SCL in erythropoiesis

The results of the colony forming assays (Figure 80) reveal an overwhelming preference for colony forming unit – erythroid (CFU-E) cell types (Figure 81) in the SCL over-expressing EBs. On average during day 4, 6, and 8 of differentiation, the percentage of CFU-E increased between 60 - 73% of the total population of CFUs, which represents a fold change of over 25. This result supports what has been published within literature regarding one the role of SCL as a potent enhancer of erythropoiesis.

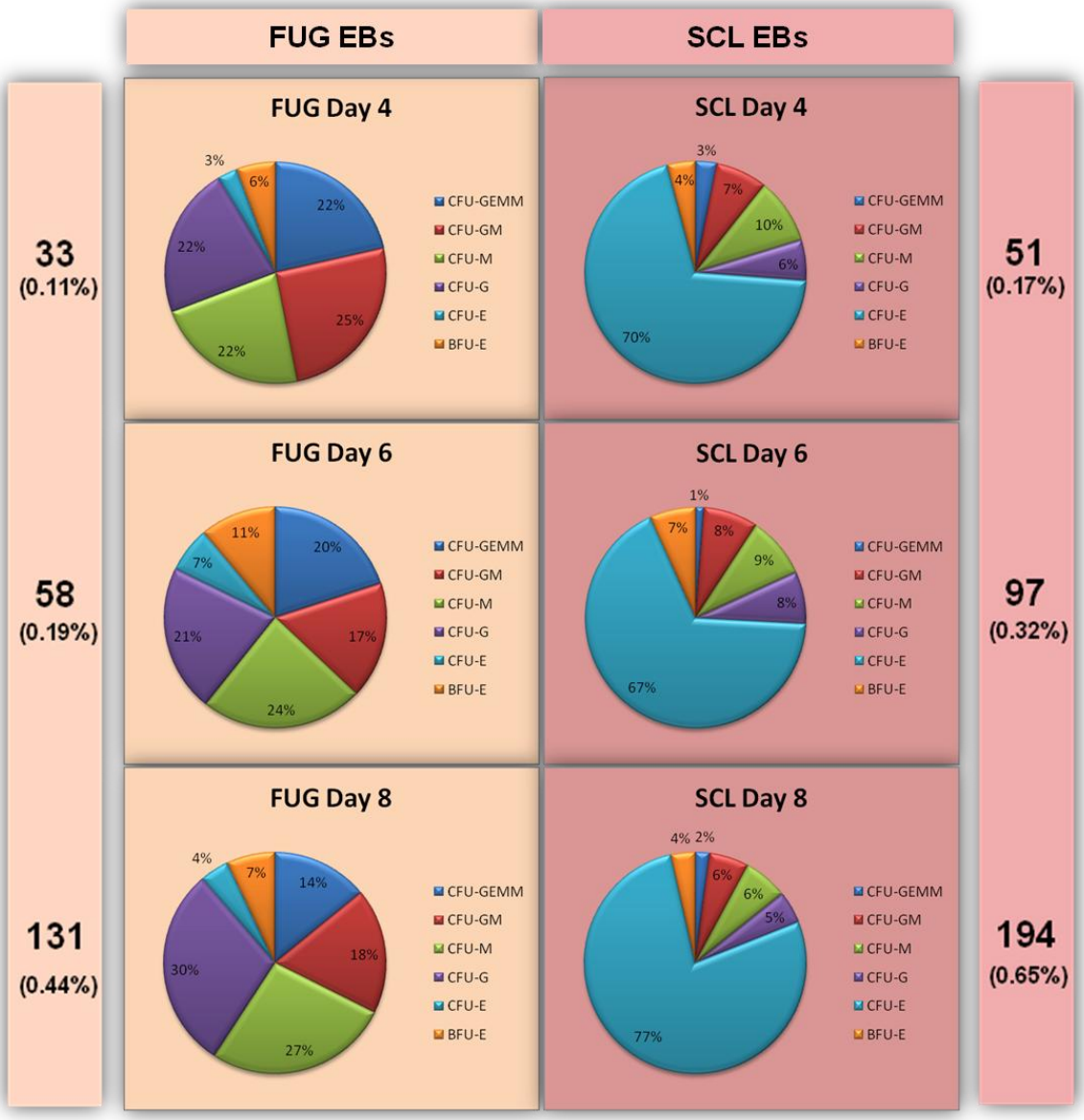


Figure 80: Distribution of Colony Forming Units from differentiated SCL and FUG cell types

The numbers and distribution of colony forming units (CFUs) produced from day 4, 6, and 8, haematopoietic differentiation of SCL-over expressing and FUG-control cells reveal a skew towards CFU-E development at the expense of the other cell types. H9 SCL, H1 SCL. H9 FUG, and H1 FUG samples were assayed in triplicate and the averages of H9 SCL and H1 SCL counts were used in the calculation of combined SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. The values displayed in the columns either side of the pie charts show the average number of total CFU recorded from an original input of 30,000 from FUG and SCL samples.

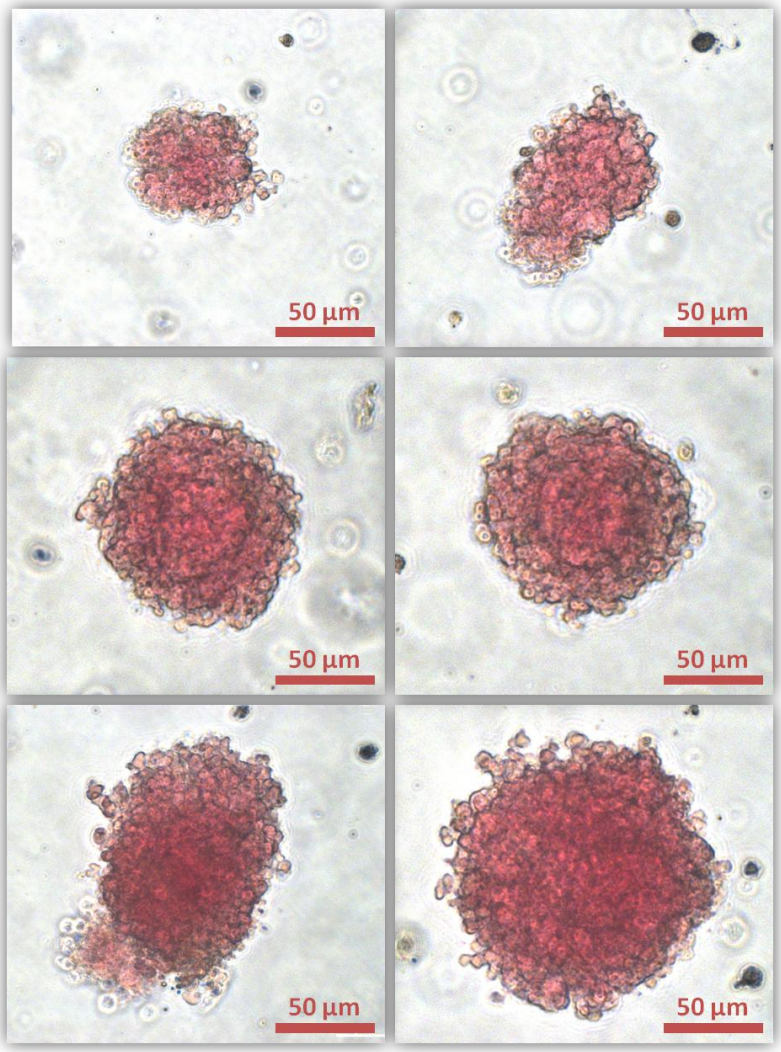


Figure 81: Images of Colony Forming Unit - Erythroid

Colony forming unit – Erythoid (CFU-E) cell types may easily be identified by the distinctive red colour produced by the haemoglobin. The majority of CFU-Es within the CFU assay appear to be composed of circular aggregates of cells in a range of sizes from 50µm to 150µm in diameter. These images were taken with an Axiocam HRC camera mounted on a Zeiss Aviovert 200M inverted microscope.

From examining the CFU-Es (Figure 81), it is hard to ignore the beautiful red colour beaming from within the cells aggregates, which leads to the question of which globin genes are expressed during development and does *Scl* over-expression affect globin expression, and if yes then which globins, if not all. As previously discussed in chapter 2 and 3, the *in vivo* stages of erythropoiesis may be tracked through the analysis of the globin expression patterns. The study of globin switching during embryonic-foetal-adult stages has been well documented (Chapter 3: Figure 66), and so by analysing the expression of the globins at any one point during differentiation *in vitro*, in theory (through differentiation as EBs capable of recapitulating developments *in vivo* [303, 388]) the equivalent stage of development *in vivo* may be identified. The qPCR results shown in Figure 82, shows the effect of *Scl* over expression

on the gene expression levels of the 6 globin genes (*HBA*, *HBB*, *HBD*, *HBE*, HBG, and *HBZ*) during the course of EB haematopoietic differentiation. An analysis of the qPCR results suggest that Scl has the greatest affect on *HBE* (the principal globin present during the embryonic stage of develop) at all stages of differentiation. Although *Scl* over-expression does to an extent result in an increase the expression of *HBA*, *HBB*, and *HBD*, this change is however not statistically significant during all days of differentiation. The effect of Scl on *HBG* and *HBZ*, appears to be chiefly negative or has little effect on these globins during this differentiation study.

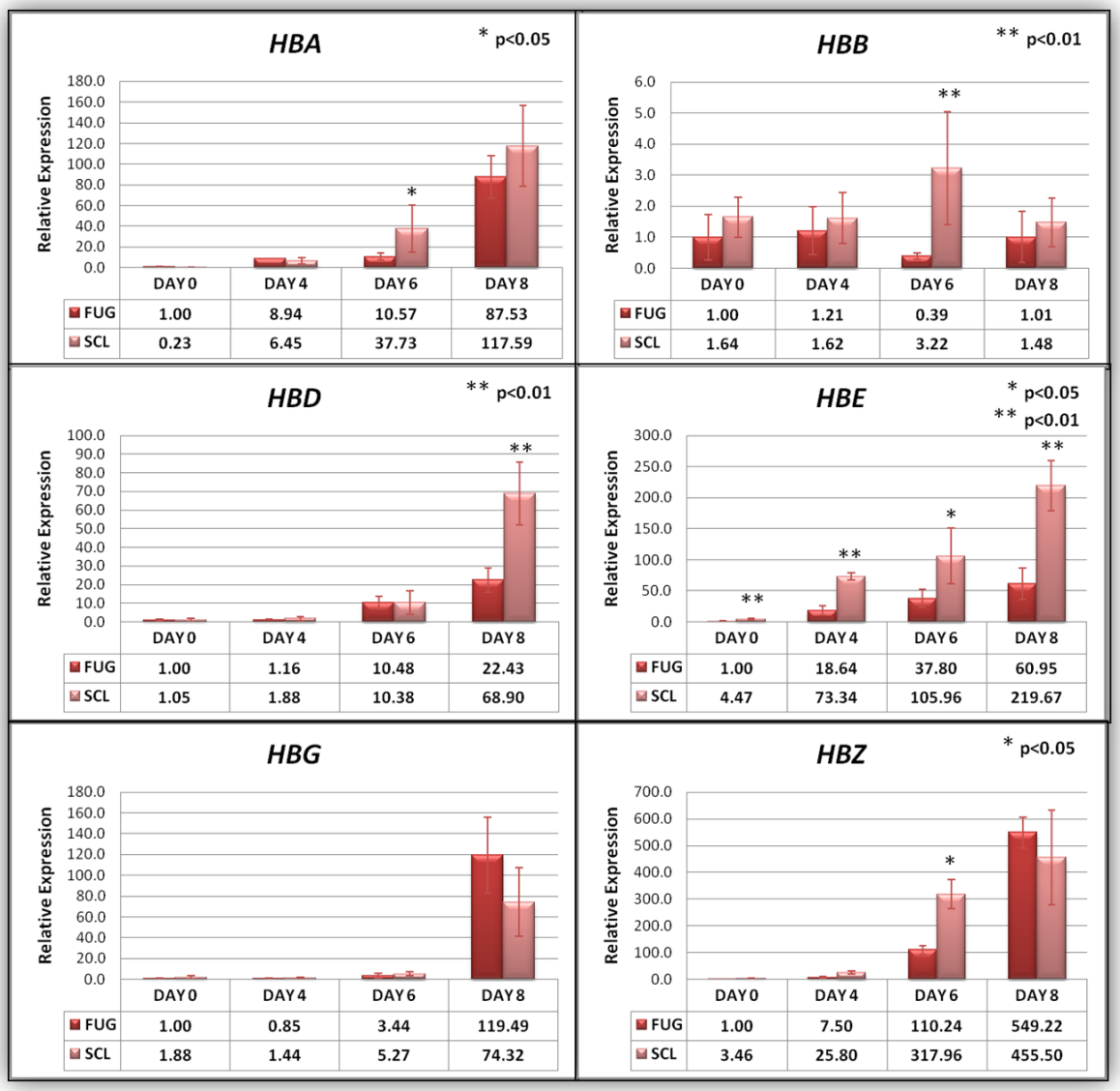


Figure 82: SCL / FUG qPCR Results of Gene Expression of Globin genes

The qPCR results show the expression levels of the globin genes during differentiation. All samples were normalised by stably expressed genes *GAPDH*, *TBP*, *RPL13A*, and *SDHA* (expression plots shown in the APPENDIX). H9 SCL, H1 SCL. H9 FUG, and H1 FUG samples were run in triplicate and the averages of H9 SCL and H1 SCL qPCR results were used in the calculation of combined SCL values, and the combined averages of H9 FUG and H1 FUG were used for FUG. Error bars were produced from the standard error of the mean with a sample size of three per cell line. The generation of p-values in the assessment of statistical significance testing was performed using a student's t-test with p-values less than 0.05 deemed significant.

4.4 Discussion

4.4.1 The Role of SCL in Humans and Marmosets

A study by Kurita *et al.* (2006) [389] examined the roles and effects of the over-expression of a number of haematopoietic genes (*scl, gata1, gata2, hoxB4,* and *lhx2*) in ESCs from a New World species, the common marmoset (CM; *Callithrix jaccchus*). Using similar methods of over-expression as in this project, Kurita *et al.* (2006) [389] also transduced their haematopoietic genes into ESCs using a lentiviral vector approach with *GFP* as a reporter gene. They discovered, out of all of the haematopoietic genes which they individually transduced into ESCs, only one was found to highly stimulate efficient haematopoietic cell differentiation, this was *scl*. This result correlates very well with what was observed in the over-expression of *Scl* in HESCs, however it must be noted that the haematopoietic differentiation efficiency of untransduced CM ESCs in this marmoset study appears to be several magnitude lower than compared with H9 / H1 or H9 / H1 FUG HESC lines. Unfortunately due to the lack of quantitative data within the marmoset study (due to the omission of FACS data and qPCR results displayed in the form of a gel image), only CFU efficiency may be quantitatively compared. CFU assay result show that CM ESCs after 14 days of differentiation generated CFUs at an efficiency of 0.05% (only CFU-M with absence all other sub-types), compared with 0.34% with H9 HESCs (Chapter 2: Figure 26; displaying the complete range of CFU sub-types) which were differentiated for 8 days, representing almost 7 fold difference in efficiency. This difference was confirmed by the qPCR results of the marmoset study which does not show a visible band for the early haematopoietic marker, *CD34,* on the agarose gel until day 18 of differentiation, compared to expression in H9 / H1 HESCs appearing from day 2 (Chapter 2: Figure 29; displaying a 6 fold increase between day 0 and day 2). The marmoset study described a number of different differentiation methods which include monolyer co-culture with mouse OP9 stromal cells and EB, however they discovered that differentiation with OP9 co-culture resulted in the absence of haematopoietic cells (MESC were used as a control), but the EB approach successfully induced haematopoietic differentiation.

From these preliminary results, it appears that the basic protocol for the differentiation of CM ESCs has not been properly addressed or established. However, post transduction with *scl* resulted in a 5-fold increase in haematopoietic differentiation efficiency, which represents a significantly greater increase in enhancement than compared to that observed between FUG and SCL in HESCs. In addition, the *scl* over-expressing CM ESC lines were able to give rise to the full range of CFU subtypes and by May-Giemsa staining, haemoglobin immunohistochemistry,

and double–esterase staining, it was demonstrated that the complete multiple lineages of haematopoietic cells were produced. It may be concluded from these studies that *scl* plays a very important role in the enhancement of haematopoietic differentiation, however within CM ESCs, *scl* appears to function to rescue haematopoiesis rather than to enhance normal haematopoietic development. The striking difference between these two studies (*scl* CM ESCs and *Scl* HESCs) is the role of *scl* in the promotion of erythropoiesis, which has not been demonstrated in *scl* CM ESCs, and this may possibly be due to the absence of other essential factors that are perhaps already expressed in *Scl* HESCs. Perhaps it would be of interest to co-transduce known erythroid co-factors *gata1* and *lmo2* together with *scl*, and observe if these two factors are the missing pieces to erythroid enhancement. In summary, these CM ESCs results have thoroughly supported the findings in our experiments of role of *scl* in the enhancement of haematopoietic differentiation and its importance for *in vivo* development in both species, however as marmoset as a model organism, at this stage and under these differentiation conditions are not truly representative of HESC differentiation

4.4.1 The Identification of SCL induced Globin Expression

It is clear from the results of this experiment that *Scl* plays a major role in erythropoiesis, and this was demonstrated by the CFU assays whereby the over-expression of *Scl* resulted in over 25 fold greater efficiency of CFU-E cell type formation. However, the exact mechanisms of *Scl* in the regulation of the various globin genes are unknown, but the results do suggest that *Scl* plays an significant role in the expression of *HBE* which plays an integral part in HbE Gower 1 (HBZ/HBE) and HbGower 2 (HBA/HBE) embryonic haemoglobin complex. The role of Scl in the regulation of expression of the other globins remains unclear, but this may be because *Scl* does not function alone but rather in a complex with a number of other proteins (LMO2, LDB1, LMO2, GATA1) [381] [390]. Comparison of the globin expressions induced by scl in marmosets [389] show the up-regulation of *HBE* (embryonic), *HBG* (foetal), and *HBB* (adult). This result adds support to the expression of *HBE*, but no correlation can be made with *HBG*, and *HBB* appears to be significantly at day 6 of differentiation only with the biological significance of this for the time being unknown. A possible reason may be that these qPCR results were obtained from whole EBs (containing 0.17-0.62% CFUs and CFU-E represents 70% of this), which may cloud the overall picture of true globin expression due to the heterogeneous population of cell types present. An improved method of assessing globin identity and frequency would be to directly sample the CFU-Es produced from the CFU assay, but as this mixed population of CFU-E may contain a percentage of enucleated erythrocytes, qPCR cannot

be used in this instance for quantitative identification. An alternative method would be to use cation exchange high performance liquid chromatography (HPLC) to identify and measure the various globin proteins that are present in the CFU-Es [391]. We have already begun a collaboration with Dr. Jo Mountford and her team in the University of Glasgow to utilising her HPLC machine in the identification of the globins within the CFU-Es at different stages of differentiation, and currently in the process of collecting enough CFU-E (approximately 0.5 million per time point) for the analysis.

4.4.2 From HESCs to Blood with help from *SCL*

Millions of units of blood are transfused in hospitals around the world each year, and the prospect of producing an unlimited supply of the essential component of blood, the oxygen transporting erythrocytes, from HESCs has been the Holy Grail ever since stem cells were discovered [392, 393]. Such is the potency of *SCL* in the generation of erythroid progenitors in CFU assays that this investigation will be continuing with these *SCL* over-expressing cell lines with the aim of improving the efficiency of CFU-E production. We are currently investigating a novel erythropoietic differentiation protocol in the generation of erythroid progenitors from HESCs, and preliminary results by our collaborator using normal HESC has shown that over 80% of the differentiated cells were CD34-positive after the 22 days differentiation period. It would therefore be of great interest to differentiated the *SCL* over-expression HESC lines within the same media and observing the enhancement (if any) of this differentiation.

4.4.3 Potential of *SCL* in the haematopoietic differentiation of HESCs

This study demonstrates the potential of *SCL* in the enhancement of haematopoietic differentiation and underlines its importance in the many roles it plays during differentiation. Referring to a study by Mead *et al.* (2001) [383] whom over-expressed *SCL, LMO2,* and *GATA1* in Xenopus and discovered a far greater synergistic capacity in the enhancement of haematopoiesis compared to SCL over-expression alone. It is clear from the results generated from this experiment that SCL alone cannot direct HESCs efficiently into HSCs, and the next step maybe to investigate and over-express the protein complex co-factors *LMO2* and *GATA1* in combination with *SCL* in following on from the research by Mead *et al.* (2001) [383]. However a simpler method would be to discover the master haematopoietic regulator (if one should exist) which would normally *in vivo* activate *SCL*, *LMO2* and *GATA1*. However, one such

master regulator has already been proposed within literature by Liu *et al.* (2008) [379] called friend leukaemia integration 1 (FLI1), which may induce a host of haemangioblast genes including *SCL, LMO2, GATA2, ETSRP*, and *KDR*. This gene was also picked up within the microarray screen for novel haematopoietic transcripts in chapter 3, however because its role within haematopoietic development was already known, this gene was not examined until its relationship with *SCL* brought it to attention. It would therefore be of immense interest to investigate the role of this FLI1 in its capacity to promote and enhance haematopoietic differentiation from HESCs.

4.5 Future Directions

The results from this chapter demonstrate the important role in which Scl plays in haematopoietic specification and development. However, as with most results, this often leads to a multitude of additional questions and queries in order to gain further insight into the precise mode of action (and of the associated gene targets) at various stages of development. The approach used in this study constitutively over-expressed the *scl* transcript within human embryonic stem cells (HESC) and the changes in differentiation were assessed based on haematopoietic marker expression by flow cytometry (protein level), and a glimpse of the genetic impact was revealed through qPCR analysis. However, as previously described at the beginning of this chapter, the mouse homolog of the *scl* gene was used within the *scl*-overexpressing HESC lines in this study. Although the human *SCL* and mouse *scl* genes are 93% identical with the functional core bHLH DNA binding domain conserved across multiple species (including human, mouse, chick, xenopus, zebrafish), it is imperative that this study be repeated with the human *SCL* gene. A minor modification should to used Flag (N-DYKDDDDK-C)-tagged *SCL* within the vector which will facilitate the identification of the SCL protein. (A number of groups, including those of Bertie Gottgen's, Pablo Menendez's, and Catherine Porcher's, have commented on the lack of a good SCL antibody to this date). The results expected from using the human form of *SCL* in a repeat of the haematopoietic differentiation experiment described in this chapter would be expected to yield very similar results due to the high level of homology observed between the two species.

With regards to the mechanistic role of SCL, cell numbers may be counted at each of the differentiation time points (Day 0, Day 4, day 6, Day 8, etc.) and be compared with the control cell line (FUG) to determine if SCL is either involved in haematopoietic specification or haematopoietic proliferation, or indeed both. The current theory as indicated by a number of studies suggested that SCL is involved in both specification and proliferation, but its role is not divided between the two processes at the same time, but rather switches depending on the stage of development [171, 177, 394]. However, even the role of SCL within one of these processes, such as specification, does not remain constant as its expression affects the differentiation capacity into certain lineages, with high *SCL* expression required for differentiation into erythrocyte, megakaryocyte, and mast lineages, and whereas low *SCL* expression is required for differentiation into monocyte and granulocyte lineages [395, 396].

The next approach would be the use of inducible vectors, which in this case means the ability to regulate *SCL* expression at different stages of the differentiation protocol in assessing

and discovering the important stages of *SCL* function. The first modification to the original differentiation protocol would be to switch *SCL* 'on' during the beginning of haematopoietic differentiation to day 4, 6 and 8, but then to switch it 'off' before placing these cells into methocult differentiation for colony forming unit assay. It is hypothesised that the number of haematopoietic progenitors would increase during differentiation from Day 0 to Day 4, to Day 6, and to Day 8 as observed in the original experiment when compared to the *SCL*-negative control. But turning 'off' *SCL* before plating these differentiated cells into colony forming unit assays, may result in lower quantities of colony forming unit erythrocytes and instead result in a more mixed and varied population of CFU-GEMM, CFU-GM, CFU-G, CFU-G, CFU-E as observed in the original *SCL*-negative control results. Thus the results of this experiment may provide evidence supporting the earlier hypotheses suggesting that SCL is involved in both proliferation and specification, by demonstrating that SCL increases the proliferation of haematopoietic progenitors and that it is also required later for the specification of erythrocytes [171, 177, 394]. The use of a *SCL*-inducible vector would also mean that the critical stage of SCL action may be pinpointed during differentiation by switching it 'on' and 'off' at varying points during differentiation, providing further clues to the role and effect of *SCL* on *in vitro* haematopoietic differentiation from HESCs. An opposite approach would involve the stably or transiently knocked-down of *SCL* expression through use of shRNA/siRNA to reveal roles that may not be evident from the over-expression experiments. In addition, this approach would also show how important *SCL* expression is at each stage of haematopoietic development and which downstream genes depend on the transcriptional activating role of *SCL*. *SCL* knock-down studies in the mouse model has resulted in the complete obliteration of primitive haematopoiesis leading to death of the animal at 9.5 d.p.c. [169, 170]. This finding is also supported by SCL-deficient EB haematopoietic differentiation experiments with mouse embryonic stem cells, which resulted in the complete absence of blast colony formation, which adds supports to the role of *SCL* in haematopoietic specification and commitment [171]. Based on these findings, one would expect the results of *SCL*-knocked down haematopoietic differentiation experiments to show a dramatic reduction in the number of haematopoietic progenitors in combination with an absence in CFU-E/BFU-E within the colony forming assays.

As previously described, *GATA1* and *LMO2* genes are expressed and mediate their function during erythroid differentiation by forming a complex with *SCL* in the regulation of erythroid specific genes such as those belonging to the globin family of genes. It may be of interest to co-overexpress these two transcripts in combination with *SCL* to examine if this would affect the efficiency of erythroid progenitor generation and other effects on erythrogenesis. One could hypothesis that over expression of these accessory transcripts

within the same vector may increase the efficiency of erythroid specification, but the effect of this in haemangioblast specification from HESCs is less clear, however results from the microarray study has suggested that *GATA1* and *LMO2* may play a role during this early stage of development. It would be of great interest to also proceed down this route and investigate the partnership between these transcripts and their role during early haematopoietic development.

The results of this project have only provided a small glimpse of the diverse role of *SCL* in haematopoiesis. Within this section a number of strategies have been proposed that may provide some clues to aid in further the understanding of the complex fundamental mechanisms surrounding this important haematopoietic regulator.

SUMMARY

This study utilised the versatility and potential of human embryonic stem cells (HESCs) in combination with the power and scope of microarray technology, in a genome-wide screen for novel haematopoietic transcripts which may assist to creak open the door to the great mysteries of haematopoietic development just that little bit wider. The potential of HESCs in the generation of haematopoietic and endothelial cells *in vitro* has been well documented in numerous studies [156, 157, 262, 397] as well as demonstrated in this study, but the haematopoietic differentiation efficiency from HESCs to this date remains an area for significant improvement. It was hypothesised that through the comparative analysis of the gene expression differences between undifferentiated and haematopoietic differentiated cell types, novel genes may be discovered which play an important role during haematopoietic development, and provide an improved understand of the underlining mechanisms involved during each developmental stage. The understand of such key mechanisms will also pave the way for improved differentiation media protocol in the controlled and directed enhancement of *in vitro* differentiation from HESCs. This is of course the Holy Grail for all embryonic stem cell research, the ability to deftly generate, with pure efficiency, any differentiated cell type of the body, with the end target ultimately aimed for clinical and therapeutic purposes.

One of the principal aims within this study was to isolate the earliest population of haematopoietic cells during embryonic development, known as the haemangioblast, which are bipotent mesodermal progenitors with the unique capacity to give rise to both haematopoietic and endothelial cell types [143, 157, 160, 172, 178, 180-184, 243, 244, 246, 247, 398, 399]. Studies have shown that this population of cell may be identified through the combination of number of early haematopoietic and endothelial cell surface markers, which includes CD31, CD34, KDR, CD144, and CD117 [157, 181, 182, 400-402]. An essential element of this study was the acquisition of sufficient quantities of this population at different stages of haematopoietic differentiation for microarray analysis, and delving into the transcript level, comparing the gene expression see within the haemangioblast with the gene expression of undifferentiated pluripotent stem cells. By simply using fold change analysis in conjunction with significance thresholds, a group of unique differentially expressing genes may be identified for further investigation. On paper, this plan sounded very straightforward, however in practise and in reality, numerous practical and technical challenges were encountered during the differentiation process, which resulted in severe delays with the full microarray data only completed at the beginning of the third year of the PhD project when the analysis began.

Besides from the numerous outbreaks of mycoplasma in the tissue culture units, the occasional bouts of bacterial and fungal infections in the incubators, the primary reason for the delay was due to the poor efficiency of differentiation, which resulted from the foetal bovine serum (FBS) based standard differentiation media. This was as the name described, the standard method of differentiation of EB HESCs, whereby the removal of pluripotency-and self-renewal inducing factors (mouse embryonic feeders and basis fibroblast growth factor) from the media in combination with the addition of serum, resulted in the spontaneous differentiation of HESC and giving rise to cell representative of all three germ layers [3, 227]. However the haematopoietic differentiation efficiency was so poor that as previous described, in order to sort for the haemangioblast population using the five key markers, would require a few thousand 6-well plates of HESC cells. A search for haematopoietic-specific differentiation protocols and media revealed a novel approach by Keller's group (Kennedy *et al.* (2007)[157]), whereby a serum-free media was used and haematopoietic differentiation was directed by the addition of cytokines at various stages of differentiation in an attempt to mimic *in vivo* development. As previously described, these key cytokines were BMP4, bFGF, and VEGF, and were added to the differentiation media sequentially, inducing the HESCs to first become cells representative of mesoderm and then through to those of the haemangioblast. However, although haematopoietic differentiation efficiencies dramatically improved, the goal of sorting cells presenting all five haemangioblast markers was still elusive with numbers of plates required still in the hundreds. A number of optimisation were carried out in an attempt to maximise the efficiency and to reduce the number of variables. These primarily consisted of optimisations in the cell line and the density of which the EBs were plated, and the results suggested that neither appeared to have a significant effect on the haematopoietic differentiation efficiency. EB size is an obvious variable that unfortunately was not addressed or controlled at the time primarily due to the absence of published protocols surround this area of EB optimisation with the first report by Elefanty's group (Ng *et al.* 2008 [253]) in the formation of spin EBs that contain 3000 cells for optimal mesoderm differentiation. Regretfully, it appeared from the analysis of the results from Ng *et al.* 2008 [253], that the advantages gained from this optimisation alone do not appear to enhance haematopoietic differentiation significantly enough to warrant the time and expense in the generation of sufficient quantities of uniform EBs. An attractive alternative was the addition of a fourth cytokine called Activin A, which is already known to play a role in the maintenance of pluripotency and self-renewal of HESCs, as it is one of the key factors which MEFs secrete large quantities of into the media. Nostro *et al.* (2008) [203] however has shown in mouse that during differentiation Activin switches roles, and in combination with Nodal/Wnt/BMP pathways assists in the formation of the primitive streak and induction of KDR+ mesoderm.

Unfortunately, due to financial constraints, only one differentiation experiment was performed with HESC H9 cell line, and the results did not produce any significant results, but as previously discussed, the role of this cytokine in haematopoietic differentiation definitely deserves a more thorough investigation in the future. As the efficiency of haematopoietic differentiation was still very low and the number of plates of HESCs required to sort the tiny population of haemangioblasts was still much too high. The purity of the haemangioblast population that was sorted for the microarray comparison was unfortunately sacrificed in an attempt to increase the percentage of sortable cells. Studies has shown that KDR and CD31 were the two markers that are indispensible in the identification of the haemangioblast population as these the earliest markers that are presented during development [75, 152, 181, 182, 403], however it also worth noting that not all cells that expressed *KDR* and *CD31* are haemangioblasts. As the haemangioblast represents a transient population during development, a possible reason for the detection of such low numbers (using the five-haemangioblast cell surface markers) maybe because this population does not exist stably in this form for long during differentiation and immediately take on the characteristics of the haemogenic endothelium. This reason was supported by the results of the KDR+CD31+ CFU assay, which questioned the reliability of using all five markers in the identification of the haemangioblast, because a much greater number of CFUs were observed in the assays than expected based on the percentage of cells presenting the five-markers of the haemangioblast. However, what the results of the KDR+CD31+ CFU assay also reiterate is that the use of just two markers in the identification of the haemangioblast is not sufficient. With the maximum efficiency of KDR+CD31+ in the formation of CFUs shown as 3.28%, which although it is a significant enhancement from the 0.34% achieved from unsorted EBs, it is still a long way from the 100% that would be expected from pure haemangioblasts. It was concluded that KDR+CD31+ represented an enriched population of the haemangioblast (20 fold more enriched compared to unsorted cells of the EB), and it was hypothesised that the majority of differentially expressed transcripts in this population would be haematopoietic-related but because this sorted population is not truly representative of pure haemangioblasts, non-haematopoietic transcripts may also be differentially expressed.

In the analysis of the microarray, which consisted for 4 different stages of differentiation (unsorted Day 0, KDR+CD31+ sorted Day 4, KDR+CD31+ sorted Day 6, and KDR+CD31+ sorted Day 8) a number of approaches were used in the identification and separation of transcripts that are haematopoietic-related from those that are not. The core gene list for the selection of novel haematopoietic transcripts, **Day0toDay4Up2fold+HaemSim+p0.5** (APPENDIX: Table 39), was composed of all transcripts that increased in expression from day 0 to day 4 (with a fold

change of 2 or higher and with a p-value of 0.5 or lower) combined with a selection of transcripts that displayed similar expression profiles to the five haemangioblast genes. Within this gene list of over 3000 transcripts, the expression of *Sclerostin* (*SOST*) was 197 fold over-expression from day 0 to day 4 of differentiation, and as this gene had not been identified in the published literature as playing a role in haematopoiesis, vasculogenesis, or angiogenesis, it was an potential haematopoietic candidate. *Sclerostin* encodes a cystine knot-containing glycoprotein, which was identified in sclerosteosis patients, where a loss-of-function mutation in this gene resulted in progressive bone overgrowth [297]. Studies have implicated Bone Morphogenetic Protein (BMP) 2, 4, and 6 in the regulation of *SOST* expression in mouse and human osteoblasts,[350] and that bone mass regulation was regulated by BMP through SOST by the inhibition of the canonical Wnt signalling pathway [351]. It may be hypothesised that SOST would function in a similar role to BMP4, which is an essential component of the optimised haematopoietic differentiation media and the introduction of SOST into the media may assist to inhibit non-haematopoietic gene expression through the canonical Wnt pathway. It would be of great interest to investigate this in greater depth and carry out an *in vitro* haematopoietic differentiation to test the effects of SOST.

The validity and reliable of this microarray experiment in using KDR+CD31+ sorted cells in the pursuit of novel haematopoietic transcripts, increased when gene ontology and pathway analysis showed that the up-regulation of a large proportion of transcripts were involved in blood vessel and vasculature development. Although not quite haematopoiesis, this result correlates perfectly with the onset of the haemogenic endothelium, which acts as an intermediate between the formation of haemangioblasts, and the generation of haematopoietic cells [192, 193]. Analysis of day4 to day 8, shows the activation of a number of erythropoietin pathways and together with sequential expression of globin genes suggests the beginnings of erythropoiesis, and hence primitive haematopoiesis. These results further support the reliability of this microarray and adds confidence to all novel discoveries, as it appears that these *in silico* results correlates well with what was already known through *in vivo* models.

EPO and hypoxia are the two other factors identified from the microarray analysis which may be worth investigating. It was noticed that a large number of transcripts (minus EPO) of the EPO-induced PI3K/AKT pathway (Figure 65) was over-expressed during the course of the haematopoietic differentiation experiment. It may therefore be hypothesised that the addition of EPO cytokine into the differentiation media may enhance erythropoiesis through the positive activation of the EPO-induced PI3K/AKT pathway, but may also possibly positively

affect haematopoietic development in general. This dual role was demostrated by *Scl* over-expression (with *in vitro* experiments) whereby overall haematopoietic differentiation efficiency increased but also significantly promoted erythropoiesis, however Scl is a multifunctional transcription factor and EPO is a receptor binding ligand, its range of regulatory roles may not be as diverse as Scl. Within the human adult, EPO release from the kidney / liver is triggered by the detection of low oxygen levels via hypoxia-inducible factors, which in turn stimulates the overproduction of red blood cells within the bone marrow [339]. It may be of interest to investigate this during HESC differentiation to observe if hypoxia may over-express EPO and have an effect on haematopoietic differentiation efficiency.

Through the analysis of the microarray data, and by the process taken in the identification and separation of transcripts with expression patterns that changed significantly from those which does not, revealed a subset of highly up-regulated transcripts with an expression profile that didn't change over the course of the 8 days of differentiation. These stably expressed genes (APPENDIX: Table 38) may be excellent candidates for used as reference genes in the normalisation of quantitative analyses between different samples (for example, qPCR) from haematopoietic differentiation experiments, which will be more reliable and consistent than common reference genes which are used to for every cell and tissue type of the body. A brief screen of this list of stably expressed genes revealed an overwhelming number of ribosomal proteins, and of which a substantial number have already been implicated as known stably expressed reference genes, playing a generic role in protein synthesis in every cell of the human body [324]. The use of these transcripts in the normalisation of qPCR experiments will improved the accuracy of the analysis especially if changes of expression between samples are small, then the availability of consistent, reliable haematopoietic stably expressed transcripts become of greater importance.

The final part of the microarray analysis reveals the over-expression of Scl and although there has been countless studies demonstrating the crucial roles it plays in the specification of the mesoderm and the formation of the haemangioblast [358], in the formation of blood vessels [359], and in erythropoiesis [360] in a number of species, there has been limited studies of its effects on haematopoietic differentiation. Our colleague, Dr. Nicholas Slater (who has since left), had generated a number of *Scl* over-expressing HESC lines, so this presented a great opportunity to test out one of the transcripts from the list of potential haematopoietic candidates (albeit quite a well studied haematopoietic master regulator). The results speak for themselves, with flow cytometry results showing enhanced haematopoietic differentiation, which was supported up by qPCR data showing significant up-regulation across

haematopoietic markers. The generation of CFU showed slightly high numbers but the most dramatic difference was the overwhelming promotion of erythroid progenitors in the form of CFU-E accounting on average for over 70% of the total CFUs. The analysis of globin expression from the qPCR results alone proves inconclusive due to the hetergenous population of the EB, the true representation of globin expression will not be very clear, and hopefully though the use of HPLC a clearer picture of the kinetics of globin expression may be presented.

It was of interest to compare the results from this study with *Scl* over-expression in human ESCs analysis compared with a study with *Scl* over-expression in marmoset ESCs [389] and because of the close relationship of these two species, the result were expected to be similar. It was concluded from these studies that *scl* plays a very important role in the enhancement of haematopoietic differentiation in both human and marmoset, however within marmoset CM ESCs, *scl* appears to function to rescue haematopoiesis rather than to enhance normal haematopoietic development. With the big difference being that it appears the over-expression of *Scl* in HESCs was sufficient to promote erythropoiesis, where as in marmoset CM ESCs, there appears to be some factors missing. It would be of interest to co-transduce known erythroid co-factors *gata1* and *lmo2* together with *scl*, and observe if these two factors are the missing pieces to erythroid enhancement and it would be interesting to see what effect these co-factors will bring toward haematopoietic differentiation efficiency. This could quite possibly be the key link to the efficient and directed differentiation of functional erythrocytes from HESCs.

All of this future work and ultimate goals in perfecting haematopoietic differentiation sounds incredibly exciting, and with the new arrival of the phenomenon that is induce pluripotent stem cells (IPSCs). All of this acquire knowledge gained from embryonic stem cell would appear (from preliminary studies in our group) to be directly translatable. Meaning that the ethical issue which has always plagued human embryonic research may now once and for all be removed and allowing research to proceed with even greater pace in the race to generate patient specific tissues for therapeutic and clinical uses in the not so distant future.

APPENDIX

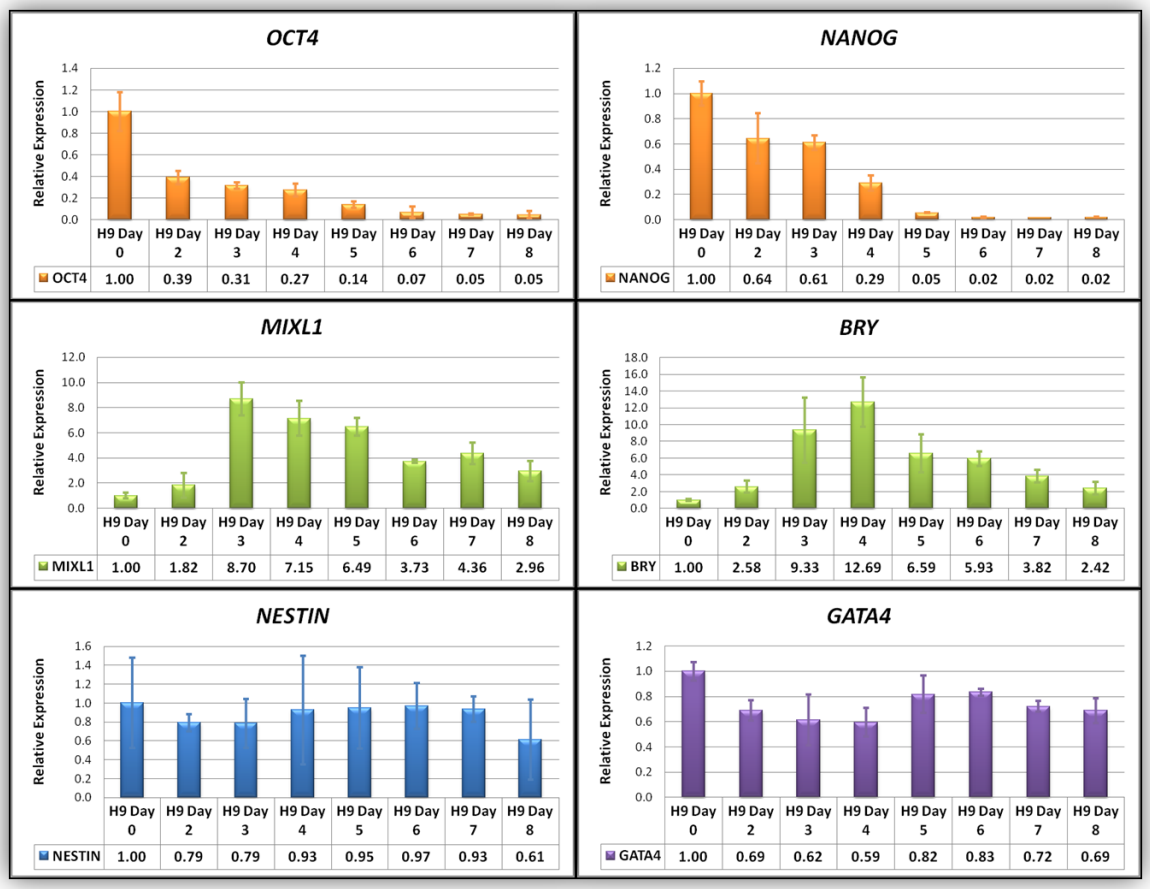


Figure 83: Real-time qPCR analysis of H9 cells from the eight day differentiation – pluripotency, mesoderm, ectoderm, and endoderm markers

A sample of genes were selected in the identification of the different lineage types (mesoderm, ectoderm, endoderm) resulting from differentiation during the 8 day period which includes genes representative of pluripotencv (*OCT4* and *NANOG*), mesoderm (*MIXL1* and *BRY*), ectoderm (*NESTIN*), and endoderm (*GATA4*). The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate with the average gene expression values plotted in a column chart.

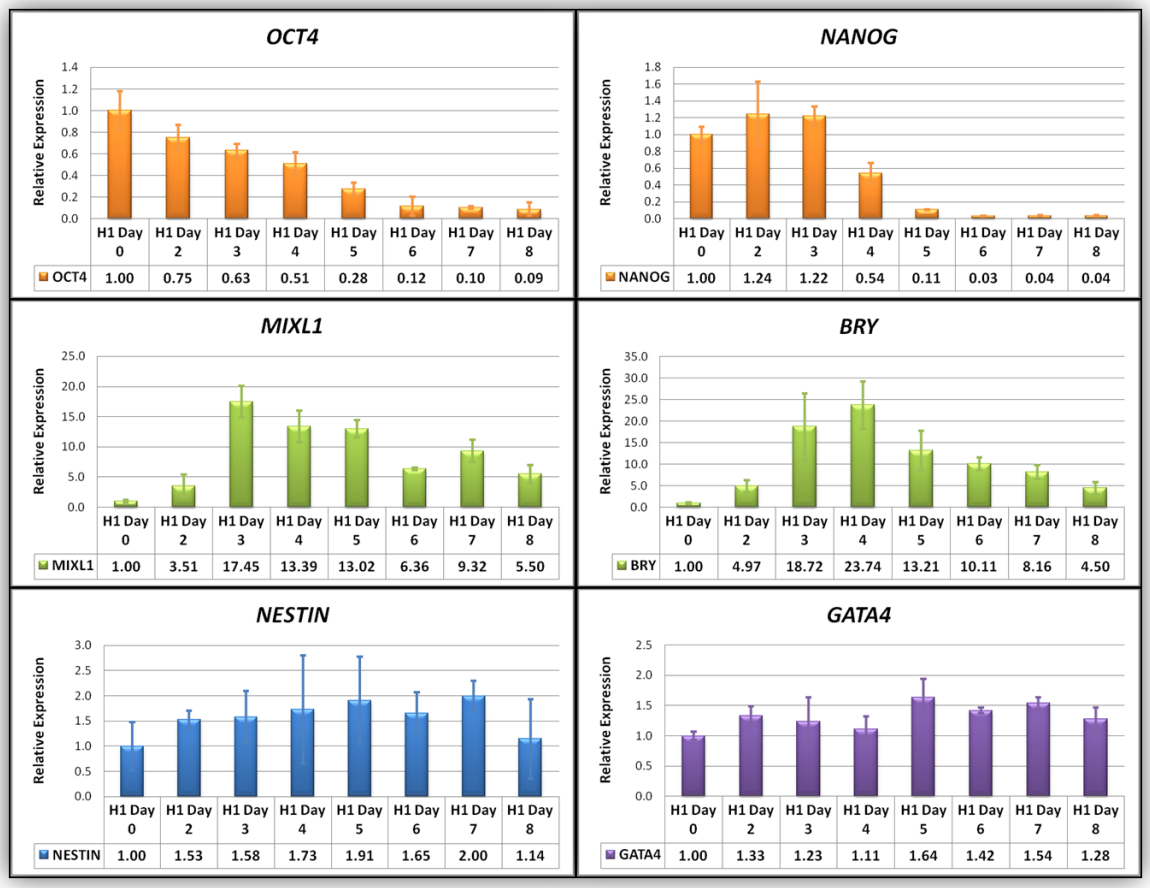


Figure 84: Real-time qPCR analysis of H1 cells from the eight day differentiation – pluripotency, mesoderm, ectoderm, and endoderm markers

A sample of genes were selected in the identification of the different lineage types (mesoderm, ectoderm, endoderm) resulting from differentiation during the 8 day period which includes genes representative of pluripotencv (*OCT4* and *NANOG*), mesoderm (*MIXL1* and *BRY*), ectoderm (*NESTIN*), and endoderm (*GATA4*). The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate with the average gene expression values plotted in a column chart.

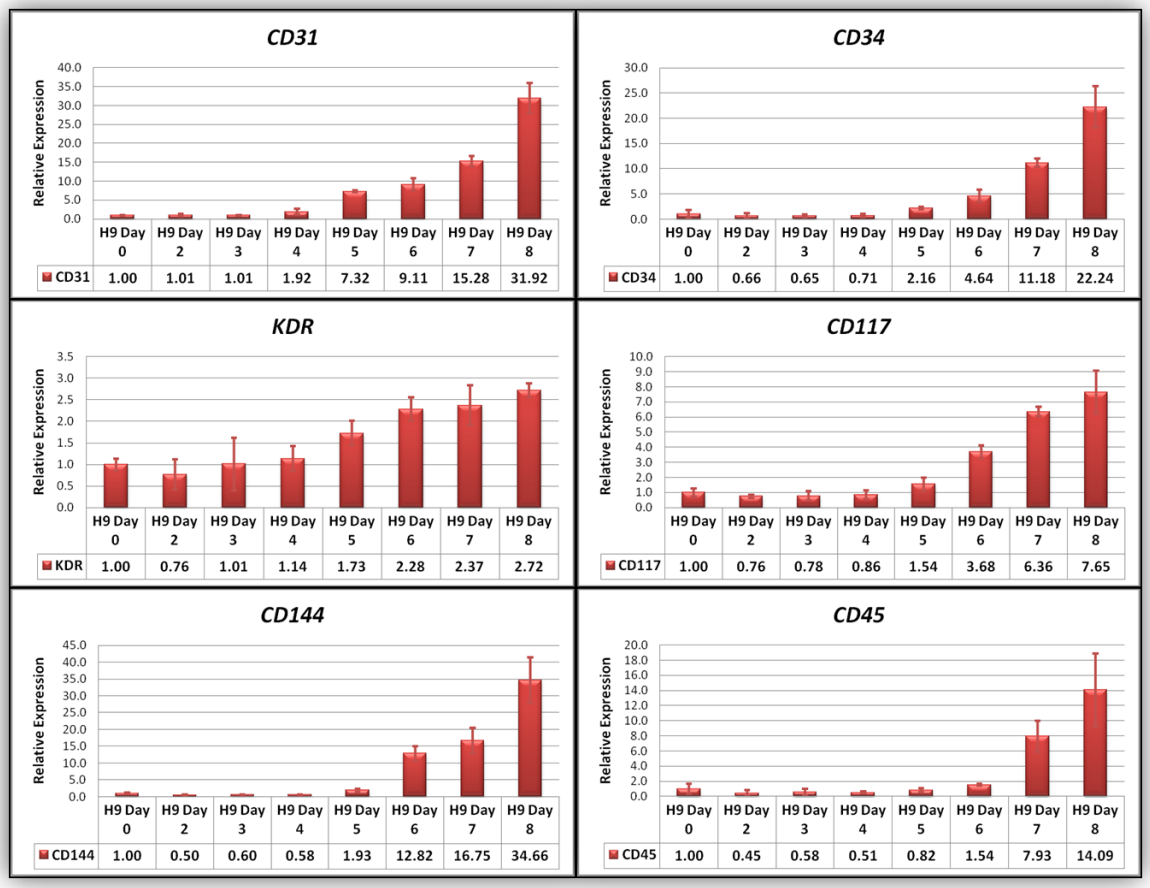


Figure 85: Real-time qPCR analysis of H9 cells from the eight day differentiation – haemangioblast markers

A sample of haematopoietic genes (*KDR, CD117, CD31, CD34, CD144,* and *CD45*) were selected in the analysis of haematopoietic differentiation during differentiation with the optimised differentiation media. The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate with the average gene expression values plotted in a column chart.

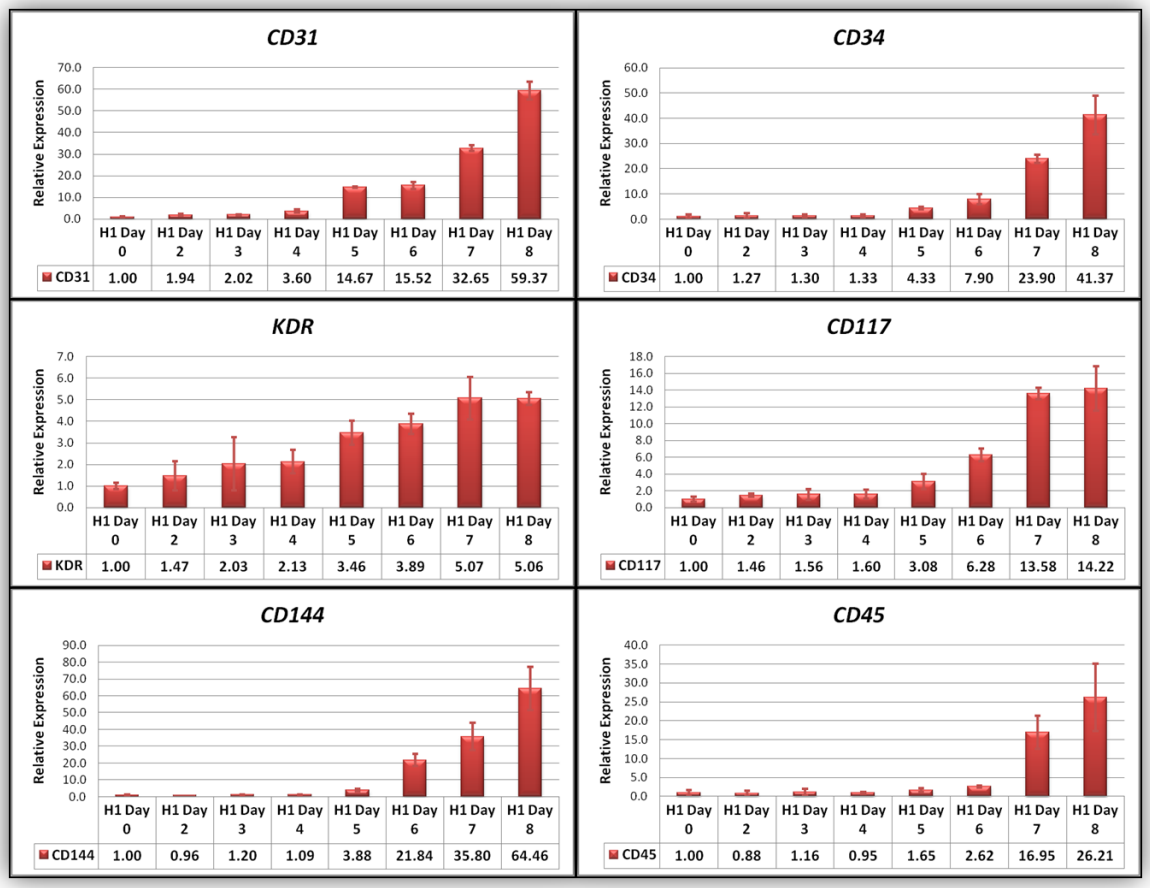


Figure 86: Real-time qPCR analysis of H1 cells from the eight day differentiation – haemangioblast markers

A sample of haematopoietic genes (*KDR, CD117, CD31, CD34, CD144,* and *CD45*) were selected in the analysis of haematopoietic differentiation during differentiation with the optimised differentiation media. The relative expression value for Day 0 for each gene was set to 1 and all values from all other samples were calculated with respect to this. Differentiation samples were run in triplicate with the average gene expression values plotted in a column chart.

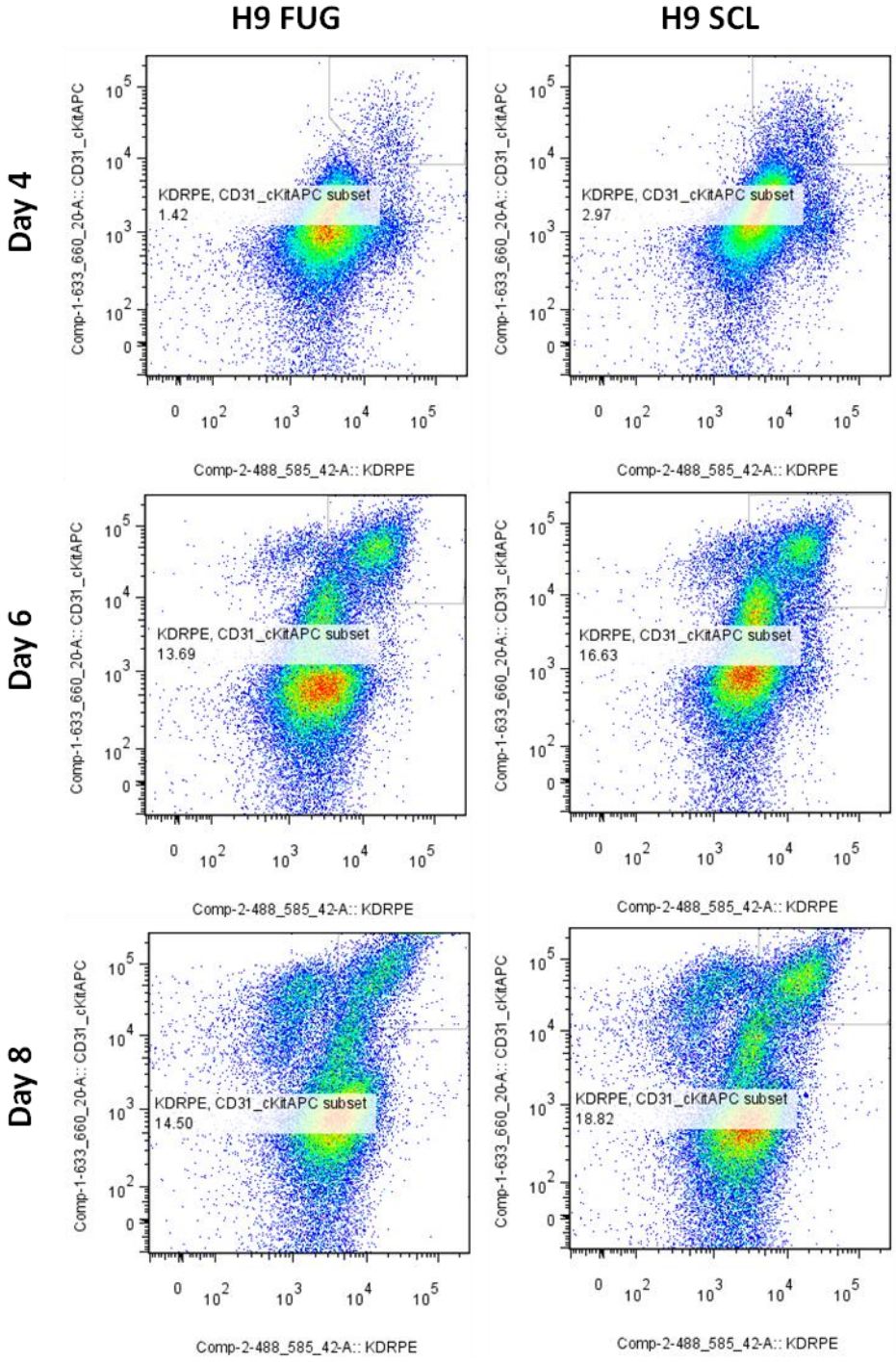


Figure 87: Flow Cytometry Plots of H9 Fug and H9 SCL showing CD31 and KDR

The flow cytometry plots display CD31 on the y-axis and KDR on the x-axis. The gates for the identification of the double positive CD31+KDR+ population were applied based on the isotype negative control plots.

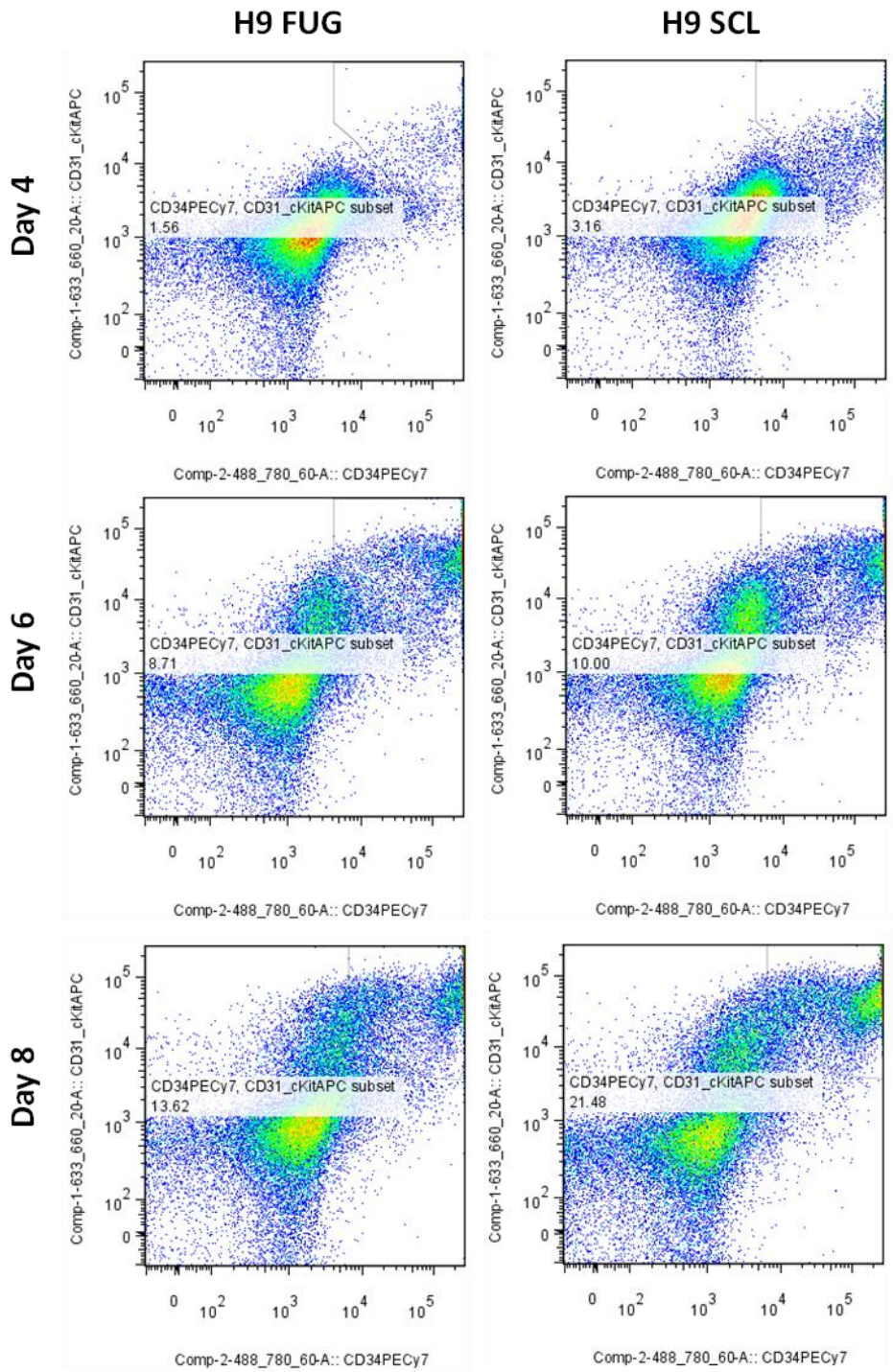


Figure 88: Flow Cytometry Plots of H9 Fug and H9 SCL showing CD31 and CD34

The flow cytometry plots display CD31 on the y-axis and CD34 on the x-axis. The gates for the identification of the double positive CD31+CD34+ population were applied based on the isotype negative control plots.

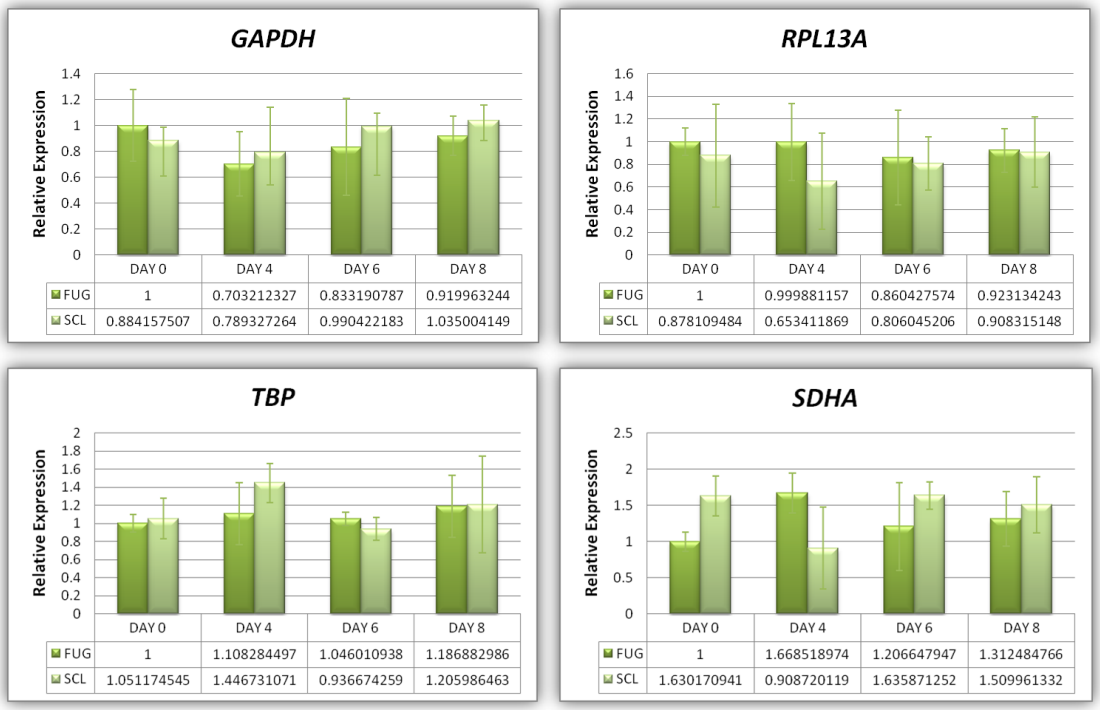


Figure 89: SCL / FUG qPCR Results of Gene Expression of Stably Expressed Reference Genes

The qPCR results show the expression levels of the four stable expressed reference genes used in the normalisation of the qPCR data. H9 / H1 SCL and H9 / H1 FUG qPCR samples were run in triplicate and results were average with error bars produced from the standard error of the mean.

Non-changing Transcripts

Probe Set ID	Day 0	Day 4	Day 6	Day 8	Gene Symbol	Gene Title
201492_s_at	33063	32605	33746	31572	RPL41	ribosomal protein L41
206559_x_at	32790	33751	31317	33456	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
201217_x_at	27369	25570	24547	25998	RPL3	ribosomal protein L3
200099_s_at	26945	26230	26206	24527	RPS3A	ribosomal protein S3A
200063_s_at	26511	24783	25901	24494	NPM1	Nucleophosmin
200926_at	26442	25860	26496	24505	RPS23	ribosomal protein S23
203107_x_at	26124	25489	24811	25771	RPS2	ribosomal protein S2
212039_x_at	25998	25111	25151	25435	RPL3	ribosomal protein L3
226131_s_at	25840	25671	26382	24712	RPS16	ribosomal protein S16
200717_x_at	24850	25288	26777	25020	RPL7	ribosomal protein L7
213801_x_at	24558	25043	24190	24097	RPSA	ribosomal protein SA
208695_s_at	24420	25693	24036	23922	RPL39	ribosomal protein L39
201049_s_at	23815	22610	21050	21254	RPS18	ribosomal protein S18
200032_s_at	23625	24129	24279	23010	RPL9	ribosomal protein L9
213890_x_at	23411	23068	22514	21147	RPS16	ribosomal protein S16
213377_x_at	23362	22279	20813	21157	RPS12	ribosomal protein S12
200016_x_at	23221	23945	24326	23361	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1
200095_x_at	23192	21657	22367	21606	RPS10	ribosomal protein S10
224594_x_at	23082	23275	23117	25143	ACTB	actin, beta
202649_x_at	22698	21932	20384	19763	RPS19	ribosomal protein S19
AFFX-HUMGAPDH	22524	20747	19427	19895	GAPDH	glyceraldehyde-3-phosphate dehydrogenase
213356_x_at	22379	23331	23305	22729	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1
212661_x_at	21951	21274	20488	21742	PPIA	peptidylprolyl isomerase A (cyclophilin A)
1553588_at	21232	21134	19412	19657	ND3	NADH dehydrogenase, subunit 3 (complex I)
213564_x_at	20315	21045	19734	20590	LDHB	lactate dehydrogenase B
208692_at	20146	20614	19499	20556	RPS3	ribosomal protein S3
200034_s_at	19207	20634	19606	18931	RPL6	ribosomal protein L6
224187_x_at	18047	18422	19208	17815	HSPA8	heat shock 70kDa protein 8
200080_s_at	16442	16690	16215	15981	H3F3A	H3 histone, family 3A
217719_at	15069	14880	15007	14150	EIF3L	eukaryotic translation initiation factor 3, subunit L
211939_x_at	14766	15144	14315	13479	BTF3	basic transcription factor 3
221476_s_at	14395	15386	14675	15536	RPL15	ribosomal protein L15
211787_s_at	13947	14296	13318	13925	EIF4A1	eukaryotic translation initiation factor 4A, isoform 1
221481_x_at	13797	14354	14591	14116	HNRNPD	heterogeneous nuclear ribonucleoprotein D
205292_s_at	13323	12656	11772	12089	HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
208517_x_at	13048	12841	12400	11518	BTF3	basic transcription factor 3
208687_x_at	12652	12435	12575	11860	HSPA8	heat shock 70kDa protein 8
208775_at	12218	12028	11855	11174	XPO1	exportin 1 (CRM1 homolog, yeast)
201385_at	11221	11161	11615	11188	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15
217092_x_at	9641	9533	9033	9104	LOC643308	ribosomal protein L7 pseudogene
211509_s_at	9509	9822	9698	10558	RTN4	reticulon 4
200040_at	9010	8754	8424	8529	KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1

201092_at	8884	8468	8525	8252	RBBP7	retinoblastoma binding protein 7
210949_s_at	8862	9058	8542	9229	EIF3C	eukaryotic translation initiation factor 3, subunit C
201121_s_at	8426	8783	8499	8104	PGRMC1	progesterone receptor membrane component 1
201993_x_at	7670	8297	8678	8078	HNRPDL	heterogeneous nuclear ribonucleoprotein D-like
215230_x_at	7660	8339	8464	8443	EIF3C	eukaryotic translation initiation factor 3, subunit C
208778_s_at	7559	7452	6779	6421	TCP1	t-complex 1
225285_at	7411	7550	7758	7150	BCAT1	branched chain aminotransferase 1, cytosolic
201327_s_at	7259	6976	6600	6135	CCT6A	chaperonin containing TCP1, subunit 6A (zeta 1)
213507_s_at	6958	6960	7235	7160	KPNB1	karyopherin (importin) beta 1
213491_x_at	6895	6925	7465	7789	RPN2	ribophorin II
200068_s_at	6512	6266	6265	6225	CANX	calnexin
200723_s_at	6173	6236	6384	6418	CAPRIN1	cell cycle associated protein 1
210024_s_at	6024	6236	5690	5764	UBE2E3	ubiquitin-conjugating enzyme E2E 3
220044_x_at	5809	5342	5799	5815	CROP	cisplatin resistance-associated overexpressed protein
209240_at	5259	5640	5527	5576	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase
200058_s_at	5048	5084	5403	5384	SNRNP200	small nuclear ribonucleoprotein
200708_at	4997	4654	4852	4590	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial
200851_s_at	4945	5149	5103	4825	KIAA0174	KIAA0174
200048_s_at	4594	4674	4374	4220	JTB	jumping translocation breakpoint
201503_at	4416	4208	4415	4053	G3BP1	GTPase activating protein (SH3 domain) binding protein 1
218047_at	4213	4204	4036	4060	OSBPL9	oxysterol binding protein-like 9
201296_s_at	4140	4271	4147	4320	WSB1	WD repeat and SOCS box-containing 1
200079_s_at	3857	3809	3512	3663	KARS	lysyl-tRNA synthetase
209108_at	3661	3975	3757	3498	TSPAN6	tetraspanin 6
217256_x_at	3529	3428	3401	3142	RPL36A	ribosomal protein L36a
222398_s_at	3385	3325	3276	3355	EFTUD2	elongation factor Tu GTP binding domain containing 2
222488_s_at	3298	3111	3083	2835	DCTN4	dynactin 4 (p62)
221767_x_at	3289	3105	3122	3010	HDLBP	high density lipoprotein binding protein
223106_at	3273	3520	3156	3083	TMEM14C	transmembrane protein 14C
209503_s_at	3263	3173	3084	2919	PSMC5	proteasome (prosome, macropain)ATPase, 5
218150_at	3256	3526	3539	3572	ARL5A	ADP-ribosylation factor-like 5A
200083_at	3250	3362	3277	3267	USP22	ubiquitin specific peptidase 22
221932_s_at	3105	2929	2869	2888	GLRX5	glutaredoxin 5
201304_at	2864	2963	2817	2684	NDUFA5	NADH dehydrogenase (ubiquinone) 1 alpha
202138_x_at	2733	2622	2495	2398	JTV1	JTV1 gene
208924_at	2723	2939	2992	2758	RNF11	ring finger protein 11
202382_s_at	2635	2479	2557	2728	GNPDA1	glucosamine-6-phosphate deaminase 1
202365_at	2538	2463	2558	2422	UNC119B	unc-119 homolog B (C. elegans)
208842_s_at	2500	2461	2441	2372	GORASP2	golgi reassembly stacking protein 2, 55kDa
225501_at	2467	2623	2414	2294	PHF6	PHD finger protein 6
211678_s_at	2408	2325	2308	2229	RNF114	ring finger protein 114
201696_at	2384	2282	2091	2170	SFRS4	splicing factor, arginine/serine-rich 4
224747_at	2380	2219	2148	1984	UBE2Q2	ubiquitin-conjugating enzyme E2Q family member 2
224607_s_at	2339	2365	2493	2382	SRP68	signal recognition particle 68kDa
216396_s_at	2324	2206	2116	1993	EI24	etoposide induced 2.4 mRNA
201036_s_at	2305	2392	2164	1991	HADH	hydroxyacyl-Coenzyme A dehydrogenase

208943_s_at	2136	2140	2282	2186	SEC62	SEC62 homolog (S. cerevisiae)
224619_at	2120	2237	2295	2208	CASC4	cancer susceptibility candidate 4
202560_s_at	1976	1951	1801	1803	C1orf77	chromosome 1 open reading frame 77
202420_s_at	1809	1887	2022	1986	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9
221479_s_at	1773	1804	1754	1636	BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like
200732_s_at	1713	1721	1822	1802	PTP4A1	protein tyrosine phosphatase type IVA, member 1
212779_at	1632	1687	1683	1693	KIAA1109	KIAA1109
208666_s_at	1626	1688	1530	1467	ST13	suppression of tumorigenicity 13
218505_at	1597	1605	1607	1720	WDR59	WD repeat domain 59
223209_s_at	1582	1608	1488	1596	SELS	selenoprotein S
227586_at	1499	1433	1441	1511	TMEM170A	transmembrane protein 170A
201928_at	1482	1591	1522	1468	PKP4	plakophilin 4

Table 38: List of stably expressed transcripts from 4 day microarray analysis

The top 100 stably expressed transcripts throughout the 4 day time course were sorted in order of expression intensity with the highest expressed transcripts at the top of the list. With the commonly used housekeeping genes *GAPDH* and *Beta Actin* also present within this list.

Selected Transcripts from Day 0 to Day 4

Affy ID	p-value	Fold change	Day0	Day4	Gene Symbol	Gene Title
204677_at	0.2033	326.2561	20.249	6924.7	CDH5	cadherin 5, type 2 (vascular endothelium)
213541_s_at	0.1974	226.2574	19.677	4452.9	ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)
236262_at	0.1974	208.75818	21.65	4565.4	MMRN2	multimerin 2
220330_s_at	0.1988	200.43456	17.515	3596.4	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1
202878_s_at	0.1974	199.17519	36.073	7368.5	CD93	CD93 molecule
223869_at	0.2229	197.18787	21.578	4661.3	SOST	sclerosteosis
205801_s_at	0.1974	160.01443	44.781	7175.5	RASGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)
208982_at	0.2051	129.13086	64.354	8184	PECAM1	platelet/endothelial cell adhesion molecule
228339_at	0.1974	122.49222	50.852	6255.1	ECSCR	endothelial cell-specific chemotaxis regulator
227874_at	0.1896	120.20736	18.46	2238.9	EMCN	endomucin
206283_s_at	0.2087	92.75034	40.189	3795	TAL1	T-cell acute lymphocytic leukemia 1
235306_at	0.2051	90.207466	21.814	2090.2	GIMAP8	GTPase, IMAP family member 8
222885_at	0.1974	89.921005	31.832	2768.3	EMCN	endomucin
231947_at	0.1974	88.75204	40.723	3635.4	MYCT1	myc target 1
204236_at	0.1974	87.4609	89.026	7748.1	FLI1	Friend leukemia virus integration 1
229910_at	0.2008	80.2986	45.614	3739.2	SHE	Src homology 2 domain containing E
228698_at	0.1785	78.847206	120.41	9527.9	SOX7	SRY (sex determining region Y)-box 7
234996_at	0.1974	76.10911	34.531	2640.9	CALCRL	calcitonin receptor-like
230478_at	0.2275	75.40476	24.093	2119.8	OIT3	oncoprotein induced transcript 3
230250_at	0.2162	75.13228	24.556	1899.7	PTPRB	protein tyrosine phosphatase, receptor type, B
212950_at	0.2334	73.28357	28.739	2284.9	GPR116	G protein-coupled receptor 116
204438_at	0.2183	65.58381	21.983	1436.6	MRC1	mannose receptor, C type 1
201721_s_at	0.1896	64.71529	58.463	3776	LAPTM5	lysosomal multispanning membrane protein 5
220637_at	0.2052	62.614372	57.107	3774.3	FAM124B	family with sequence similarity 124B
227780_s_at	0.1974	60.821896	94.021	5739.2	ECSCR	endothelial cell-specific chemotaxis regulator
206995_x_at	0.1974	57.310642	17.791	1011.4	SCARF1	scavenger receptor class F, member 1
213592_at	0.1769	55.181446	150.17	8287.3	APLNR	apelin receptor
213620_s_at	0.2008	52.77545	152.55	7859	ICAM2	intercellular adhesion molecule 2
40687_at	0.2216	51.832413	44.437	2643.3	GJA4	gap junction protein, alpha 4, 37kDa
204468_s_at	0.2051	48.83415	34.388	1671	TIE1	tyrosine kinase with immunoglobulin-like and EGF-like domains 1
225763_at	0.1974	48.41866	21.975	1072.9	RCSD1	RCSD domain containing 1
242931_at	0.1974	46.535164	19.062	877.18	LONRF3	LON peptidase N-terminal domain and ring finger 3
211597_s_at	0.2122	45.701244	102.16	5157.3	HOPX	HOP homeobox
205866_at	0.2162	45.458473	117.83	5851.6	FCN3	ficolin (collagen/fibrinogen domain containing) 3

226701_at	0.2329	44.6091	25.392	1286.7	GJA5	gap junction protein, alpha 5, 40kDa
203548_s_at	0.2162	44.13397	67.37	3477.7	LPL	lipoprotein lipase
214651_s_at	0.2739	43.275833	32.507	1921.3	HOXA9	homeobox A9
226955_at	0.2097	43.161556	81.829	3694.9	AFAP1L1	actin filament associated protein 1-like 1
219091_s_at	0.2075	42.54642	56.203	2438.2	MMRN2	multimerin 2
205269_at	0.2075	42.29471	12.725	579.88	LCP2	lymphocyte cytosolic protein 2
204249_s_at	0.1979	41.905083	119.51	5149.1	LMO2	LIM domain only 2 (rhombotin-like 1)
227289_at	0.1974	41.374565	108.15	4483	PCDH17	protocadherin 17
209199_s_at	0.2055	40.643066	120.75	4975.1	MEF2C	myocyte enhancer factor 2C
205612_at	0.2104	40.376953	17.552	746	MMRN1	multimerin 1
203887_s_at	0.2008	38.769604	67.447	2619	THBD	thrombomodulin
209795_at	0.2356	37.67864	35.958	1490.8	CD69	CD69 molecule
230061_at	0.2205	36.76924	60.225	2179.4	TM4SF18	Transmembrane 4 L six family member 18
203549_s_at	0.2162	35.497284	153.51	5933	LPL	lipoprotein lipase
208981_at	0.2223	35.170044	101.81	3553.1	PECAM1	platelet/endothelial cell adhesion molecule
228167_at	0.1996	35.136524	13.256	454.46	KLHL6	kelch-like 6 (Drosophila)
208983_s_at	0.2087	34.771168	39.781	1436.3	PECAM1	platelet/endothelial cell adhesion molecule
210665_at	0.1769	34.44757	253.29	8721.9	TFPI	tissue factor pathway inhibitor
235489_at	0.1974	33.754974	70.576	2371.1	RHOJ	ras homolog gene family, member J
227297_at	0.2223	33.35993	186.56	6804.1	ITGA9	integrin, alpha 9
209200_at	0.1974	32.624138	155.87	5107	MEF2C	myocyte enhancer factor 2C
222106_at	0.2547	31.983282	23.921	1025.5	PRND	prion protein 2 (dublet)
225369_at	0.2055	31.71041	58.081	1841.6	ESAM	endothelial cell adhesion molecule
228863_at	0.1974	31.4981	160.59	5005.4	PCDH17	protocadherin 17
221529_s_at	0.2356	31.427761	69.026	2473.4	PLVAP	plasmalemma vesicle associated protein
206145_at	0.2344	31.343185	18.679	704.97	RHAG	Rh-associated glycoprotein
215447_at	0.2011	30.37045	27.654	880.63	TFPI	Tissue factor pathway inhibitor
1552316_a_at	0.2077	30.271324	17.078	528.77	GIMAP1	GTPase, IMAP family member 1
1555638_a_at	0.2353	30.068472	54.275	1683.1	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1
230261_at	0.1974	29.36529	27.886	818.3	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4
219436_s_at	0.1974	27.718817	24.487	694.68	EMCN	endomucin
219777_at	0.2162	27.490158	11.528	331.4	GIMAP6	GTPase, IMAP family member 6
238488_at	0.2075	27.338062	45.305	1241.1	LRRC70	leucine rich repeat containing 70
209676_at	0.2181	26.991695	88.411	2373.1	TFPI	tissue factor pathway inhibitor
241926_s_at	0.1994	26.78391	12.575	340.45	ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)
230895_at	0.2033	26.740746	394.81	10679	HAPLN1	hyaluronan and proteoglycan link protein 1
226420_at	0.2287	26.452623	128.48	3907.5	EVI1	ecotropic viral integration site 1
205656_at	0.1769	26.278677	138.3	3632.8	PCDH17	protocadherin 17
210815_s_at	0.1974	25.879143	31.425	802.37	CALCRL	calcitonin receptor-like
204683_at	0.1668	25.873224	66.854	1729.5	ICAM2	intercellular adhesion molecule 2
210786_s_at	0.2181	25.812332	22.977	614.98	FLI1	Friend leukemia virus integration 1
231841_s_at	0.2184	25.702272	28.189	714.96	KIAA1462	KIAA1462
203934_at	0.1974	25.445131	530.25	13462	KDR	kinase insert domain receptor
224013_s_at	0.1974	25.328306	21.276	550.56	SOX7	SRY (sex determining region Y)-box 7
219778_at	0.1785	25.251823	56.047	1415.8	ZFPM2	zinc finger protein, multitype 2
230943_at	0.1974	24.70031	54.443	1343.9	SOX17	SRY (sex determining region Y)-box 17

227779_at	0.1668	24.611181	184.37	4539.8	ECSCR	Endothelial cell-specific chemotaxis regulator
206009_at	0.1974	23.88073	75.655	1790	ITGA9	integrin, alpha 9
228568_at	0.2097	23.548391	63.334	1509.8	GCOM1	GRINL1A complex locus
205984_at	0.1974	23.406197	229.94	5390	CRHBP	corticotropin releasing hormone binding protein
230147_at	0.1974	23.393465	448.51	10455	F2RL2	coagulation factor II (thrombin) receptor-like 2
221884_at	0.2376	23.355942	161.41	4589.3	EVI1	ecotropic viral integration site 1
204904_at	0.2565	22.976894	242.55	6380.1	GJA4	gap junction protein, alpha 4, 37kDa
235976_at	0.2326	22.96766	30.526	715.58	SLITRK6	SLIT and NTRK-like family, member 6
203477_at	0.2184	22.918978	26.633	673.26	COL15A1	collagen, type XV, alpha 1
206670_s_at	0.2087	22.808134	59.873	1443.7	GAD1	glutamate decarboxylase 1 (brain, 67kDa)
206795_at	0.1974	22.636108	18.366	417.7	F2RL2	coagulation factor II (thrombin) receptor-like 2
206669_at	0.2363	22.266542	84.624	1975.3	GAD1	glutamate decarboxylase 1 (brain, 67kDa)
221031_s_at	0.2125	22.17797	93.537	2246	APOLD1	apolipoprotein L domain containing 1
230836_at	0.2061	21.48411	30.039	634.56	ST8SIA4	ST8 alpha-N-acetyl-neuraminidealpha-2,8-sialyltransferase 4
205749_at	0.2053	21.211643	163.22	3501.2	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
226950_at	0.2087	21.099663	82.932	1801.5	ACVRL1	activin A receptor type II-like 1
219315_s_at	0.1974	20.790037	99.265	2066.1	TMEM204	transmembrane protein 204
204642_at	0.2008	20.686087	60.577	1268.7	S1PR1	sphingosine-1-phosphate receptor 1
228489_at	0.1974	20.448097	120.3	2480	TM4SF18	transmembrane 4 L six family member 18
205934_at	0.1974	20.386135	16.662	345.37	PLCL1	phospholipase C-like 1
213316_at	0.2055	20.311823	59.457	1198	KIAA1462	KIAA1462
218353_at	0.2219	19.866278	181.04	3698.3	RGS5	regulator of G-protein signaling 5
218589_at	0.1785	19.793259	230.07	4555.5	P2RY5	purinergic receptor P2Y, G-protein coupled, 5
226992_at	0.2162	19.602444	34.998	642.86	NOSTRIN	nitric oxide synthase trafficker
211434_s_at	0.2492	19.57276	28.258	565.48	CCRL2	chemokine (C-C motif) receptor-like 2
205159_at	0.1987	19.533447	28.269	554.72	CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity
219243_at	0.1769	19.358768	41.479	803.78	GIMAP4	GTPase, IMAP family member 4
241986_at	0.1974	19.206175	25.04	485.77	BMPER	BMP binding endothelial regulator
203888_at	0.1974	19.00431	70.925	1343.9	THBD	thrombomodulin
226673_at	0.1974	18.903494	67.111	1267.3	SH2D3C	SH2 domain containing 3C
238906_s_at	0.1769	18.549995	27.723	514.18	RHOJ	ras homolog gene family, member J
242943_at	0.1974	18.475048	22.827	422.43	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase4
204220_at	0.1974	18.395945	104	1933.3	GMFG	glia maturation factor, gamma
230204_at	0.1974	18.35572	721.13	13235	HAPLN1	hyaluronan and proteoglycan link protein 1
201288_at	0.1785	18.292955	189.45	3468.9	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta
205523_at	0.1896	18.229298	828.05	15073	HAPLN1	hyaluronan and proteoglycan link protein 1
214591_at	0.2229	18.222376	32.507	659.88	KLHL4	kelch-like 4 (Drosophila)
206331_at	0.2097	18.09374	97.096	1772.9	CALCRL	calcitonin receptor-like
205524_s_at	0.1974	17.814808	807.47	14435	HAPLN1	hyaluronan and proteoglycan link protein 1
210664_s_at	0.1979	17.788155	809.75	14509	TFPI	tissue factor pathway inhibitor

228372_at	0.1974	17.71886	30.513	542.61	C10orf128	chromosome 10 open reading frame 128
214265_at	0.1987	17.649551	66.541	1168.2	ITGA8	integrin, alpha 8
206211_at	0.2551	17.545464	16.925	351.92	SELE	selectin E
219134_at	0.1974	17.451262	372.38	6463.7	ELTD1	EGF, latrophilin and seven transmembrane domain containing 1
219761_at	0.1974	17.402481	46.551	825.57	CLEC1A	C-type lectin domain family 1, member A
209543_s_at	0.167	17.328342	150.02	2600.2	CD34	CD34 molecule
231842_at	0.2008	17.280003	72.142	1237.1	KIAA1462	KIAA1462
236295_s_at	0.2087	17.259386	41.877	725.3	NLRC3	NLR family, CARD domain containing 3
205366_s_at	0.227	17.25801	44.398	827.26	HOXB6	homeobox B6
214438_at	0.2075	17.166708	95.699	1730.8	HLX	H2.0-like homeobox
206702_at	0.1974	16.782217	508.98	8537.8	TEK	TEK tyrosine kinase, endothelial
205278_at	0.2087	16.749695	210.72	3637.7	GAD1	glutamate decarboxylase 1 (brain, 67kDa)
205846_at	0.234	16.60154	37.079	596.44	PTPRB	protein tyrosine phosphatase, receptor type, B
222717_at	0.1974	16.556358	15.406	254.28	SDPR	serum deprivation response
205898_at	0.234	16.525808	19.746	346.69	CX3CR1	chemokine (C-X3-C motif) receptor 1
204115_at	0.2008	16.513813	552.87	9260.2	GNG11	guanine nucleotide binding protein (G protein), gamma 11
207691_x_at	0.2008	16.50619	36.11	611.55	ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1
230367_at	0.2008	16.456188	92.974	1557.5	SMTNL1	smoothelin-like 1
204482_at	0.2341	16.282593	120.79	2222.6	CLDN5	claudin 5
203508_at	0.2087	16.113134	53.697	896.57	TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B
207016_s_at	0.2342	16.017344	47.618	871.78	ALDH1A2	aldehyde dehydrogenase 1 family, member A2
204723_at	0.1974	15.900242	42.912	677.45	SCN3B	sodium channel, voltage-gated, type III, beta
212750_at	0.2008	15.482925	103.06	1627.8	PPP1R16B	protein phosphatase 1, regulatory (inhibitor) subunit 16B
1553635_s_at	0.2326	15.456221	32.923	537.35	TCTEX1D1	Tctex1 domain containing 1
221840_at	0.2162	15.306811	79.065	1239.1	PTPRE	protein tyrosine phosphatase, receptor type, E
226122_at	0.2087	14.855433	79.962	1268.2	PLEKHG1	pleckstrin homology domain containing, family G member 1
212325_at	0.1974	14.786374	140.2	2076.9	LIMCH1	LIM and calponin homology domains 1
218660_at	0.1029	14.785619	364.37	5387.4	DYSF	dysferlin, limb girdle muscular dystrophy 2B
209905_at	0.3021	14.771102	69.785	1419.2	HOXA9	homeobox A9
41577_at	0.2008	14.636342	118.57	1762.6	PPP1R16B	protein phosphatase 1, regulatory (inhibitor) subunit 16B
212951_at	0.2177	14.479265	51.639	816.24	GPR116	G protein-coupled receptor 116
232068_s_at	0.2326	14.4420595	24.341	359.18	TLR4	toll-like receptor 4
237252_at	0.1769	14.1702385	28.929	410.07	THBD	thrombomodulin
210139_s_at	0.2391	14.168494	560.22	8749.9	PMP22	peripheral myelin protein 22
223754_at	0.2008	14.159286	58.434	844.49	C2orf88	chromosome 2 open reading frame 88
224341_x_at	0.2055	14.13086	42.224	601.73	TLR4	toll-like receptor 4
204722_at	0.1896	14.113259	40.251	566.49	SCN3B	sodium channel, voltage-gated, type III, beta
1555233_at	0.1974	13.987184	52.012	734.63	RHOJ	ras homolog gene family, member J
208394_x_at	0.4183	13.940657	33.647	1423.8	ESM1	endothelial cell-specific molecule 1

219993_at	0.2075	13.920034	54.633	773.16	SOX17	SRY (sex determining region Y)-box 17
213258_at	0.1769	13.785161	842.56	11613	TFPI	tissue factor pathway inhibitor
209070_s_at	0.2087	13.768181	317.99	4489.8	RGS5	regulator of G-protein signaling 5
204694_at	0.2657	13.663281	280.37	4946.7	AFP	alpha-fetoprotein
235182_at	0.2055	13.6526	56.56	789.3	ISM1	isthmin 1 homolog (zebrafish)
203708_at	0.1785	13.598154	317.5	4315.9	PDE4B	phosphodiesterase 4B, cAMP-specific
240532_at	0.234	13.585779	24.808	370.16	SLC32A1	solute carrier family 32 member 1
201720_s_at	0.1896	13.514512	97.045	1315.9	LAPTM5	lysosomal multispanning membrane protein 5
220027_s_at	0.2052	13.37601	181.88	2375.4	RASIP1	Ras interacting protein 1
205466_s_at	0.2175	13.262622	38.802	547.76	HS3ST1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1
213115_at	0.1974	13.174369	147.61	1954.4	ATG4A	ATG4 autophagy related 4 homolog A (S. cerevisiae)
208966_x_at	0.1974	13.149601	603.85	7994.2	IFI16	interferon, gamma-inducible protein 16
214329_x_at	0.2157	13.14249	17.413	238.82	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10
207002_s_at	0.1996	13.066617	299.33	3977.3	PLAGL1	pleiomorphic adenoma gene-like 1
212328_at	0.2055	13.037352	420.5	5524.4	LIMCH1	LIM and calponin homology domains 1
204567_s_at	0.2162	13.033267	16.17	222.74	ABCG1	ATP-binding cassette, sub-family G (WHITE), member 1
226814_at	0.2338	12.9957	179.05	2596.9	ADAMTS9	ADAM metallopeptidase thrombospondin type 1 motif 9
218806_s_at	0.1785	12.9875	89.485	1161.2	VAV3	vav 3 guanine nucleotide exchange factor
228311_at	0.2162	12.974192	199.98	2691.9	BCL6B	B-cell CLL/lymphoma 6, member B (zinc finger protein)
216973_s_at	0.2343	12.9535	42.14	560.18	HOXB7	homeobox B7
220300_at	0.1769	12.91542	37.942	489.61	RGS3	regulator of G-protein signaling 3
204731_at	0.2162	12.753868	240.38	3174.3	TGFBR3	transforming growth factor, beta receptor III
213150_at	0.2521	12.722925	27.658	426.88	HOXA10	homeobox A10
204779_s_at	0.2343	12.613849	67.328	884.98	HOXB7	homeobox B7
205978_at	0.227	12.603537	15.867	212.35	KL	klotho
204689_at	0.2175	12.464438	191.22	2495.1	HHEX	hematopoietically expressed homeobox

Table 39: List of Transcripts from Day0toDay4Up2fold+HaemSim+p0.5 Gene List

The top 200 transcripts within the gene list Day0toDay4Up2fold+HaemSim+p0.5 are displayed and sorted in order of fold change (highest at the top).

Selected Transcripts from Day 4 to Day 6

Affy ID	p-value	Fold change	Day 4	Day 6	Gene Symbol	Gene Title
223916_s_at	0.028	26.466	11.552	319.837	BCOR	BCL6 co-repressor
242352_at	0.089	12.694	152.095	1813.21	NIPBL	Nipped-B homolog (Drosophila)
243995_at	0.070	12.421	55.383	766.014	PTAR1	protein prenyltransferase alpha subunit repeat containing 1
221765_at	0.153	8.694	65.135	389.380	UGCG	UDP-glucose ceramide glucosyltransferase
206762_at	0.235	8.664	87.547	685.360	KCNA5	potassium voltage-gated channel, shaker-related subfamily, member 5
236207_at	0.072	8.413	8.951	76.549	SSFA2	sperm specific antigen 2
233827_s_at	0.182	7.762	256.253	1524.03	SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae)
208701_at	0.062	7.651	7.188	51.033	APLP2	Amyloid beta (A4) precursor-like protein 2
219387_at	0.395	7.404	215.682	560.849	CCDC88A	coiled-coil domain containing 88A
206765_at	0.093	7.277	162.108	1025.12	KCNJ2	potassium inwardly-rectifying channel, subfamily J, member 2
204270_at	0.146	6.982	34.714	273.241	SKI	v-ski sarcoma viral oncogene homolog (avian)
226663_at	0.412	6.906	769.104	1679.25	ANKRD10	ankyrin repeat domain 10
227647_at	0.114	6.795	50.632	419.767	KCNE3	potassium voltage-gated channel, Isk-related family, member 3
1554260_at	0.467	6.474	92.620	279.125	FRYL	FRY-like
215092_s_at	0.468	6.421	164.634	392.229	NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive
222562_s_at	0.294	6.379	135.754	397.160	TNKS2	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2
236930_at	0.340	6.140	77.567	213.094	NUMB	Numb homolog (Drosophila)
1552482_at	0.082	6.031	57.209	284.523	RAPH1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1
211993_at	0.207	5.953	194.094	850.445	WNK1	WNK lysine deficient protein kinase 1
219717_at	0.392	5.738	50.003	122.601	C4orf30	chromosome 4 open reading frame 30
216109_at	0.483	5.738	68.075	176.192	MED13L	Mediator complex subunit 13-like
235852_at	0.057	5.639	23.651	123.629	STON2	Stonin 2
1563947_a_at	0.472	5.600	200.145	364.621	ERC1	ELKS/RAB6-interacting/CAST family member 1
229010_at	0.157	5.577	62.897	388.151	CBL	Cas-Br-M (murine) ecotropic retroviral transforming sequence
243683_at	0.411	5.537	206.655	438.965	MORF4L2	Mortality factor 4 like 2
222880_at	0.008	5.371	28.503	154.516	AKT3	v-akt murine thymoma viral oncogene homolog 3
230106_at	0.010	5.371	31.643	167.382	ZXDC	ZXD family zinc finger C
216493_s_at	0.220	5.349	29.695	151.181	LOC645468	insulin-like growth factor 2 mRNA binding protein 3
218457_s_at	0.056	5.327	62.287	314.029	DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha
1569607_s_at	0.301	5.292	42.078	144.902	LOC728783	ankyrin repeat domain 20 family
213286_at	0.273	5.232	158.232	497.442	ZFR	zinc finger RNA binding protein
1553122_s_at	0.100	5.191	9.223	53.301	RBAK	RB-associated KRAB zinc finger
242918_at	0.101	5.099	220.190	1086.74	NASP	Nuclear autoantigenic sperm protein (histone-binding)

212777_at	0.281	4.996	15.989	148.896	SOS1	son of sevenless homolog 1 (Drosophila)
211081_s_at	0.119	4.936	75.898	418.602	MAP4K5	mitogen-activated protein kinase kinase kinase kinase 5
243857_at	0.367	4.850	221.661	535.538	MORF4L2	Mortality factor 4 like 2
203056_s_at	0.039	4.793	17.575	81.507	PRDM2	PR domain containing 2, with ZNF domain
224828_at	0.034	4.771	10.603	51.616	CPEB4	cytoplasmic polyadenylation element binding protein 4
229966_at	0.393	4.744	498.796	955.314	EWSR1	Ewing sarcoma breakpoint region 1
1557227_s_at	0.170	4.737	57.895	222.222	TPR	translocated promoter region (to activated MET oncogene)
202773_s_at	0.106	4.729	101.907	439.256	SFRS8	splicing factor, arginine/serine-rich 8 (suppressor-of-white-apricot homolog, Drosophila)
1555125_at	0.294	4.602	4.974	30.468	C21orf66	chromosome 21 open reading frame 66
1557388_at	0.488	4.580	35.601	101.154	RTTN	rotatin
235216_at	0.013	4.572	78.779	361.015	ESCO1	establishment of cohesion 1 homolog 1 (S. cerevisiae)
219138_at	0.105	4.562	253.114	980.566	RPL14	ribosomal protein L14
235645_at	0.280	4.525	107.857	492.320	ESCO1	establishment of cohesion 1 homolog 1 (S. cerevisiae)
203096_s_at	0.240	4.433	582.023	1558.97	RAPGEF2	Rap guanine nucleotide exchange factor (GEF) 2
231235_at	0.230	4.368	132.544	347.650	NKTR	natural killer-tumor recognition sequence
212076_at	0.323	4.365	180.464	372.262	MLL	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
242701_at	0.138	4.338	5.803	26.279	TBRG1	transforming growth factor beta regulator 1
219232_s_at	0.416	4.329	76.109	223.912	EGLN3	egl nine homolog 3 (C. elegans)
224517_at	0.030	4.273	37.875	155.089	POLR2J4	polymerase (RNA) II (DNA directed) polypeptide J4, pseudogene
216563_at	0.268	4.258	75.026	273.005	ANKRD12	Ankyrin repeat domain 12
209203_s_at	0.425	4.244	56.796	98.387	BICD2	bicaudal D homolog 2 (Drosophila)
244766_at	0.049	4.233	16.844	67.619	LOC440354	PI-3-kinase-related kinase SMG-1
210407_at	0.488	4.188	52.884	120.497	PPM1A	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
229467_at	0.278	4.179	243.413	778.507	PCBP2	Poly(rC) binding protein 2
224580_at	0.164	4.170	12.066	44.262	SLC38A1	solute carrier family 38, member 1
232238_at	0.397	4.148	325.410	1735.68	ASPM	asp (abnormal spindle) homolog, microcephaly associated
241425_at	0.455	4.129	206.826	334.258	NUPL1	Nucleoporin like 1
33148_at	0.436	4.125	224.788	452.580	ZFR	zinc finger RNA binding protein
230057_at	0.293	4.119	22.516	92.911	LOC285178	hypothetical protein LOC285178
211074_at	0.095	4.093	129.212	591.847	FOLR1	folate receptor 1 (adult)
235019_at	0.321	4.080	42.165	135.736	CPM	carboxypeptidase M
204750_s_at	0.462	4.074	41.660	127.285	DSC2	desmocollin 2
206552_s_at	0.462	4.073	27.164	341.473	TAC1	tachykinin, precursor 1
220946_s_at	0.459	4.067	113.962	200.779	SETD2	SET domain containing 2
223701_s_at	0.245	4.051	95.884	285.992	USP47	ubiquitin specific peptidase 47
215169_at	0.398	4.043	31.369	195.776	SLC35E2	solute carrier family 35, member E2
204840_s_at	0.492	4.025	109.008	205.001	EEA1	early endosome antigen 1
242408_at	0.242	4.022	36.887	173.397	STYX	serine/threonine/tyrosine interacting protein

1556066_at	0.318	4.002	6.190	40.829	KDM6B	lysine (K)-specific demethylase 6B
240728_at	0.029	3.992	11.611	48.121	PLCB4	Phospholipase C, beta 4
239757_at	0.387	3.971	238.408	437.864	ZFAND6	Zinc finger, AN1-type domain 6
209013_x_at	0.328	3.967	14.288	92.128	TRIO	triple functional domain (PTPRF interacting)
223577_x_at	0.012	3.964	554.372	2196.48	MALAT1	metastasis associated lung adenocarcinoma transcript 1
1560105_at	0.001	3.964	10.182	40.317	PTPRB	Protein tyrosine phosphatase, receptor type, B
209875_s_at	0.490	3.937	10.503	89.862	SPP1	secreted phosphoprotein 1
212720_at	0.098	3.931	247.614	987.509	PAPOLA	poly(A) polymerase alpha
202765_s_at	0.075	3.929	21.290	75.896	FBN1	fibrillin 1
227223_at	0.201	3.915	594.496	1609.97	RBM39	RNA binding motif protein 39
1570259_at	0.373	3.914	139.766	297.096	LIMS1	LIM and senescent cell antigen-like domains 1
203000_at	0.373	3.904	44.221	114.212	STMN2	stathmin-like 2
1557729_at	0.093	3.885	8.408	31.699	LOC10013184	FP2025
233882_s_at	0.115	3.875	38.219	139.203	SEMA6D	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D
201101_s_at	0.376	3.846	656.011	1608.25	BCLAF1	BCL2-associated transcription factor 1
212524_x_at	0.386	3.829	4.385	24.694	H2AFX	H2A histone family, member X
213075_at	0.362	3.824	14.094	62.185	OLFML2A	olfactomedin-like 2A
213446_s_at	0.191	3.814	455.053	1746.19	IQGAP1	IQ motif containing GTPase activating protein 1
1566129_at	0.405	3.813	61.765	119.722	LIMS1	LIM and senescent cell antigen-like domains 1
233874_at	0.379	3.792	14.777	120.310	SLAIN2	SLAIN motif family, member 2
228986_at	0.169	3.785	278.251	1190.87	OSBPL8	oxysterol binding protein-like 8
214415_at	0.288	3.768	95.125	244.423	PLGLB1	plasminogen-like B1
213350_at	0.163	3.707	2900.18	9079.04	RPS11	Ribosomal protein S11
1557278_s_at	0.093	3.697	25.225	85.880	TNPO1	Transportin 1
236264_at	0.410	3.682	12.177	79.809	LPHN3	latrophilin 3
218273_s_at	0.207	3.673	139.924	409.281	PPM2C	protein phosphatase 2C, magnesium-dependent, catalytic subunit
234701_at	0.025	3.670	36.363	137.545	ANKRD11	ankyrin repeat domain 11
230630_at	0.189	3.670	97.191	447.422	AK3L1	Adenylate kinase 3-like 1
225420_at	0.411	3.669	45.251	90.827	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial
1553768_a_at	0.237	3.641	2350.32	7654.44	DCBLD1	discoidin, CUB and LCCL domain containing 1
219925_at	0.105	3.639	87.788	310.078	ZMYM6	zinc finger, MYM-type 6
223122_s_at	0.316	3.635	38.576	256.915	SFRP2	secreted frizzled-related protein 2
239742_at	0.061	3.634	649.720	2352.20	TULP4	Tubby like protein 4
236613_at	0.212	3.631	162.568	456.788	RBM25	RNA binding motif protein 25
215220_s_at	0.361	3.625	195.040	339.375	TPR	translocated promoter region (to activated MET oncogene)
244546_at	0.453	3.612	44.275	120.901	CYCS	cytochrome c, somatic
210172_at	0.077	3.609	70.825	228.498	SF1	splicing factor 1
1556049_at	0.215	3.580	117.439	319.979	RTN4	reticulon 4
227101_at	0.209	3.570	115.125	351.436	ZNF800	zinc finger protein 800
206848_at	0.058	3.524	58.353	201.318	HOXA7	homeobox A7
1556277_at	0.238	3.514	252.711	621.300	PAPD4	PAP associated domain containing 4
209327_s_at	0.173	3.494	15.718	70.164	NOP16	NOP16 nucleolar protein homolog (yeast)
226310_at	0.120	3.490	176.897	571.029	RICTOR	rapamycin-insensitive companion of mTOR
235461_at	0.381	3.483	16.597	66.618	TET2	tet oncogene family member 2

201299_s_at	0.006	3.477	104.858	364.414	MOBKL1B	MOB1, Mps One Binder kinase activator-like 1B (yeast)
209286_at	0.076	3.461	98.448	354.898	CDC42EP3	CDC42 effector protein (Rho GTPase binding) 3
212178_s_at	0.008	3.456	202.943	695.319	POM121	POM121 membrane glycoprotein (rat)
235412_at	0.187	3.447	831.796	2050.80	ARHGEF7	Rho guanine nucleotide exchange factor (GEF) 7
231735_s_at	0.073	3.423	1652.56	5722.19	MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)
210479_s_at	0.124	3.404	36.149	146.234	RORA	RAR-related orphan receptor A
205511_at	0.318	3.399	150.263	281.206	FLJ10038	hypothetical protein FLJ10038
212650_at	0.218	3.390	103.188	304.579	EHBP1	EH domain binding protein 1
208003_s_at	0.448	3.376	106.397	169.893	NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive
1559954_s_at	0.313	3.368	64.596	133.909	DDX42	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
1561361_at	0.333	3.354	9.949	37.088	ZNF660	zinc finger protein 660
210426_x_at	0.074	3.352	58.946	216.023	RORA	RAR-related orphan receptor A
1553252_at	0.151	3.343	361.911	1114.14	BRWD3	bromodomain and WD repeat domain containing 3
212107_s_at	0.100	3.340	676.119	2136.93	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9
205756_s_at	0.085	3.318	15.172	49.278	F8	coagulation factor VIII, procoagulant component
1557963_at	0.316	3.317	305.686	652.404	CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)
202687_s_at	0.448	3.308	179.784	475.537	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10
210742_at	0.390	3.296	91.584	213.772	CDC14A	CDC14 cell division cycle 14 homolog A (S. cerevisiae)
1557049_at	0.455	3.285	7.323	28.183	LOC149478	Hypothetical protein LOC149478
203424_s_at	0.449	3.285	165.566	1068.22	IGFBP5	insulin-like growth factor binding protein 5
203623_at	0.448	3.278	9.548	46.644	PLXNA3	plexin A3
212105_s_at	0.497	3.254	53.863	100.715	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9
1553448_at	0.307	3.242	108.355	492.633	FLJ34503	hypothetical FLJ34503
201224_s_at	0.226	3.237	206.253	880.355	SRRM1	serine/arginine repetitive matrix 1
202756_s_at	0.107	3.236	20.220	67.886	GPC1	glypican 1
212384_at	0.288	3.232	411.273	972.802	BAT1	HLA-B associated transcript 1
205257_s_at	0.350	3.193	8.740	40.511	AMPH	amphiphysin
213926_s_at	0.412	3.190	102.250	300.286	AGFG1	ArfGAP with FG repeats 1
1565867_at	0.310	3.181	85.024	186.702	LOC10013140	hypothetical protein LOC100131402
236023_at	0.149	3.178	44.287	116.503	CDK9	cyclin-dependent kinase 9
1557685_at	0.379	3.175	46.495	118.962	C4orf38	Chromosome 4 open reading frame 38
238419_at	0.223	3.174	41.162	104.697	PHLDB2	pleckstrin homology-like domain, family B, member 2
213998_s_at	0.194	3.162	720.029	1682.69	DDX17	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17
239035_at	0.252	3.154	56.274	141.878	MTHFR	5,10-methylenetetrahydrofolate reductase (NADPH)
1557350_at	0.383	3.129	18.877	40.258	G3BP1	GTPase activating protein (SH3 domain) binding protein 1
230742_at	0.361	3.128	108.977	226.517	RBM6	RNA binding motif protein 6
219049_at	0.076	3.117	582.735	1667.59	CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1
242248_at	0.113	3.114	34.096	122.940	PHKB	phosphorylase kinase, beta

237741_at	0.168	3.103	231.305	579.543	SLC25A36	Solute carrier family 25, member 36
205769_at	0.237	3.096	65.757	219.138	SLC27A2	solute carrier family 27 (fatty acid transporter), member 2
60815_at	0.086	3.094	40.356	112.017	POLR2J4	polymerase (RNA) II (DNA directed) polypeptide J4
241414_at	0.247	3.091	14.100	51.791	ANKRD10	ankyrin repeat domain 10
1557384_at	0.437	3.090	463.206	1029.02	ZNF131	Zinc finger protein 131
229033_s_at	0.295	3.088	139.809	342.362	MUM1	melanoma associated antigen (mutated) 1
220459_at	0.404	3.062	98.445	225.301	MCM3APAS	MCM3AP antisense RNA (non-protein coding)
225173_at	0.262	3.054	304.899	692.555	ARHGAP18	Rho GTPase activating protein 18
207711_at	0.183	3.050	21.045	52.896	C20orf117	chromosome 20 open reading frame 117
234268_at	0.201	3.037	8.610	27.825	SLC2A13	solute carrier family 2 (facilitated glucose transporter), member 13
219234_x_at	0.020	3.032	13.655	41.186	SCRN3	secernin 3
1562442_at	0.272	3.030	39.044	87.087	SSBP1	single-stranded DNA binding protein 1
1559470_at	0.288	3.026	7.411	24.425	D21S2088E	D21S2088E
235476_at	0.072	3.025	297.077	885.100	TRIM59	tripartite motif-containing 59
204597_x_at	0.049	3.019	185.817	573.874	STC1	stanniocalcin 1
243361_at	0.336	3.011	70.681	127.286	SFRS12	splicing factor, arginine/serine-rich 12
227062_at	0.428	3.009	668.026	859.286	NCRNA00084	non-protein coding RNA 84
240859_at	0.283	2.990	247.610	484.633	ZFYVE16	zinc finger, FYVE domain containing 16
241701_at	0.060	2.980	22.766	66.561	ARHGAP21	Rho GTPase activating protein 21
236862_at	0.119	2.975	248.600	659.040	GOPC	Golgi associated PDZ and coiled-coil motif containing
233304_at	0.493	2.966	6.798	35.871	NFIB	Nuclear factor I/B
205691_at	0.228	2.956	19.708	45.552	SYNGR3	synaptogyrin 3
214786_at	0.392	2.949	9.735	38.965	MAP3K1	mitogen-activated protein kinase kinase kinase 1
217878_s_at	0.251	2.939	649.317	2079.46	CDC27	cell division cycle 27 homolog (S. cerevisiae)
204913_s_at	0.074	2.933	568.702	1835.67	SOX11	SRY (sex determining region Y)-box 11
208859_s_at	0.265	2.919	689.635	1560.05	ATRX	alpha thalassemia/mental retardation syndrome X-linked
239014_at	0.271	2.914	224.584	459.038	CCAR1	Cell division cycle and apoptosis regulator 1
214250_at	0.053	2.912	68.048	213.150	NUMA1	nuclear mitotic apparatus protein 1
206967_at	0.324	2.909	42.470	146.267	CCNT1	cyclin T1
202549_at	0.372	2.897	129.544	231.786	VAPB	VAMP (vesicle-associated membrane protein)-associated protein B and C
235576_at	0.234	2.891	42.839	110.272	WDR27	WD repeat domain 27
1556203_at	0.384	2.876	459.573	860.333	SRGAP2	SLIT-ROBO Rho GTPase activating protein 2
227931_at	0.235	2.869	400.338	944.791	INO80D	INO80 complex subunit D
242829_x_at	0.162	2.856	286.394	691.808	FBXL3	F-box and leucine-rich repeat protein 3
227129_x_at	0.065	2.852	441.956	1163.81	tcag7.907	hypothetical LOC402483
239695_at	0.367	2.850	77.415	151.686	JAK1	Janus kinase 1
1559101_at	0.340	2.849	129.738	244.147	FYN	FYN oncogene related to SRC, FGR, YES
1556202_at	0.281	2.846	122.557	259.070	SRGAP2	SLIT-ROBO Rho GTPase activating protein 2
229982_at	0.188	2.845	654.301	1720.47	QSER1	glutamine and serine rich 1
230566_at	0.028	2.839	46.943	132.966	C22orf27	chromosome 22 open reading frame 27
212808_at	0.068	2.838	70.418	204.418	NFATC2IP	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein

239665_at	0.491	2.829	13.585	50.979	LOC441179	hypothetical LOC441179
235789_at	0.351	2.825	4.625	17.656	KDM4B	lysine (K)-specific demethylase 4B
219568_x_at	0.391	2.819	6.241	25.444	SOX18	SRY (sex determining region Y)-box 18
240890_at	0.392	2.814	68.789	375.862	LOC643733	hypothetical LOC643733
218430_s_at	0.373	2.812	128.056	267.176	RFX7	regulatory factor X, 7

Table 40: List of Transcripts from Day4toDay6Up2fold+HaemSim+p0.5 Gene List

The top 200 transcripts within the gene list Day4toDay6Up2fold+HaemSim+p0.5 are displayed and sorted in order of fold change (highest at the top).

Selected Transcripts from Day 6 to Day 8

Affy ID	p-value	Fold change	Day 6	Day 8	Gene Symbol	Gene Title
213515_x_at	0.49	54.44	4539.9	13709.7	HBG1 /// HBG2	hemoglobin, gamma A /// hemoglobin, gamma G
214146_s_at	0.45	43.54	4.33	2400.13	PPBP	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)
204419_x_at	0.48	32.22	644.14	2152.32	HBG1 /// HBG2	hemoglobin, gamma A /// hemoglobin, gamma G
206390_x_at	0.45	28.77	141.63	3347.65	PF4	platelet factor 4
204551_s_at	0.13	27.89	10.82	263.01	AHSG	alpha-2-HS-glycoprotein
210929_s_at	0.35	26.47	395.87	2669.81	AHSG	alpha-2-HS-glycoprotein
205919_at	0.46	24.92	4546.3	16319.6	HBE1	hemoglobin, epsilon 1
214414_x_at	0.48	24.65	677.72	2981.06	HBA1 /// HBA2	hemoglobin, alpha 1 /// hemoglobin, alpha 2
219612_s_at	0.33	24.05	62.80	605.83	FGG	fibrinogen gamma chain
204694_at	0.44	23.16	3881.9	17266.8	AFP	alpha-fetoprotein
204848_x_at	0.48	18.71	326.67	962.73	HBG1 /// HBG2	hemoglobin, gamma A /// hemoglobin, gamma G
220114_s_at	0.11	16.34	4.61	83.81	STAB2	stabilin 2
203471_s_at	0.38	16.12	4.79	255.21	PLEK	pleckstrin
211298_s_at	0.20	15.14	6.30	138.08	ALB	albumin
206647_at	0.45	14.92	82.54	545.32	HBZ	hemoglobin, zeta
1553808_at	0.07	13.41	5.38	75.88	NKX2-3	NK2 transcription factor related, locus 3 (Drosophila)
216361_s_at	0.26	13.37	13.81	191.20	MYST3	MYST histone acetyltransferase (monocytic leukemia) 3
205650_s_at	0.42	13.35	182.21	611.40	FGA	fibrinogen alpha chain
210809_s_at	0.27	12.75	46.00	309.62	POSTN	periostin, osteoblast specific factor
219140_s_at	0.42	12.52	99.29	454.55	RBP4	retinol binding protein 4, plasma
205649_s_at	0.19	12.31	20.11	185.06	FGA	fibrinogen alpha chain
202833_s_at	0.13	12.05	7.55	96.45	SERPINA1	serpin peptidase inhibitor, clade A
211429_s_at	0.15	11.75	42.97	473.45	SERPINA1	serpin peptidase inhibitor, clade A
206834_at	0.37	11.70	61.99	281.24	HBD	hemoglobin, delta
228708_at	0.32	11.19	76.55	1106.89	RAB27B	RAB27B, member RAS oncogene family
219059_s_at	0.05	10.64	6.05	64.09	LYVE1	lymphatic vessel endothelial hyaluronan receptor 1
206493_at	0.43	9.93	147.25	695.67	ITGA2B	integrin, alpha 2b
219466_s_at	0.47	9.87	1482.7	3435.99	APOA2	apolipoprotein A-II
211560_s_at	0.32	9.42	22.04	118.75	ALAS2	aminolevulinate, delta-, synthase 2
207269_at	0.50	9.06	3.71	154.38	DEFA4	defensin, alpha 4, corticostatin
230908_at	0.09	8.82	9.30	92.22	TACR1	tachykinin receptor 1
206373_at	0.50	8.70	4.29	164.43	ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)
205837_s_at	0.47	8.68	224.02	538.75	GYPA	glycophorin A (MNS blood group)
1552848_at	0.24	8.43	59.80	446.05	PTCHD1	patched domain containing 1
213150_at	0.06	8.43	22.01	162.73	HOXA10	homeobox A10
207815_at	0.49	7.94	7.22	224.81	PF4V1	platelet factor 4 variant 1
205108_s_at	0.37	7.90	164.79	635.01	APOB	apolipoprotein B (including Ag(x) antigen)
219602_s_at	0.19	7.77	173.80	1007.06	FAM38B	family with sequence similarity 38, member B
232090_at	0.16	7.48	28.72	202.59	LOC100128178	similar to hCG2041313

201110_s_at	0.42	7.33	26.41	624.10	THBS1	thrombospondin 1
230560_at	0.02	7.03	12.31	88.13	STXBP6	syntaxin binding protein 6 (amisyn)
210839_s_at	0.40	6.94	36.82	190.30	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2
213317_at	0.50	6.86	6.69	161.16	CLIC5	chloride intracellular channel 5
222881_at	0.19	6.76	31.58	177.62	HPSE	heparanase
203949_at	0.50	6.73	5.86	136.04	MPO	myeloperoxidase
226068_at	0.15	6.50	9.38	57.79	SYK	spleen tyrosine kinase
219465_at	0.41	6.41	151.99	466.68	APOA2	apolipoprotein A-II
203936_s_at	0.13	6.36	805.09	4046.60	MMP9	matrix metallopeptidase 9
235591_at	0.02	6.32	30.36	188.00	SSTR1	somatostatin receptor 1
220560_at	0.25	6.30	5.24	36.44	C11orf21	chromosome 11 open reading frame 21
204988_at	0.44	6.13	353.45	1164.56	FGB	fibrinogen beta chain
209116_x_at	0.29	6.00	20.49	95.58	HBB	hemoglobin, beta
209458_x_at	0.38	5.84	9.39	58.19	HBA1 /// HBA2	hemoglobin, alpha 1 /// hemoglobin, alpha 2
203447_at	0.49	5.74	304.64	544.84	PSMD5	proteasome (prosome, macropain)
207414_s_at	0.32	5.68	91.41	441.31	PCSK6	proprotein convertase subtilisin/kexin type 6
204638_at	0.07	5.62	10.17	60.57	ACP5	acid phosphatase 5, tartrate resistant
217414_x_at	0.26	5.62	6.85	37.64	HBA1 /// HBA2	hemoglobin, alpha 1 /// hemoglobin, alpha 2
218723_s_at	0.43	5.55	17.97	99.16	C13orf15	chromosome 13 open reading frame 15
202982_s_at	0.33	5.47	30.56	97.21	ACOT1	acyl-CoA thioesterase 1
233814_at	0.43	5.42	6.82	71.95	EFNA5	Ephrin-A5
218711_s_at	0.23	5.34	5.61	35.75	SDPR	serum deprivation response
216604_s_at	0.01	5.31	108.76	571.95	SLC7A8	solute carrier family 7
216603_at	0.00	5.25	32.09	168.59	SLC7A8	solute carrier family 7
210401_at	0.03	5.21	10.74	55.65	P2RX1	purinergic receptor P2X, ligand-gated ion channel, 1
242626_at	0.47	5.21	8.36	127.77	SAMD5	sterile alpha motif domain containing 5
214651_s_at	0.02	5.16	408.30	2063.56	HOXA9	homeobox A9
227488_at	0.24	5.14	101.16	433.18	MGC16121	hypothetical protein MGC16121
209905_at	0.03	5.05	294.95	1436.93	HOXA9	homeobox A9
211743_s_at	0.50	5.04	9.17	63.57	PRG2	proteoglycan 2, bone marrow
203948_s_at	0.50	5.03	3.90	51.39	MPO	myeloperoxidase
207074_s_at	0.11	5.02	123.92	521.93	SLC18A1	solute carrier family 18 (vesicular monoamine), member 1
1556423_at	0.43	4.96	25.45	80.47	VASH1	Vasohibin 1
214778_at	0.48	4.92	168.39	311.73	MEGF8	multiple EGF-like-domains 8
222717_at	0.19	4.91	302.06	1346.91	SDPR	serum deprivation response
223809_at	0.50	4.90	5.10	63.89	RGS18	regulator of G-protein signaling 18
225540_at	0.49	4.89	12.54	183.04	MAP2	microtubule-associated protein 2
204339_s_at	0.49	4.88	5.64	68.86	RGS4	regulator of G-protein signaling 4
209582_s_at	0.36	4.86	20.67	59.79	CD200	CD200 molecule
209660_at	0.42	4.85	109.63	328.39	TTR	transthyretin
206935_at	0.50	4.78	8.07	110.09	PCDH8	protocadherin 8
209735_at	0.38	4.76	518.88	1337.06	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
235412_at	0.31	4.75	2050.8	5542.28	ARHGEF7	Rho guanine nucleotide exchange factor (GEF) 7
202917_s_at	0.50	4.69	3.95	45.59	S100A8	S100 calcium binding protein A8
219195_at	0.45	4.64	241.22	466.90	PPARGC1A	peroxisome proliferator-activated receptor

						gamma, coactivator 1 alpha
205350_at	0.49	4.53	8.52	74.40	CRABP1	cellular retinoic acid binding protein 1
211745_x_at	0.46	4.52	31.72	127.14	HBA1 /// HBA2	hemoglobin, alpha 1 /// hemoglobin, alpha 2
209189_at	0.09	4.51	313.51	1291.13	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
222549_at	0.44	4.37	108.99	474.85	CLDN1	claudin 1
219403_s_at	0.12	4.33	108.97	421.48	HPSE	heparanase
229095_s_at	0.45	4.32	345.59	761.78	LIMS3-LOC440895	LIMS3-LOC440895 read-through
57588_at	0.43	4.29	7.74	61.09	SLC24A3	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3
205898_at	0.37	4.26	245.26	751.74	CX3CR1	chemokine (C-X3-C motif) receptor 1
1554079_at	0.49	4.25	1022.2	1577.99	GALNTL4	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 4
236244_at	0.38	4.18	184.20	412.31	HNRNPU	Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)
1554485_at	0.34	4.14	105.19	403.47	TMEM37	transmembrane protein 37
205033_s_at	0.50	4.14	5.42	49.05	DEFA1	defensin, alpha 1
209469_at	0.41	4.12	8.47	69.70	GPM6A	glycoprotein M6A
213147_at	0.07	4.10	29.26	128.94	HOXA10	homeobox A10
206932_at	0.38	4.07	6.84	48.66	CH25H	cholesterol 25-hydroxylase
238375_at	0.09	4.06	18.57	67.40	RPL22	Ribosomal protein L22
209392_at	0.42	4.00	451.92	1598.75	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2
202877_s_at	0.48	3.89	1173.5	2537.60	CD93	CD93 molecule
202767_at	0.39	3.88	104.99	253.43	ACP2	acid phosphatase 2, lysosomal
38487_at	0.39	3.87	388.99	736.79	STAB1	stabilin 1
231666_at	0.49	3.85	7.08	61.48	PAX3	paired box 3
238850_at	0.50	3.84	7.96	62.51	LOC645323	hypothetical LOC645323
242738_s_at	0.47	3.78	198.85	400.15	ZFHX3	zinc finger homeobox 3
203924_at	0.36	3.76	48.60	128.08	GSTA1	glutathione S-transferase alpha 1
207558_s_at	0.29	3.75	278.58	1034.18	PITX2	paired-like homeodomain 2
206542_s_at	0.36	3.74	192.62	655.64	SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
204850_s_at	0.50	3.74	5.35	40.26	DCX	doublecortin
203305_at	0.50	3.70	4.69	34.49	F13A1	coagulation factor XIII, A1 polypeptide
202409_at	0.01	3.66	3296.2	12149.7	IGF2	insulin-like growth factor 2 (somatomedin A)
1565817_at	0.33	3.64	8.30	48.33	IKZF1	IKAROS family zinc finger 1 (Ikaros)
218746_at	0.07	3.64	8.53	31.60	TAPBPL	TAP binding protein-like
210270_at	0.20	3.63	40.16	127.34	RGS6	regulator of G-protein signaling 6
206049_at	0.48	3.63	3.67	24.31	SELP	selectin P
209919_x_at	0.04	3.63	16.13	57.73	GGT1	gamma-glutamyltransferase 1
240440_at	0.06	3.63	13.02	46.05	NPL	N-acetylneuraminate pyruvate lyase
212353_at	0.23	3.61	305.13	946.05	SULF1	sulfatase 1
202416_at	0.31	3.58	170.53	571.10	DNAJC7	DnaJ (Hsp40) homolog, subfamily C, member 7
205779_at	0.50	3.54	285.34	506.09	RAMP2	receptor (G protein-coupled) activity modifying protein 2

221029_s_at	0.30	3.50	105.91	272.85	WNT5B	wingless-type MMTV integration site family, member 5B
203756_at	0.48	3.49	246.67	376.57	ARHGEF17	Rho guanine nucleotide exchange factor (GEF) 17
217073_x_at	0.43	3.49	459.40	938.15	APOA1	apolipoprotein A-I
240757_at	0.40	3.48	348.18	788.23	CLASP1	Cytoplasmic linker associated protein 1
212810_s_at	0.38	3.48	754.94	1654.74	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4
229159_at	0.30	3.45	52.23	138.43	THSD7A	Thrombospondin, type I, domain containing 7A
209505_at	0.17	3.45	164.34	631.79	NR2F1	Nuclear receptor subfamily 2, group F, member 1
205935_at	0.03	3.44	76.28	257.41	FOXF1	forkhead box F1
223283_s_at	0.30	3.44	122.87	331.55	TSHZ1	teashirt zinc finger homeobox 1
228821_at	0.50	3.43	6.85	42.94	ST6GAL2	ST6 beta-galactosamide alpha-2,6-sialyltranferase 2
202800_at	0.30	3.43	8.67	32.39	SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3
220059_at	0.19	3.42	14.11	46.18	STAP1	signal transducing adaptor family member 1
209583_s_at	0.39	3.42	919.44	2336.00	CD200	CD200 molecule
229125_at	0.36	3.42	13.38	41.64	KANK4	KN motif and ankyrin repeat domains 4
242662_at	0.12	3.41	7.25	25.61	PCSK6	Proprotein convertase subtilisin/kexin type 6
206157_at	0.47	3.41	8.35	50.04	PTX3	pentraxin-related gene, rapidly induced by IL-1 beta
204150_at	0.38	3.41	642.89	1212.82	STAB1	stabilin 1
216956_s_at	0.42	3.40	6.21	28.58	ITGA2B	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)
204959_at	0.50	3.39	6.52	40.72	MNDA	myeloid cell nuclear differentiation antigen
202768_at	0.25	3.39	273.15	883.15	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
218559_s_at	0.04	3.36	1273.7	4317.07	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
244029_at	0.47	3.33	62.70	97.74	LOC100131043	hypothetical LOC100131043
236029_at	0.08	3.32	115.01	411.88	FAT3	FAT tumor suppressor homolog 3 (Drosophila)
205883_at	0.49	3.32	51.92	102.01	ZBTB16	zinc finger and BTB domain containing 16
212354_at	0.08	3.31	218.11	679.50	SULF1	sulfatase 1
222891_s_at	0.25	3.31	9.97	36.96	BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein)
205820_s_at	0.05	3.31	7.87	25.41	APOC3	apolipoprotein C-III
207067_s_at	0.33	3.31	20.21	85.07	HDC	histidine decarboxylase
204537_s_at	0.46	3.28	51.04	78.89	GABRE	gamma-aminobutyric acid (GABA) A receptor, epsilon
204482_at	0.49	3.27	7050.5	11791.9	CLDN5	claudin 5
233825_s_at	0.23	3.26	24.34	69.83	CD99L2	CD99 molecule-like 2
214708_at	0.36	3.26	12.01	41.70	SNTB1	syntrophin, beta 1
228933_at	0.49	3.26	52.68	300.29	NHS	Nance-Horan syndrome
204712_at	0.00	3.24	6.65	21.56	WIF1	WNT inhibitory factor 1
235057_at	0.45	3.23	113.87	184.52	ITCH	itchy E3 ubiquitin protein ligase homolog (mouse)
205609_at	0.35	3.20	136.77	439.06	ANGPT1	angiopoietin 1
228653_at	0.18	3.20	225.28	624.70	SAMD5	sterile alpha motif domain containing 5
201149_s_at	0.44	3.19	81.83	151.65	TIMP3	TIMP metallopeptidase inhibitor 3
204851_s_at	0.50	3.19	5.17	28.89	DCX	doublecortin
230008_at	0.37	3.18	134.22	266.81	THSD7A	thrombospondin, type I, domain containing 7A

207131_x_at	0.03	3.17	19.21	60.97	GGT1	gamma-glutamyltransferase 1
205522_at	0.50	3.16	4.02	22.06	HOXD4	homeobox D4
1567224_at	0.20	3.15	66.37	169.78	HMGA2	high mobility group AT-hook 2
206207_at	0.50	3.15	3.88	21.24	CLC	Charcot-Leyden crystal protein
206858_s_at	0.00	3.14	100.57	313.98	HOXC6	homeobox C6
228402_at	0.49	3.11	57.98	131.04	ZBED3	zinc finger, BED-type containing 3
204304_s_at	0.25	3.08	75.70	254.26	PROM1	prominin 1
203939_at	0.47	3.08	2441.1	4832.70	NT5E	5'-nucleotidase, ecto (CD73)
204446_s_at	0.39	3.06	28.43	61.74	ALOX5	arachidonate 5-lipoxygenase
224673_at	0.03	3.06	9.83	29.62	LENG8	leukocyte receptor cluster (LRC) member 8
221667_s_at	0.36	3.04	123.67	314.37	HSPB8	heat shock 22kDa protein 8
229168_at	0.23	3.03	60.26	141.93	COL23A1	collagen, type XXIII, alpha 1
203908_at	0.13	3.03	40.21	120.66	SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4
1555745_at	0.49	3.03	8.09	40.14	LYZ	lysozyme (renal amyloidosis)
206584_at	0.09	3.02	6.83	20.63	LY96	lymphocyte antigen 96
224566_at	0.24	3.00	136.74	335.08	NCRNA00084	non-protein coding RNA 84
222670_s_at	0.03	2.99	528.82	1592.74	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
208096_s_at	0.41	2.99	19.25	78.11	COL21A1	collagen, type XXI, alpha 1
225095_at	0.45	2.98	131.90	291.36	SPTLC2	Serine palmitoyltransferase, long chain base subunit 2
211250_s_at	0.34	2.98	66.13	132.68	SH3BP2	SH3-domain binding protein 2
221210_s_at	0.03	2.98	16.97	49.76	NPL	N-acetylneuraminate pyruvate lyase (dihydrodipicolinate synthase)
236892_s_at	0.17	2.96	106.23	317.12	LOC100130740	similar to hCG2042068
212951_at	0.49	2.96	1261.6	2381.51	GPR116	G protein-coupled receptor 116
222908_at	0.09	2.95	8.46	25.37	FAM38B	family with sequence similarity 38, member B
226703_at	0.24	2.94	20.21	46.83	NEURL4	neuralized homolog 4 (Drosophila)
217232_x_at	0.05	2.93	10.53	30.97	HBB	hemoglobin, beta
232898_at	0.28	2.91	42.68	108.62	DAB2	disabled homolog 2, mitogen-responsive phosphoprotein
204018_x_at	0.43	2.89	5.78	19.90	HBA1 /// HBA2	hemoglobin, alpha 1 /// hemoglobin, alpha 2
220918_at	0.01	2.89	5.61	16.20	C21orf96	chromosome 21 open reading frame 96
235795_at	0.50	2.88	7.13	32.87	PAX6	paired box 6
203477_at	0.46	2.88	718.26	2035.50	COL15A1	collagen, type XV, alpha 1
217419_x_at	0.30	2.87	379.17	757.98	AGRN	agrin
209560_s_at	0.19	2.87	1542.2	4499.97	DLK1	delta-like 1 homolog (Drosophila)
209581_at	0.32	2.86	122.77	324.81	PLA2G16	phospholipase A2, group XVI
205984_at	0.41	2.85	374.43	1266.51	CRHBP	corticotropin releasing hormone binding protein
229127_at	0.48	2.85	12.49	53.89	JAM2	junctional adhesion molecule 2
218693_at	0.18	2.83	211.33	567.51	TSPAN15	tetraspanin 15

Table 41: List of Transcripts from Day6toDay8Up2fold+HaemSim+p0.5 Gene List

The top 200 transcripts within the gene list Day6toDay8Up2fold+HaemSim+p0.5 are displayed and sorted in order of fold change (highest at the top).

qPCR Primer Sequences

Primer Name	Forward Sequence	Reverse Sequence	Product Length	AT	MT
AML1 (Iso A)	AAGACACAGCACCCTGGAGA	GCCTTCCTCATAACGTGCAT	151	62	80
BRACHYURY	TCAGCAAAGTCAAGCTCACCA	CCCCAACTCTCACTATGTGGATT	102	65	80
CCND1	ACCAGCTCCTGTGCTGCGAAGTG	GACGGCAGGACCTCCTTCTGCACA	214	64	80
CD117	gagtgaagtgaatgttgctgag	caagtgaattgcagtccttcc	169	60	81
CD144	GATCAAGTCAAGCGTGAGTCG	AGCCTCTCAATGGCGAACAC	181	58	81
CD31	GCTGACCCTTCTGCTCTGTT	TGAGAGGTGGTGCTGACATC	186	55	83
CD34	GAGGCCTCAGTGTCTACTGC	CGATTCATCAGGAAATAGCC	228	65	80
CER1	GAGAGCATCCGGAAGAAGC	TGGCCGACGTTCATCATGC	184	61	80
FOXA2	AAGGCATACGAACAGGCACTG	TACACACCTTGGTAGTACGCC	140	65	85
FOXD3	CCTAAGCTGGTCGAGCAAAC	TCGGTTTTCGGTTTTACCTG	73	65	80
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG	86	53	81
GATA4	ACACCCCAATCTCGATATGTTTG	GTTGCACAGATAGTGACCCGT	210	58	81
GFAP	AAGAGATCCGCACGCAGTAT	AAGAGATCCGCACGCAGTAT	174	68	86
HBA	ACTCTTCTGGTCCCCACAG	GTGGGGAAGGACAGGAAC	153	57	85
HBB	GCAACCTCAAACAGACACCA	TTAGGGTTGCCCATAACAGC	198	65	84
HBD	TCCTGAGGAGAAGACTGCTGTCAATGC	GAGAGGACAGATCCCCAAAGGACTCAAA	140	57	83
HBE1	ATATCTGCTTCCGACACAGC	GCTTGAGGTTGTCCATGTTT	303	65	81
HBG	GGACAAGGCTACTATCACAAGC	GGAAGTCAGCACCTTCTTGC	193	65	86
HBG1	CACAGAGGAGGACAAGGCTA	CTTGAGATCATCCAGGTGCT	238	65	85
KDR	CACCACTCAAACGCTGACATGTA	GCTCGTTGGCGCACTCTT	95	63	81
MIXL1	CAGAGTGGGAAATCCTTCCA	TTCAGAGAGAGGGGAACAGG	207	65	82
NANOG	GATTTGTGGGCCTGAAGAAA	AAGTGGGTTGTTTGCCTTTG	155	65	81
NESTIN	CAGGAGAAACAGGGCCTACA	TGGGAGCAAAGATCCAAGAC	243	65	88
OCT4-TV1	CGTGAAGCTGGAGAAGGAGA	CTTGGCAAATTGCTCGAGTT	91	65	82
PAX6-TV1	GAGGTCAGGCTTCGCTAATG	TGGTGATGGCTCAAGTGTGT	91	52	85
RPLI3A	CCTGGAGGAGAAGAGGAAAGAGA	TTGAGGACCTCTGTGTATTTGTCAA	126	65	80
SCL	AGCCGGATGCCTTCCCTAT	CCGCACAACTTTGGTGTGG	104	53	80
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG	84	65	76
TBP (GEN)	ATGTTTTTCCCCATGAACCA	TCTGGCACAGAAATAACCCC	112	65	73

Table 42: Table of Real-time Quantitative Polymerase Chain Reaction (qPCR) Primers

These qPCR primers were designed using Invitrogen's OligoPerfect Designer [234] . The annealing temperature (AT) and melting temperature (MT) were both verified and optimised by Dr. Stuart Atkinson using reference cDNA.

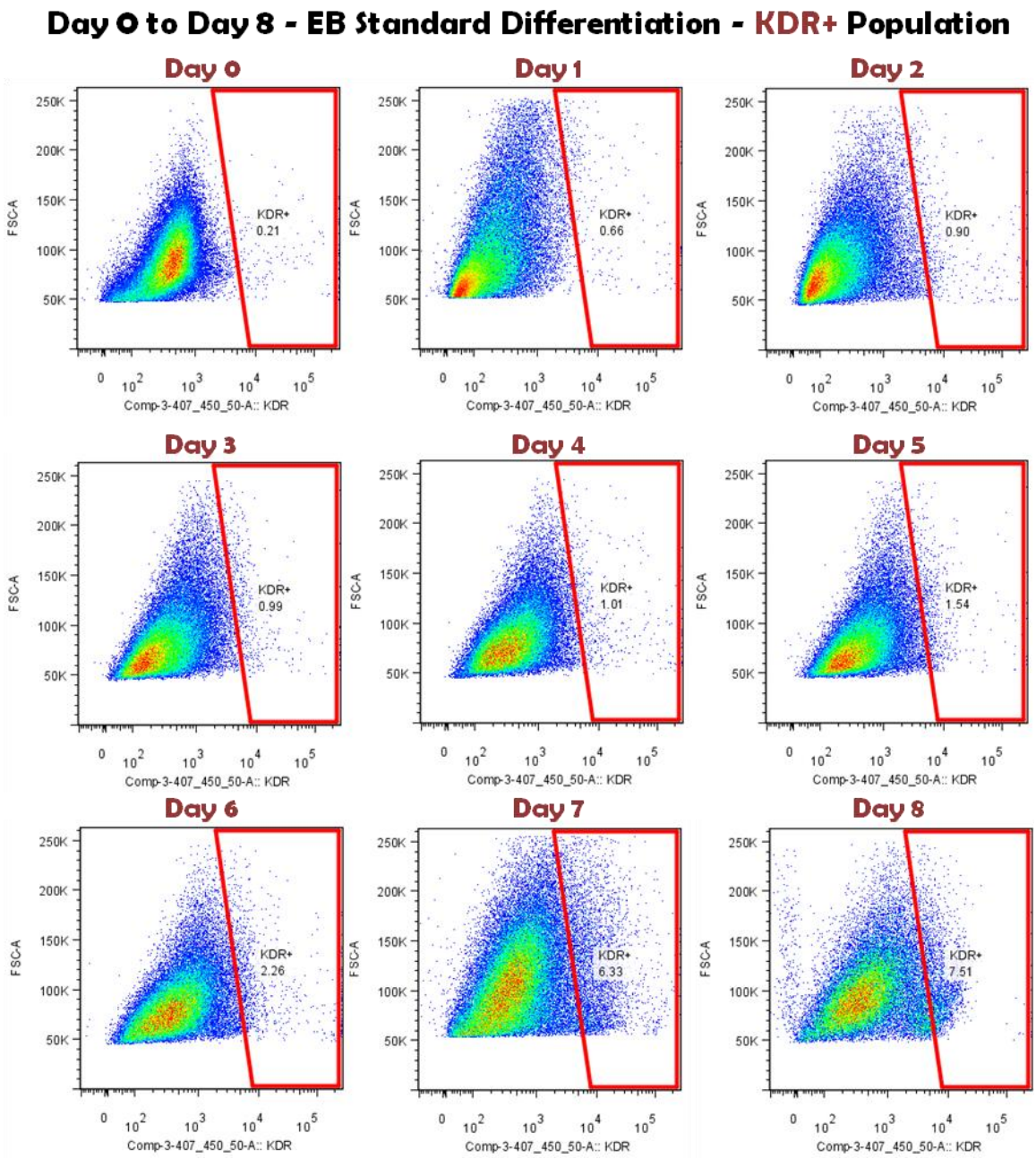


Figure 90: Kinetics of KDR+ population during FBS-mediated differentiation

The percentages of KDR-positive cells were measured by flow cytometry over the course of 8 days of differentiation using FBS based differentiation media. Isotypes were used to represent the negative population and for the application of the positive gates. 200,000 events were recorded.

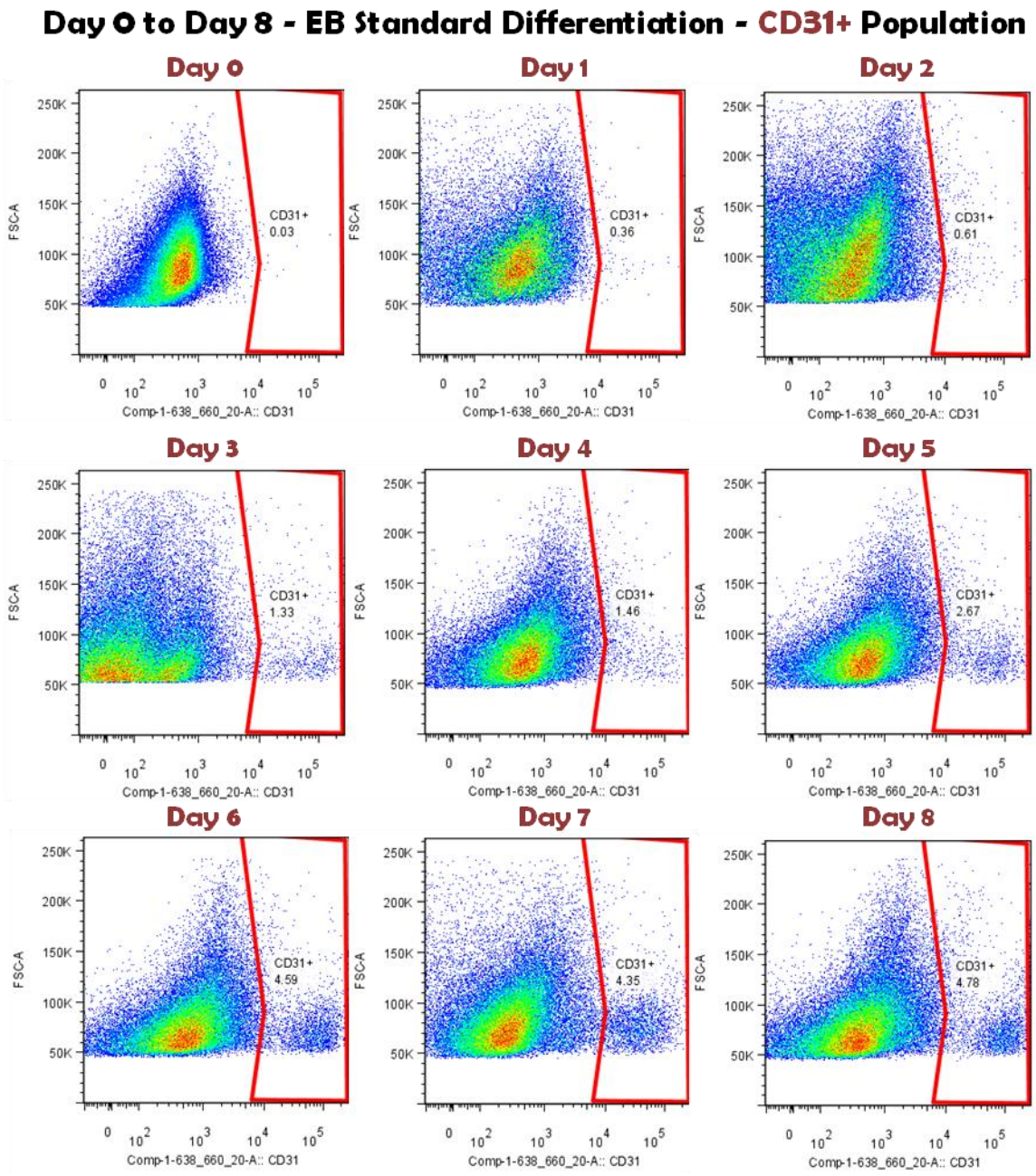


Figure 91: Kinetics of CD31+ population during FBS-mediated differentiation

The percentages of CD31-positive cells were measured by flow cytometry over the course of 8 days of differentiation using FBS based differentiation media. Isotypes were used to represent the negative population and for the application of the positive gates. 200,000 events were recorded.

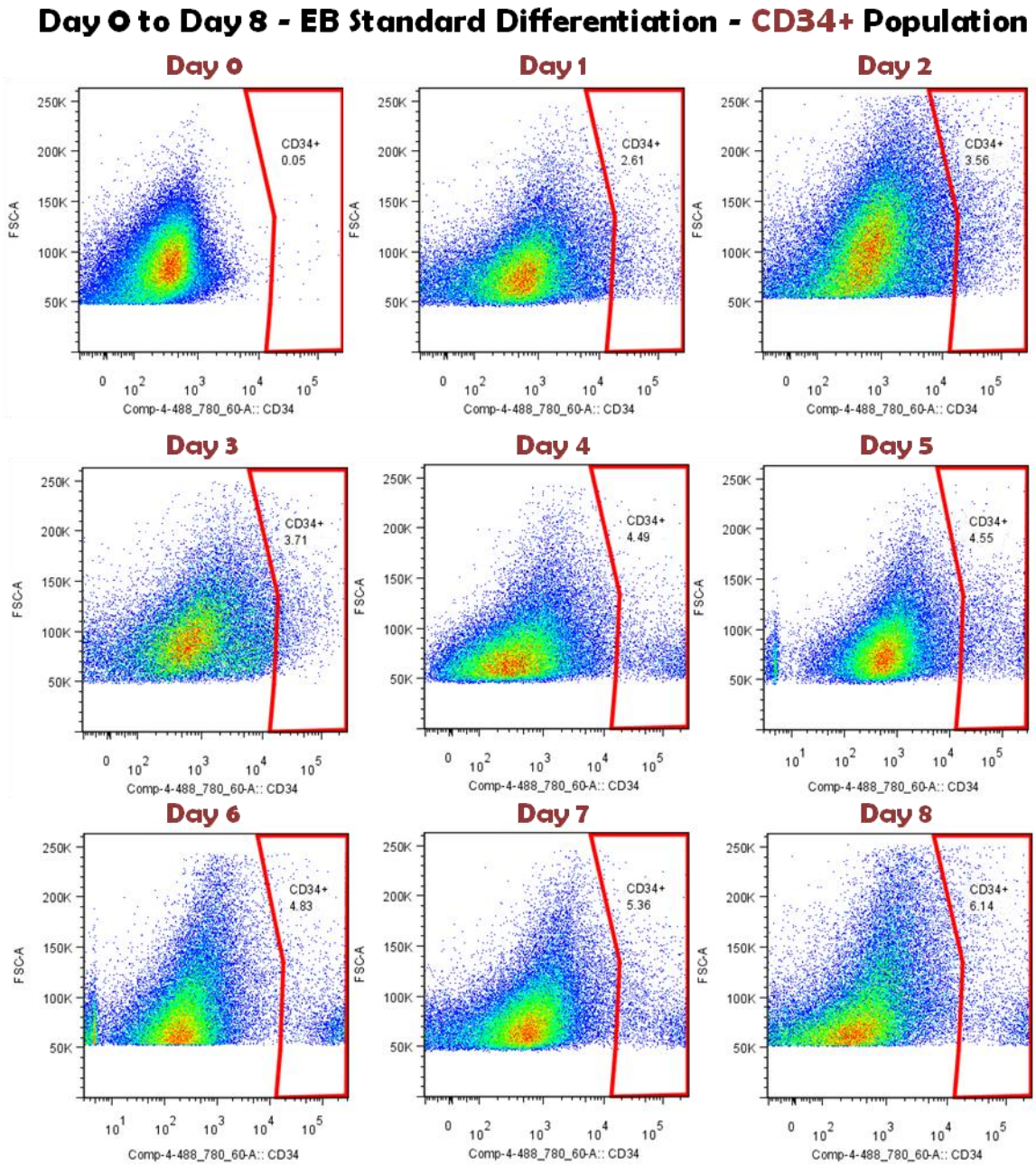


Figure 92: Kinetics of CD34+ population during FBS-mediated differentiation

The percentages of CD34-positive cells were measured by flow cytometry over the course of 8 days of differentiation using FBS based differentiation media. Isotypes were used to represent the negative population and for the application of the positive gates. 200,000 events were recorded.

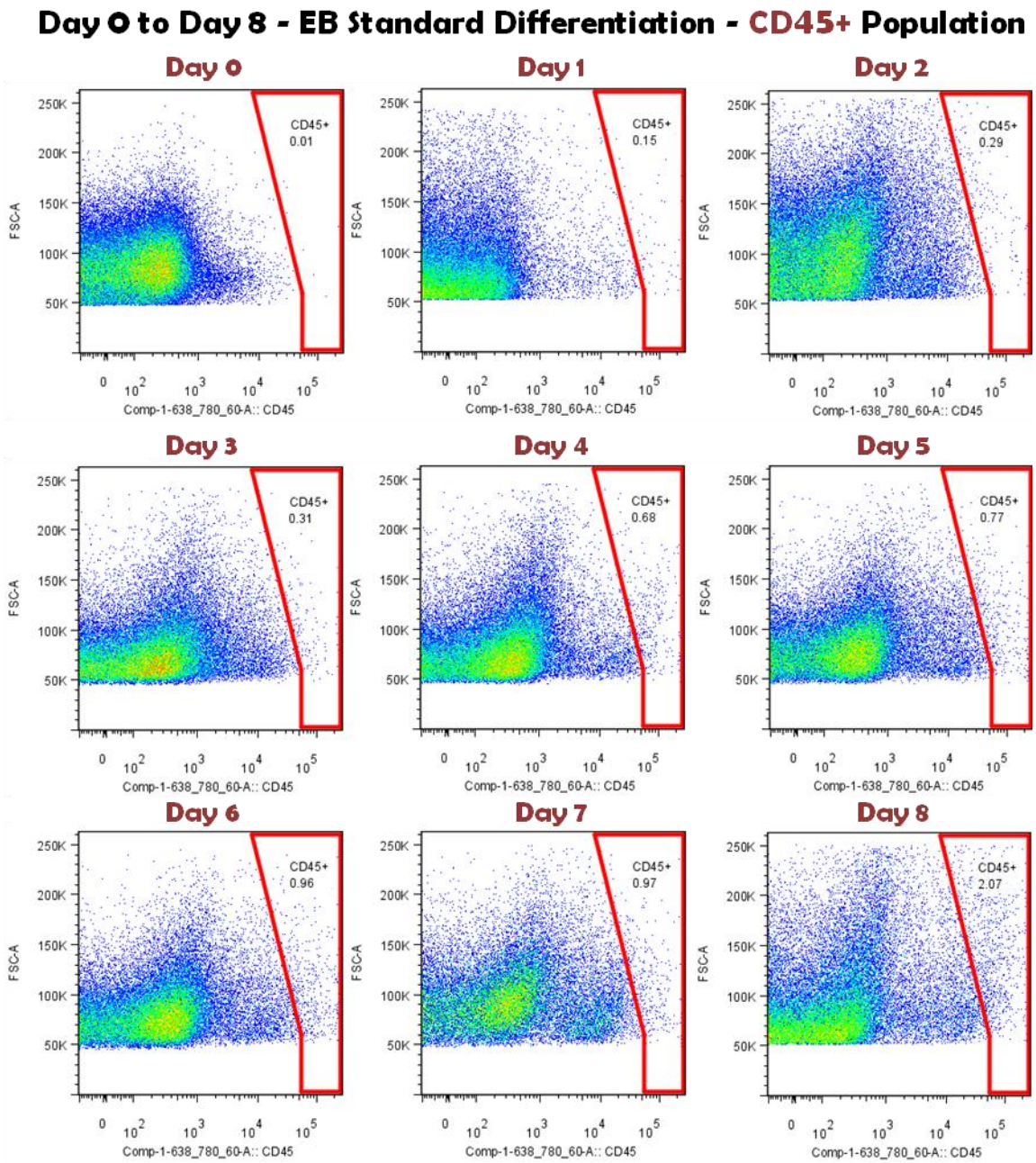


Figure 93: Kinetics of CD45+ population during FBS-mediated differentiation

The percentages of CD45-positive cells were measured by flow cytometry over the course of 8 days of differentiation using FBS based differentiation media. Isotypes were used to represent the negative population and for the application of the positive gates. 200,000 events were recorded.

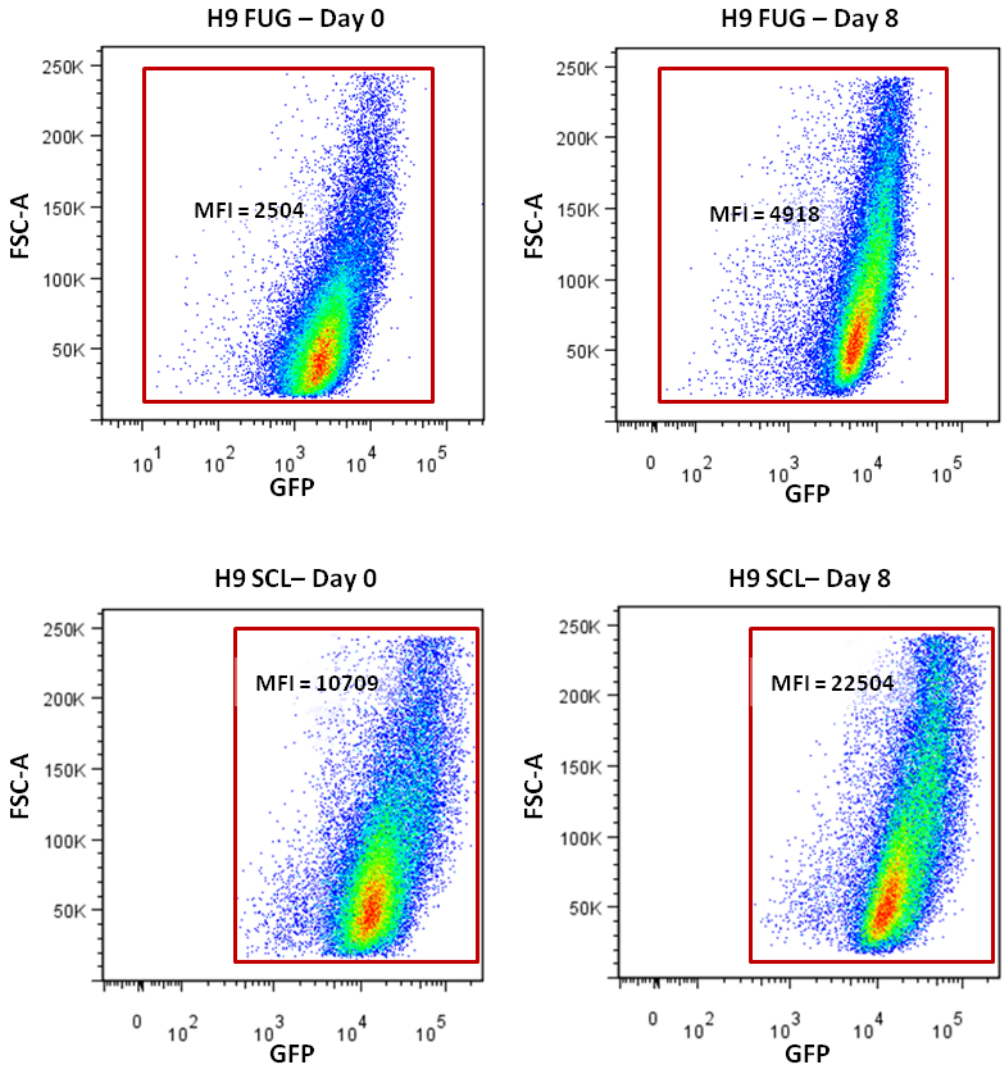


Figure 94: Mean Fluorescence Intensity of GFP between undifferentiated and differentiated H9 FUG/H9 SCL cell lines.

Flow cytometry analysis between undifferentiated (day 0) and differentiated (day 8) H9SCL/H9 FUG cells reveal a large difference between the strength of GFP fluorescence base on the mean fluorescence intensity (MFI) of the ungated population of cells. The MFI value of H9 SCL cell line is approx. 5 fold greater than that of H9 FUG, and both increases by over 2 fold from day 0 to day 8 of differentiation

REFERENCES

1. Bibikova, M., et al., *Human embryonic stem cells have a unique epigenetic signature.* Genome Res, 2006. **16**(9): p. 1075-83.
2. Findikli, N., et al., *Establishment and characterization of new human embryonic stem cell lines.* Reprod Biomed Online, 2005. **10**(5): p. 617-27.
3. Reubinoff, B.E., et al., *Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro.* Nat Biotechnol, 2000. **18**(4): p. 399-404.
4. Wikipedia. *Ectoderm*. 2007 [cited 2007 13th June]; Available from: http://en.wikipedia.org/wiki/Ectoderm.
5. Pera, M.F., B. Reubinoff, and A. Trounson, *Human embryonic stem cells.* J Cell Sci, 2000. **113 (Pt 1)**: p. 5-10.
6. Winslow, T. and C. Duckwall. *Stem Cells: Scientific Progress and Future Research Directions*. 2001 [cited 2007 13th June]; Available from: http://stemcells.nih.gov/info/scireport/chapter1.asp.
7. NIH. *Stem Cell Information: Stem Cells and Diseases*. 2007; Available from: http://stemcells.nih.gov/info/health.asp.
8. McWilliams, R.R., et al., *Phase II Trial of Oxaliplatin/Irinotecan/5-Fluorouracil/Leucovorin for Metastatic Colorectal Cancer.* Clin Colorectal Cancer, 2007. **7**(1): p. 516-21.
9. NIH. *Stem Cell Information: Stem Cell Basics*. 2007 [cited 2007 17th June]; Available from: http://stemcells.nih.gov/info/basics/basics6.asp.
10. Jackson, K.A., et al., *Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells.* J Clin Invest, 2001. **107**(11): p. 1395-402.
11. Geron. *Geron to Proceed with First Human Clinical Trial of Embryonic Stem Cell-Based Therapy*. 2010 [cited 2010 08/09]; Available from: http://www.geron.com/media/pressview.aspx?id=1229.
12. Keirstead, H.S., et al., *Human embryonic stem cell-derived oligodendrocyte progenitor cell transplants remyelinate and restore locomotion after spinal cord injury.* J Neurosci, 2005. **25**(19): p. 4694-705.
13. Klimanskaya, I., et al., *Human embryonic stem cells derived without feeder cells.* Lancet, 2005. **365**(9471): p. 1636-41.
14. Amit, M. and J. Itskovitz-Eldor, *Feeder-free culture of human embryonic stem cells.* Methods Enzymol, 2006. **420**: p. 37-49.
15. Xu, C., et al., *Feeder-free growth of undifferentiated human embryonic stem cells.* Nat Biotechnol, 2001. **19**(10): p. 971-4.
16. Rosler, E.S., et al., *Long-term culture of human embryonic stem cells in feeder-free conditions.* Dev Dyn, 2004. **229**(2): p. 259-74.
17. Amit, M., et al., *Human feeder layers for human embryonic stem cells.* Biol Reprod, 2003. **68**(6): p. 2150-6.
18. Hovatta, O., et al., *A culture system using human foreskin fibroblasts as feeder cells allows production of human embryonic stem cells.* Hum Reprod, 2003. **18**(7): p. 1404-9.
19. Lee, J.B., et al., *Establishment and maintenance of human embryonic stem cell lines on human feeder cells derived from uterine endometrium under serum-free condition.* Biol Reprod, 2005. **72**(1): p. 42-9.
20. Richards, M., et al., *Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells.* Nat Biotechnol, 2002. **20**(9): p. 933-6.
21. Unger, C., et al., *Good manufacturing practice and clinical-grade human embryonic stem cell lines.* Hum Mol Genet, 2008. **17**(R1): p. R48-53.
22. Sidhu, K.S., S. Walke, and B.E. Tuch, *Derivation and propagation of hESC under a therapeutic environment.* Curr Protoc Stem Cell Biol, 2008. **Chapter 1**: p. Unit 1A 4.
23. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
24. Veeck, L. and N. Zaninovic, *An Atlas of Human Blastocysts*. 2003: Parthenon Publishing.
25. Simon, C., et al., *First derivation in Spain of human embryonic stem cell lines: use of long-term cryopreserved embryos and animal-free conditions.* Fertil Steril, 2005. **83**(1): p. 246-9.
26. Thomson, J.A. and J.S. Odorico, *Human embryonic stem cell and embryonic germ cell lines.* Trends Biotechnol, 2000. **18**(2): p. 53-7.

27. Draper, J.S., et al., *Surface antigens of human embryonic stem cells: changes upon differentiation in culture.* J Anat, 2002. **200**(Pt 3): p. 249-58.
28. Shevinsky, L.H., et al., *Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells.* Cell, 1982. **30**(3): p. 697-705.
29. Kannagi, R., et al., *Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells.* Embo J, 1983. **2**(12): p. 2355-61.
30. Yin, A.H., et al., *AC133, a novel marker for human hematopoietic stem and progenitor cells.* Blood, 1997. **90**(12): p. 5002-12.
31. Hoffman, L.M. and M.K. Carpenter, *Characterization and culture of human embryonic stem cells.* Nat Biotechnol, 2005. **23**(6): p. 699-708.
32. Rosnet, O., et al., *Expression and signal transduction of the FLT3 tyrosine kinase receptor.* Acta Haematol, 1996. **95**(3-4): p. 218-23.
33. Takeda, J., S. Seino, and G.I. Bell, *Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues.* Nucleic Acids Res, 1992. **20**(17): p. 4613-20.
34. Scholer, H.R., et al., *A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor.* Embo J, 1989. **8**(9): p. 2543-50.
35. Rosner, M.H., et al., *A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo.* Nature, 1990. **345**(6277): p. 686-92.
36. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
37. Nichols, J., et al., *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. **95**(3): p. 379-91.
38. Pan, G. and J.A. Thomson, *Nanog and transcriptional networks in embryonic stem cell pluripotency.* Cell Res, 2007. **17**(1): p. 42-9.
39. Richards, M., et al., *The transcriptome profile of human embryonic stem cells as defined by SAGE.* Stem Cells, 2004. **22**(1): p. 51-64.
40. Schuldiner, M., et al., *Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells.* Proc Natl Acad Sci U S A, 2000. **97**(21): p. 11307-12.
41. Itskovitz-Eldor, J., et al., *Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.* Mol Med, 2000. **6**(2): p. 88-95.
42. Dvash, T., et al., *Temporal gene expression during differentiation of human embryonic stem cells and embryoid bodies.* Hum Reprod, 2004. **19**(12): p. 2875-83.
43. Xu, C., et al., *Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells.* Circ Res, 2002. **91**(6): p. 501-8.
44. Mummery, C., et al., *Cardiomyocyte differentiation of mouse and human embryonic stem cells.* J Anat, 2002. **200**(Pt 3): p. 233-42.
45. Kehat, I., et al., *Development of cardiomyocytes from human ES cells.* Methods Enzymol, 2003. **365**: p. 461-73.
46. He, J.Q., et al., *Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization.* Circ Res, 2003. **93**(1): p. 32-9.
47. Satin, J., et al., *Mechanism of spontaneous excitability in human embryonic stem cell derived cardiomyocytes.* J Physiol, 2004. **559**(Pt 2): p. 479-96.
48. Xue, T., et al., *Functional integration of electrically active cardiac derivatives from genetically engineered human embryonic stem cells with quiescent recipient ventricular cardiomyocytes: insights into the development of cell-based pacemakers.* Circulation, 2005. **111**(1): p. 11-20.
49. Kaufman, D.S., et al., *Hematopoietic colony-forming cells derived from human embryonic stem cells.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10716-21.
50. Vodyanik, M.A., et al., *Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential.* Blood, 2005. **105**(2): p. 617-26.
51. Zhan, X., et al., *Functional antigen-presenting leucocytes derived from human embryonic stem cells in vitro.* Lancet, 2004. **364**(9429): p. 163-71.
52. Levenberg, S., et al., *Endothelial cells derived from human embryonic stem cells.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4391-6.

53. Wang, L., et al., *Endothelial and hematopoietic cell fate of human embryonic stem cells originates from primitive endothelium with hemangioblastic properties.* Immunity, 2004. **21**(1): p. 31-41.
54. Gerecht-Nir, S., et al., *Vascular gene expression and phenotypic correlation during differentiation of human embryonic stem cells.* Dev Dyn, 2005. **232**(2): p. 487-97.
55. Gerecht-Nir, S., et al., *Vascular development in early human embryos and in teratomas derived from human embryonic stem cells.* Biol Reprod, 2004. **71**(6): p. 2029-36.
56. Gerecht-Nir, S., et al., *Three-dimensional porous alginate scaffolds provide a conducive environment for generation of well-vascularized embryoid bodies from human embryonic stem cells.* Biotechnol Bioeng, 2004. **88**(3): p. 313-20.
57. Schuldiner, M., et al., *Induced neuronal differentiation of human embryonic stem cells.* Brain Res, 2001. **913**(2): p. 201-5.
58. Carpenter, M.K., et al., *Enrichment of neurons and neural precursors from human embryonic stem cells.* Exp Neurol, 2001. **172**(2): p. 383-97.
59. Zhang, S.C., et al., *In vitro differentiation of transplantable neural precursors from human embryonic stem cells.* Nat Biotechnol, 2001. **19**(12): p. 1129-33.
60. Reubinoff, B.E., et al., *Neural progenitors from human embryonic stem cells.* Nat Biotechnol, 2001. **19**(12): p. 1134-40.
61. Schulz, T.C., et al., *Directed neuronal differentiation of human embryonic stem cells.* BMC Neurosci, 2003. **4**: p. 27.
62. Ben-Hur, T., et al., *Transplantation of human embryonic stem cell-derived neural progenitors improves behavioral deficit in Parkinsonian rats.* Stem Cells, 2004. **22**(7): p. 1246-55.
63. Pera, M.F., et al., *Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin.* J Cell Sci, 2004. **117**(Pt 7): p. 1269-80.
64. Perrier, A.L., et al., *Derivation of midbrain dopamine neurons from human embryonic stem cells.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12543-8.
65. Schulz, T.C., et al., *Differentiation of human embryonic stem cells to dopaminergic neurons in serum-free suspension culture.* Stem Cells, 2004. **22**(7): p. 1218-38.
66. Li, X.J., et al., *Specification of motoneurons from human embryonic stem cells.* Nat Biotechnol, 2005. **23**(2): p. 215-21.
67. Nistor, G.I., et al., *Human embryonic stem cells differentiate into oligodendrocytes in high purity and myelinate after spinal cord transplantation.* Glia, 2005. **49**(3): p. 385-96.
68. Segev, H., et al., *Differentiation of human embryonic stem cells into insulin-producing clusters.* Stem Cells, 2004. **22**(3): p. 265-74.
69. Rambhatla, L., et al., *Generation of hepatocyte-like cells from human embryonic stem cells.* Cell Transplant, 2003. **12**(1): p. 1-11.
70. Xu, R.H., et al., *BMP4 initiates human embryonic stem cell differentiation to trophoblast.* Nat Biotechnol, 2002. **20**(12): p. 1261-4.
71. Gerami-Naini, B., et al., *Trophoblast differentiation in embryoid bodies derived from human embryonic stem cells.* Endocrinology, 2004. **145**(4): p. 1517-24.
72. Clark, A.T., et al., *Spontaneous differentiation of germ cells from human embryonic stem cells in vitro.* Hum Mol Genet, 2004. **13**(7): p. 727-39.
73. Cerdan, C., A. Rouleau, and M. Bhatia, *VEGF-A165 augments erythropoietic development from human embryonic stem cells.* Blood, 2004. **103**(7): p. 2504-12.
74. Chadwick, K., et al., *Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells.* Blood, 2003. **102**(3): p. 906-15.
75. Chung, Y.S., et al., *Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression.* Development, 2002. **129**(23): p. 5511-20.
76. Conley, B.J., et al., *Derivation, propagation and differentiation of human embryonic stem cells.* Int J Biochem Cell Biol, 2004. **36**(4): p. 555-67.
77. Zhang, X., et al., *Derivation of human embryonic stem cells from developing and arrested embryos.* Stem Cells, 2006. **24**(12): p. 2669-76.
78. Pearson, H. and A. Abbott. *Stem cells derived from 'dead' human embryo*. 2006 27 September 2006 [cited 2007 6th June]; Available from: http://www.nature.com/nature/journal/v443/n7110/full/443376a.html.
79. Landry, D.W., et al., *Hypocellularity and absence of compaction as criteria for embryonic death.* Regen Med, 2006. **1**(3): p. 367-71.

80. BBC. *Ministers bow to hybrid pressure*. 2007 [cited 2007 9th June]; Ministers have bowed to pressure to allow the creation of human-animal hybrid embryos for research.]. Available from: http://news.bbc.co.uk/1/hi/health/6661717.stm.
81. BBC. *Q&A: Hybrid embryos*. 2007 [cited 2007 09th June]; Available from: http://news.bbc.co.uk/1/hi/health/6233415.stm.
82. Telegraph.co.uk. *Hybrid embryos made by UK scientists*. 2008 [cited 2010 09/09/10]; Available from: http://www.telegraph.co.uk/science/science-news/3338000/Hybrid-embryos-made-by-UK-scientists.html.
83. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
84. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007.
85. Wernig, M., et al., *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.* Nature, 2007.
86. Maherali, N., et al., *Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution.* Cell Stem Cell, 2007. **1**: p. 55-70.
87. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
88. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
89. Dean, L., *Blood Groups and Red Cell Antigens*. 2005, Bethesda: NCBI.
90. Wan, J., W.D. Ristenpart, and H.A. Stone, *Dynamics of shear-induced ATP release from red blood cells.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16432-7.
91. Jiang, N., et al., *Respiratory protein-generated reactive oxygen species as an antimicrobial strategy.* Nat Immunol, 2007. **8**(10): p. 1114-22.
92. Roitt, I., Brostoff, J., Male, D, *Immunology (6th edition)*. 2001: St. Louis: Mosby.
93. Kufe, D.W., Pollock R.E., Weichselbaum, R.R., Bast, R.C., Gansler, T.S., Holland, J.F., Frei, E., *Holland-Frei Cancer Medicine, 6th edition*. 2003: American Cancer Society.
94. Schroeder, J.T., *Basophils beyond effector cells of allergic inflammation.* Adv Immunol, 2009. **101**: p. 123-61.
95. Reece, J., Campbell, N., *Biology*. 2002, San Francisco: Benjamin Cummings.
96. Ziegler-Heitbrock, L., *The CD14+ CD16+ blood monocytes: their role in infection and inflammation.* J Leukoc Biol, 2007. **81**(3): p. 584-92.
97. Prussin, C. and D.D. Metcalfe, *4. IgE, mast cells, basophils, and eosinophils.* J Allergy Clin Immunol, 2003. **111**(2 Suppl): p. S486-94.
98. Jolly, J., Ferester-Tadie, M., *Recherches sur l'oeuf du rat et de la souris.* Archs Anat. microsc. Morpho. exp, 1936. **32**: p. 67.
99. Tam, P.P. and R.S. Beddington, *The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis.* Development, 1987. **99**(1): p. 109-26.
100. Funayama, N., et al., *Coelom formation: binary decision of the lateral plate mesoderm is controlled by the ectoderm.* Development, 1999. **126**(18): p. 4129-38.
101. Willey, S., et al., *Acceleration of mesoderm development and expansion of hematopoietic progenitors in differentiating ES cells by the mouse Mix-like homeodomain transcription factor.* Blood, 2006. **107**(8): p. 3122-30.
102. McGrath, K.E. and J. Palis, *Hematopoiesis in the yolk sac: more than meets the eye.* Exp Hematol, 2005. **33**(9): p. 1021-8.
103. Carlson, B.M., *Patten's Foundations of Embryology*. 1981, New York: McGraw-Hill.
104. Krens, S.F., et al., *Distinct functions for ERK1 and ERK2 in cell migration processes during zebrafish gastrulation.* Dev Biol, 2008. **319**(2): p. 370-83.
105. Yao, Y., et al., *Extracellular signal-regulated kinase 2 is necessary for mesoderm differentiation.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12759-64.
106. Krens, S.F., et al., *ERK1 and ERK2 MAPK are key regulators of distinct gene sets in zebrafish embryogenesis.* BMC Genomics, 2008. **9**: p. 196.
107. Sadlon, T.J., I.D. Lewis, and R.J. D'Andrea, *BMP4: its role in development of the hematopoietic system and potential as a hematopoietic growth factor.* Stem Cells, 2004. **22**(4): p. 457-74.
108. Pearson, S., et al., *The stepwise specification of embryonic stem cells to hematopoietic fate is driven by sequential exposure to Bmp4, activin A, bFGF and VEGF.* Development, 2008. **135**(8): p. 1525-35.

109. Amaya, E., et al., *FGF signalling in the early specification of mesoderm in Xenopus.* Development, 1993. **118**(2): p. 477-87.
110. Schulte-Merker, S. and J.C. Smith, *Mesoderm formation in response to Brachyury requires FGF signalling.* Curr Biol, 1995. **5**(1): p. 62-7.
111. Zhang, P., et al., *Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells.* Blood, 2008. **111**(4): p. 1933-41.
112. Zhou, Q., et al., *ERK signaling is a central regulator for BMP-4 dependent capillary sprouting.* Cardiovasc Res, 2007. **76**(3): p. 390-9.
113. Koster, M., et al., *Bone morphogenetic protein 4 (BMP-4), a member of the TGF-beta family, in early embryos of Xenopus laevis: analysis of mesoderm inducing activity.* Mech Dev, 1991. **33**(3): p. 191-9.
114. Miyazono, K., K. Kusanagi, and H. Inoue, *Divergence and convergence of TGF-beta/BMP signaling.* J Cell Physiol, 2001. **187**(3): p. 265-76.
115. Winnier, G., et al., *Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse.* Genes Dev, 1995. **9**(17): p. 2105-16.
116. Anderson, G.J. and D. Darshan, *Small-molecule dissection of BMP signaling.* Nat Chem Biol, 2008. **4**(1): p. 15-6.
117. Chang, H., et al., *Smad5 knockout mice die at mid-gestation due to multiple embryonic and extraembryonic defects.* Development, 1999. **126**(8): p. 1631-42.
118. Lechleider, R.J., et al., *Targeted mutagenesis of Smad1 reveals an essential role in chorioallantoic fusion.* Dev Biol, 2001. **240**(1): p. 157-67.
119. McReynolds, L.J., et al., *Smad1 and Smad5 differentially regulate embryonic hematopoiesis.* Blood, 2007. **110**(12): p. 3881-90.
120. Edwards, Y.H., et al., *The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27.* Genome Res, 1996. **6**(3): p. 226-33.
121. Beddington, R.S., P. Rashbass, and V. Wilson, *Brachyury--a gene affecting mouse gastrulation and early organogenesis.* Dev Suppl, 1992: p. 157-65.
122. Halpern, M.E., et al., *Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation.* Cell, 1993. **75**(1): p. 99-111.
123. Yamaguchi, T.P., et al., *T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification.* Genes Dev, 1999. **13**(24): p. 3185-90.
124. Technology, C.S. *Wnt/β-Catenin Signaling* 2010 [cited 2010 15/02]; Available from: http://www.cellsignal.com/reference/pathway/Wnt_beta_Catenin.html.
125. Hart, A.H., et al., *Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo.* Development, 2002. **129**(15): p. 3597-608.
126. Mead, P.E., et al., *BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1.* Nature, 1996. **382**(6589): p. 357-60.
127. Hart, A.H., et al., *Transcriptional regulation of the homeobox gene Mixl1 by TGF-beta and FoxH1.* Biochem Biophys Res Commun, 2005. **333**(4): p. 1361-9.
128. Guo, W., et al., *A human Mix-like homeobox gene MIXL shows functional similarity to Xenopus Mix.1.* Blood, 2002. **100**(1): p. 89-95.
129. Swain, J.L., et al., *Disruption of the purine nucleotide cycle by inhibition of adenylosuccinate lyase produces skeletal muscle dysfunction.* J Clin Invest, 1984. **74**(4): p. 1422-7.
130. Moore, M.A. and J.J. Owen, *Chromosome marker studies on the development of the haemopoietic system in the chick embryo.* Nature, 1965. **208**(5014): p. 956 passim.
131. Houssaint, E., *Differentiation of the mouse hepatic primordium. II. Extrinsic origin of the haemopoietic cell line.* Cell Differ, 1981. **10**(5): p. 243-52.
132. Dieterlen-Lievre, F., *On the origin of haemopoietic stem cells in the avian embryo: an experimental approach.* J Embryol Exp Morphol, 1975. **33**(3): p. 607-19.
133. de Bruijn, M.F., et al., *Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo.* EMBO J, 2000. **19**(11): p. 2465-74.
134. Godin, I., F. Dieterlen-Lievre, and A. Cumano, *Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus.* Proc Natl Acad Sci U S A, 1995. **92**(3): p. 773-7.
135. Tavian, M., M.F. Hallais, and B. Peault, *Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo.* Development, 1999. **126**(4): p. 793-803.
136. Kinder, S.J., et al., *The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo.* Development, 1999. **126**(21): p. 4691-701.

137. Lassila, O., et al., *The origin of lymphoid stem cells studied in chick yold sac-embryo chimaeras.* Nature, 1978. **272**(5651): p. 353-4.
138. Kau, C.L. and J.B. Turpen, *Dual contribution of embryonic ventral blood island and dorsal lateral plate mesoderm during ontogeny of hemopoietic cells in Xenopus laevis.* J Immunol, 1983. **131**(5): p. 2262-6.
139. Samokhvalov, I.M., N.I. Samokhvalova, and S. Nishikawa, *Cell tracing shows the contribution of the yolk sac to adult haematopoiesis.* Nature, 2007. **446**(7139): p. 1056-61.
140. Palis, J., et al., *Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse.* Development, 1999. **126**(22): p. 5073-84.
141. Sabin, F.R., *Studies on the origin of blood-vessels and of red corpuscles as seen in the living blastoderm of chicks during the second day of incubation.* Carnegie Contrib Embryol, 1920. **9**: p. 49.
142. Fina, L., et al., *Expression of the CD34 gene in vascular endothelial cells.* Blood, 1990. **75**(12): p. 2417-26.
143. Fehling, H.J., et al., *Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation.* Development, 2003. **130**(17): p. 4217-27.
144. Choi, K., et al., *A common precursor for hematopoietic and endothelial cells.* Development, 1998. **125**(4): p. 725-32.
145. Nishikawa, S.I., et al., *Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages.* Development, 1998. **125**(9): p. 1747-57.
146. GenBank. *Homo sapiens kinase insert domain receptor (a type III receptor tyrosine kinase) (KDR), RefSeqGene on chromosome 4*. 2010 [cited 2010 15/07/10]; Available from: http://www.ncbi.nlm.nih.gov/nuccore/NG_012004.1?&from=5000&to=52336&report=genbank.
147. Zachary, I. and G. Gliki, *Signaling transduction mechanisms mediating biological actions of the vascular endothelial growth factor family.* Cardiovasc Res, 2001. **49**(3): p. 568-81.
148. Giles, F.J., *The vascular endothelial growth factor (VEGF) signaling pathway: a therapeutic target in patients with hematologic malignancies.* Oncologist, 2001. **6 Suppl 5**: p. 32-9.
149. McMahon, G., *VEGF receptor signaling in tumor angiogenesis.* Oncologist, 2000. **5 Suppl 1**: p. 3-10.
150. Biocarta. *Pathways - VEGF, Hypoxia, and Angiogenesis*. 2010 [cited 2010 21/07]; Available from: http://www.biocarta.com/pathfiles/h_vegfPathway.asp.
151. Shalaby, F., et al., *Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice.* Nature, 1995. **376**(6535): p. 62-6.
152. Shalaby, F., et al., *A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis.* Cell, 1997. **89**(6): p. 981-90.
153. Gene, E. *CDH5 cadherin 5, type 2 (vascular endothelium)*. 2010 [cited 2010 15/07/10]; Available from: http://www.ncbi.nlm.nih.gov/gene/1003#geneGeneral gene info.
154. Carmeliet, P., et al., *Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis.* Cell, 1999. **98**(2): p. 147-57.
155. Gene, E. *PTPRC protein tyrosine phosphatase, receptor type, C*. 2010 [cited 2010 15/07/10]; Available from: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=5788.
156. Keller, G., et al., *Hematopoietic commitment during embryonic stem cell differentiation in culture.* Mol Cell Biol, 1993. **13**(1): p. 473-86.
157. Kennedy, M., et al., *Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures.* Blood, 2007. **109**(7): p. 2679-87.
158. Newman, P.J., *The biology of PECAM-1.* J Clin Invest, 1997. **99**(1): p. 3-8.
159. Woodfin, A., M.B. Voisin, and S. Nourshargh, *PECAM-1: a multi-functional molecule in inflammation and vascular biology.* Arterioscler Thromb Vasc Biol, 2007. **27**(12): p. 2514-23.
160. Furuta, C., et al., *Discordant developmental waves of angioblasts and hemangioblasts in the early gastrulating mouse embryo.* Development, 2006. **133**(14): p. 2771-9.
161. Furness, S.G. and K. McNagny, *Beyond mere markers: functions for CD34 family of sialomucins in hematopoiesis.* Immunol Res, 2006. **34**(1): p. 13-32.
162. Fackler, M.J., et al., *Full-length but not truncated CD34 inhibits hematopoietic cell differentiation of M1 cells.* Blood, 1995. **85**(11): p. 3040-7.

163. Ueno, H. and I.L. Weissman, *Clonal analysis of mouse development reveals a polyclonal origin for yolk sac blood islands.* Dev Cell, 2006. **11**(4): p. 519-33.
164. Lugus, J.J., et al., *Both primitive and definitive blood cells are derived from Flk-1+ mesoderm.* Blood, 2009. **113**(3): p. 563-6.
165. Yamashita, J., et al., *Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors.* Nature, 2000. **408**(6808): p. 92-6.
166. Ema, M., et al., *Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse.* Genes Dev, 2003. **17**(3): p. 380-93.
167. Le Clech, M., et al., *PU.1/Spi-1 binds to the human TAL-1 silencer to mediate its activity.* J Mol Biol, 2006. **355**(1): p. 9-19.
168. Lecuyer, E. and T. Hoang, *SCL: from the origin of hematopoiesis to stem cells and leukemia.* Exp Hematol, 2004. **32**(1): p. 11-24.
169. Robb, L., et al., *Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7075-9.
170. Shivdasani, R.A., E.L. Mayer, and S.H. Orkin, *Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL.* Nature, 1995. **373**(6513): p. 432-4.
171. Robertson, S.M., et al., *A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1.* Development, 2000. **127**(11): p. 2447-59.
172. Lacaud, G., et al., *Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro.* Blood, 2002. **100**(2): p. 458-66.
173. Bailey, A.S., et al., *Transplanted adult hematopoietic stems cells differentiate into functional endothelial cells.* Blood, 2004. **103**(1): p. 13-9.
174. Zovein, A.C. and M.L. Iruela-Arispe, *My O'Myeloid, a tale of two lineages.* Proc Natl Acad Sci U S A, 2006. **103**(35): p. 12959-60.
175. Angioworld. *Angiogenesis in cancer*. 2010 [cited 2010 25.07.10]; Available from: http://www.angioworld.com/cancer.htm.
176. Metacore. *Development_VEGF signaling via VEGFR2 - generic cascades*. 2010 [cited 2010 27.07.10]; Available from: http://www.genego.com/map_533.php.
177. Gering, M., et al., *The SCL gene specifies haemangioblast development from early mesoderm.* EMBO J, 1998. **17**(14): p. 4029-45.
178. D'Souza, S.L., A.G. Elefanty, and G. Keller, *SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development.* Blood, 2005. **105**(10): p. 3862-70.
179. Wilson, N.K., et al., *The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development.* Blood, 2009. **113**(22): p. 5456-65.
180. Ribatti, D., *Hemangioblast does exist.* Leuk Res, 2008. **32**(6): p. 850-4.
181. Lu, S.J., et al., *Robust generation of hemangioblastic progenitors from human embryonic stem cells.* Regen Med, 2008. **3**(5): p. 693-704.
182. Basak, G.W., et al., *Human embryonic stem cells hemangioblast express HLA-antigens.* J Transl Med, 2009. **7**: p. 27.
183. Gehling, U.M., *Hemangioblasts and their progeny.* Methods Enzymol, 2006. **419**: p. 179-93.
184. Lacaud, G., G. Keller, and V. Kouskoff, *Tracking mesoderm formation and specification to the hemangioblast in vitro.* Trends Cardiovasc Med, 2004. **14**(8): p. 314-7.
185. Cogle, C.R. and E.W. Scott, *The hemangioblast: cradle to clinic.* Exp Hematol, 2004. **32**(10): p. 885-90.
186. Semenza, G.L., *Vasculogenesis, angiogenesis, and arteriogenesis: mechanisms of blood vessel formation and remodeling.* J Cell Biochem, 2007. **102**(4): p. 840-7.
187. Velazquez, O.C., *Angiogenesis and vasculogenesis: inducing the growth of new blood vessels and wound healing by stimulation of bone marrow-derived progenitor cell mobilization and homing.* J Vasc Surg, 2007. **45 Suppl A**: p. A39-47.
188. Tavian, M., et al., *Aorta-associated CD34+ hematopoietic cells in the early human embryo.* Blood, 1996. **87**(1): p. 67-72.
189. Pardanaud, L., et al., *Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis.* Development, 1996. **122**(5): p. 1363-71.
190. Lancrin, C., et al., *Blood cell generation from the hemangioblast.* J Mol Med. **88**(2): p. 167-72.
191. Jaffredo, T., et al., *Tracing the progeny of the aortic hemangioblast in the avian embryo.* Dev Biol, 2000. **224**(2): p. 204-14.
192. Eilken, H.M., S. Nishikawa, and T. Schroeder, *Continuous single-cell imaging of blood generation from haemogenic endothelium.* Nature, 2009. **457**(7231): p. 896-900.

193. Lancrin, C., et al., *The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage.* Nature, 2009. **457**(7231): p. 892-5.
194. Pardanaud, L. and F. Dieterlen-Lievre, *Manipulation of the angiopoietic/hemangiopoietic commitment in the avian embryo.* Development, 1999. **126**(4): p. 617-27.
195. Belaoussoff, M., S.M. Farrington, and M.H. Baron, *Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo.* Development, 1998. **125**(24): p. 5009-18.
196. Zhang, X.M., M. Ramalho-Santos, and A.P. McMahon, *Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node.* Cell, 2001. **105**(6): p. 781-92.
197. Sheng, G., *Primitive and definitive erythropoiesis in the yolk sac: a birds eye view.* Int J Dev Biol.
198. Barker, J.E., *Development of the mouse hematopoietic system. I. Types of hemoglobin produced in embryonic yolk sac and liver.* Dev Biol, 1968. **18**(1): p. 14-29.
199. Martin, J.S., et al., *Analysis of homozygous TGF beta 1 null mouse embryos demonstrates defects in yolk sac vasculogenesis and hematopoiesis.* Ann N Y Acad Sci, 1995. **752**: p. 300-8.
200. Cheng, X., et al., *Numb mediates the interaction between Wnt and Notch to modulate primitive erythropoietic specification from the hemangioblast.* Development, 2008. **135**(20): p. 3447-58.
201. Brou, C., et al., *A novel proteolytic cleavage involved in Notch signaling: the role of the disintegrin-metalloprotease TACE.* Mol Cell, 2000. **5**(2): p. 207-16.
202. Hadland, B.K., et al., *A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development.* Blood, 2004. **104**(10): p. 3097-105.
203. Nostro, M.C., et al., *Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood.* Cell Stem Cell, 2008. **2**(1): p. 60-71.
204. Rainis, L., et al., *Mutations in exon 2 of GATA1 are early events in megakaryocytic malignancies associated with trisomy 21.* Blood, 2003. **102**(3): p. 981-6.
205. Klimchenko, O., et al., *A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis.* Blood, 2009. **114**(8): p. 1506-17.
206. Eisbacher, M., et al., *Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic expression of megakaryocyte-specific genes through cooperative DNA binding.* Mol Cell Biol, 2003. **23**(10): p. 3427-41.
207. Guo, Y., et al., *c-Myc-mediated control of cell fate in megakaryocyte-erythrocyte progenitors.* Blood, 2009. **114**(10): p. 2097-106.
208. Lai, A.Y. and M. Kondo, *Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors.* J Exp Med, 2006. **203**(8): p. 1867-73.
209. Lux, C.T., et al., *All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac.* Blood, 2008. **111**(7): p. 3435-8.
210. Lacaud, G., et al., *Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells.* Blood, 2004. **103**(3): p. 886-9.
211. Nagai, R., et al., *RUNX1 suppression induces megakaryocytic differentiation of UT-7/GM cells.* Biochem Biophys Res Commun, 2006. **345**(1): p. 78-84.
212. Myoshi, H., K. Shimizu, and T. Kozu, *t(8,21) breakpoints on chromosome 21 in acute myeloid leukaemia are clustered with a limited region of a single gene, AML1.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 10431-10434.
213. Lacaud, G., et al., *Regulation of hemangioblast development.* Ann N Y Acad Sci, 2001. **938**: p. 96-107; discussion 108.
214. Burns, C.E., et al., *Hematopoietic stem cell fate is established by the Notch-Runx pathway.* Genes Dev, 2005. **19**(19): p. 2331-42.
215. Mandal, L., U. Banerjee, and V. Hartenstein, *Evidence for a fruit fly hemangioblast and similarities between lymph-gland hematopoiesis in fruit fly and mammal aorta-gonadal-mesonephros mesoderm.* Nat Genet, 2004. **36**(9): p. 1019-23.
216. Wang, L., et al., *Hematopoietic development from human embryonic stem cell lines.* Exp Hematol, 2005. **33**(9): p. 987-96.
217. Wang, L., et al., *Derivation and characterization of hematopoietic cells from human embryonic stem cells.* Methods Mol Biol, 2006. **331**: p. 179-200.
218. Tucker, K.L., et al., *A transgenic mouse strain expressing four drug-selectable marker genes.* Nucleic Acids Res, 1997. **25**(18): p. 3745-6.

219. Lim, J.W. and A. Bodnar, *Proteome analysis of conditioned medium from mouse embryonic fibroblast feeder layers which support the growth of human embryonic stem cells.* Proteomics, 2002. **2**(9): p. 1187-203.
220. Vallier, L., M. Alexander, and R.A. Pedersen, *Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells.* J Cell Sci, 2005. **118**(Pt 19): p. 4495-509.
221. Eiselleova, L., et al., *Comparative study of mouse and human feeder cells for human embryonic stem cells.* Int J Dev Biol, 2008. **52**(4): p. 353-63.
222. Zhou, D., et al., *Three key variables involved in feeder preparation for the maintenance of human embryonic stem cells.* Cell Biol Int, 2009. **33**(7): p. 796-800.
223. WiCell. *WiCell research Institute*. 2010 [cited 2010 20/08]; Available from: http://www.wicell.org/index.php?option=com_content&task=blogsection&id=11&Itemid=148.
224. Wicell. *Overview of NSCB Characterization Studies*. 2010 [cited 2010 20/08]; Available from: http://www.wicell.org/index.php?option=com_content&task=view&id=231&Itemid=250.
225. Vajta, G., et al., *Open Pulled Straw (OPS) vitrification: a new way to reduce cryoinjuries of bovine ova and embryos.* Mol Reprod Dev, 1998. **51**(1): p. 53-8.
226. Watanabe, K., et al., *A ROCK inhibitor permits survival of dissociated human embryonic stem cells.* Nat Biotechnol, 2007. **25**(6): p. 681-6.
227. Levenstein, M.E., et al., *Basic fibroblast growth factor support of human embryonic stem cell self-renewal.* Stem Cells, 2006. **24**(3): p. 568-74.
228. Keller, G.M., *In vitro differentiation of embryonic stem cells.* Curr Opin Cell Biol, 1995. **7**(6): p. 862-9.
229. Ledran, M.H., et al., *Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches.* Cell Stem Cell, 2008. **3**(1): p. 85-98.
230. Invitrogen. *StemPro 34 Media*. 2010 [cited 2010 22/08/10]; Available from: http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cell-Culture/Stem-Cell-Research/Hematopoietic-Stem-Cells/HSC-Stem-Cell-Culture-Media-Reagents/stempro-34-sfm.html.
231. Invitrogen. *StemPro®-34 Serum-free Medium*. 2010 [cited 2010 15th May]; Available from: http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cell-Culture/Stem-Cell-Research/Hematopoietic-Stem-Cells/HSC-Stem-Cell-Culture-Media-Reagents/stempro-34-sfm.html.
232. Green, F.J., *Sigma-Aldrich Handbook of Stains, Dyes and Indicators*. 1990.
233. Technologies, S.C. *Colony Assays of Hematopoietic Cells Using Methylcellulose Medium*. 2010 [cited 2010 19th April]; Available from: http://www.stemcell.com/technical/28404_methocult%20H.pdf.
234. Invitrogen. *Custom Primers - OligoPerfect™ Designer*. 2009 [cited 2009 25/08/09]; Available from: http://tools.invitrogen.com/content.cfm?pageid=9716.
235. NCBI. *Entrez Nucleotide*. 2009 [cited 2009 25/08/09]; Available from: http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore&cmd=search&term=.
236. Ensembl. *e!Ensembl*. 2009 [cited 2009 26/08/09]; Available from: http://www.ensembl.org/.
237. Biosystems, A.A. *TechNotes 13 (4): Top Ten Pitfalls in Quantitative Real-time PCR Primer/Probe Design and Use*. 2009 [cited 2009 27/08/09]; Available from: http://www.ambion.com/techlib/tn/134/13.html.
238. Wilfinger, W.W., K. Mackey, and P. Chomczynski, *Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity.* Biotechniques, 1997. **22**(3): p. 474-6, 478-81.
239. Ambion. *TURBO DNase*. 2010 [cited 2010 02/04/10]; Available from: http://www.ambion.com/techlib/spec/sp_2238.pdf.
240. Invitrogen. *SYBR® Green I Nucleic Acid Gel Stain*. 2010 [cited 2010 05/04]; Available from: http://probes.invitrogen.com/media/pis/mp07567.pdf.
241. Lieber, R.L., *Statistical significance and statistical power in hypothesis testing.* J Orthop Res, 1990. **8**(2): p. 304-9.
242. Bollerot, K., C. Pouget, and T. Jaffredo, *The embryonic origins of hematopoietic stem cells: a tale of hemangioblast and hemogenic endothelium.* APMIS, 2005. **113**(11-12): p. 790-803.
243. Choi, K., *Hemangioblast development and regulation.* Biochem Cell Biol, 1998. **76**(6): p. 947-56.
244. Jaffredo, T., et al., *From hemangioblast to hematopoietic stem cell: an endothelial connection?* Exp Hematol, 2005. **33**(9): p. 1029-40.
245. Park, C., Y.D. Ma, and K. Choi, *Evidence for the hemangioblast.* Exp Hematol, 2005. **33**(9): p. 965-70.

246. Xiong, J.W., *Molecular and developmental biology of the hemangioblast.* Dev Dyn, 2008. **237**(5): p. 1218-31.
247. Jaffredo, T., et al., *Tracing the hemangioblast during embryogenesis: developmental relationships between endothelial and hematopoietic cells.* Int J Dev Biol, 2005. **49**(2-3): p. 269-77.
248. Kennedy, M., et al., *A common precursor for primitive erythropoiesis and definitive haematopoiesis.* Nature, 1997. **386**(6624): p. 488-93.
249. Yokomizo, T., et al., *Runx1 is involved in primitive erythropoiesis in the mouse.* Blood, 2008. **111**(8): p. 4075-80.
250. Chandler, K.J., R.L. Chandler, and D.P. Mortlock, *Identification of an ancient Bmp4 mesoderm enhancer located 46 kb from the promoter.* Dev Biol, 2009. **327**(2): p. 590-602.
251. Martin, R., et al., *SCL interacts with VEGF to suppress apoptosis at the onset of hematopoiesis.* Development, 2004. **131**(3): p. 693-702.
252. Ramos-Mejia, V., et al., *Nodal/Activin Signaling Predicts Human Pluripotent Stem Cell Lines Prone to Differentiate Toward the Hematopoietic Lineage.* Mol Ther.
253. Ng, E.S., et al., *Directed differentiation of human embryonic stem cells as spin embryoid bodies and a description of the hematopoietic blast colony forming assay.* Curr Protoc Stem Cell Biol, 2008. **Chapter 1**: p. Unit 1D 3.
254. Chaubal, A., et al., *CD34 immunoreactivity in nervous system tumors.* Acta Neuropathol, 1994. **88**(5): p. 454-8.
255. Affymetrix. *GeneChip® One-Cycle cDNA Synthesis Kit*. 2009 [cited 2009 19/08]; Available from: http://www.affymetrix.com/products_services/reagents/specific/cdna1.affx.
256. Goldman, D.C., et al., *BMP4 regulates the hematopoietic stem cell niche.* Blood, 2009. **114**(20): p. 4393-401.
257. Willems, E. and L. Leyns, *Patterning of mouse embryonic stem cell-derived pan-mesoderm by Activin A/Nodal and Bmp4 signaling requires Fibroblast Growth Factor activity.* Differentiation, 2008. **76**(7): p. 745-59.
258. Yamada, G., et al., *Regulated expression of Brachyury(T), Nkx1.1 and Pax genes in embryoid bodies.* Biochem Biophys Res Commun, 1994. **199**(2): p. 552-63.
259. Liang, D., et al., *The role of vascular endothelial growth factor (VEGF) in vasculogenesis, angiogenesis, and hematopoiesis in zebrafish development.* Mech Dev, 2001. **108**(1-2): p. 29-43.
260. Goldman, O., et al., *A boost of BMP4 accelerates the commitment of human embryonic stem cells to the endothelial lineage.* Stem Cells, 2009. **27**(8): p. 1750-9.
261. Kulesh, D.A., et al., *Identification of interferon-modulated proliferation-related cDNA sequences.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8453-7.
262. Lu, S.J., et al., *GeneChip analysis of human embryonic stem cell differentiation into hemangioblasts: an in silico dissection of mixed phenotypes.* Genome Biol, 2007. **8**(11): p. R240.
263. Chen, D., et al., *A microarray analysis of the emergence of embryonic definitive hematopoiesis.* Exp Hematol, 2007. **35**(9): p. 1344-57.
264. Bhattacharya, B., et al., *Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies.* BMC Dev Biol, 2005. **5**: p. 22.
265. Newman, P.J. and D.K. Newman, *Signal transduction pathways mediated by PECAM-1: new roles for an old molecule in platelet and vascular cell biology.* Arterioscler Thromb Vasc Biol, 2003. **23**(6): p. 953-64.
266. Jackson, D.E., *The unfolding tale of PECAM-1.* FEBS Lett, 2003. **540**(1-3): p. 7-14.
267. Case, J., et al., *Human CD34+AC133+VEGFR-2+ cells are not endothelial progenitor cells but distinct, primitive hematopoietic progenitors.* Exp Hematol, 2007. **35**(7): p. 1109-18.
268. Hong, D.L., et al., *Vascular Endothelial Growth Factor and Its Receptor KDR/flk-1 Play Important Roles in Hematopoiesis.* Zhongguo Shi Yan Xue Ye Xue Za Zhi, 2001. **9**(3): p. 268-272.
269. Biosystems, A.A. *Denaturing Agarose Gel Electrophoresis of RNA*. 2009 [cited 2009 19/08]; Available from: http://www.ambion.com/techlib/append/supp/rna_gel.html.
270. Schroeder, A., et al., *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Mol Biol, 2006. **7**: p. 3.
271. Affymetrix. *GeneChip® Two-Cycle cDNA Synthesis Kit*. 2009 [cited 2009 19/08]; Available from: http://www.affymetrix.com/products_services/reagents/specific/cdna2.affx.
272. Affymetrix. *Affymetrix Datasheet: Genechip Human Genome 133 Arrays*. 2009 [cited 2009 4th December]; Available from: http://www.affymetrix.com/support/technical/datasheets/hgu133arrays_datasheet.pdf.

273. Agilent. *GeneSpring GX Software* 2009 [cited 2009 20th November]; Available from: http://www.chem.agilent.com/en-US/products/software/lifesciencesinformatics/genespringgx/Pages/default.aspx.
274. Agilent, *Principle Component Analysis*. 2009.
275. Affymetrix. *Gene Expression Assay and Data Analysis*. 2009 [cited 2009 2nd December]; Available from: http://www.affymetrix.com/support/help/faqs/ge_assays/faq_17.jsp.
276. (NCBI), N.C.f.B.I. *Entrez Gene*. 2010 [cited 2010 10/01/10]; Available from: http://www.ncbi.nlm.nih.gov/gene.
277. Lee, P.D., et al., *Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies.* Genome Res, 2002. **12**(2): p. 292-7.
278. Armstrong, L., et al., *The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis.* Hum Mol Genet, 2006. **15**(11): p. 1894-913.
279. GeneGo. *Metacore*. 2009 [cited 2009 28th November]; Available from: http://www.genego.com/metacore.php.
280. Hasegawa, T., et al., *Gastrointestinal stromal tumor: consistent CD117 immunostaining for diagnosis, and prognostic classification based on tumor size and MIB-1 grade.* Hum Pathol, 2002. **33**(6): p. 669-76.
281. Breccia, M., et al., *Atypical chronic myeloid leukaemia with CD117-positive blast cells treated with imatinib: A report of two cases.* Acta Haematol, 2006. **116**(3): p. 211-2.
282. GeneGo. *Metacore*. 2009 [cited 2009 30th November]; Available from: https://portal.genego.com.
283. EMBL-EBI. *GO @ EBI*. 2010 [cited 2010 21/02/10]; Available from: http://www.ebi.ac.uk/GO/.
284. Yang, L., et al., *Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population.* Nature, 2008. **453**(7194): p. 524-8.
285. DeRisi, J., et al., *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nat Genet, 1996. **14**(4): p. 457-60.
286. Huggins, C.E., et al., *Functional and metabolic remodelling in GLUT4-deficient hearts confers hyper-responsiveness to substrate intervention.* J Mol Cell Cardiol, 2008. **44**(2): p. 270-80.
287. Patterson, T.A., et al., *Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project.* Nat Biotechnol, 2006. **24**(9): p. 1140-50.
288. Raouf, A., et al., *Transcriptome analysis of the normal human mammary cell commitment and differentiation process.* Cell Stem Cell, 2008. **3**(1): p. 109-18.
289. Loughran, S.J., et al., *The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells.* Nat Immunol, 2008. **9**(7): p. 810-9.
290. Christian, S., et al., *Molecular cloning and characterization of EndoGlyx-1, an EMILIN-like multisubunit glycoprotein of vascular endothelium.* J Biol Chem, 2001. **276**(51): p. 48588-95.
291. Claudio, J.O., et al., *HACS1 encodes a novel SH3-SAM adaptor protein differentially expressed in normal and malignant hematopoietic cells.* Oncogene, 2001. **20**(38): p. 5373-7.
292. Roberts, D.M., et al., *A vascular gene trap screen defines RasGRP3 as an angiogenesis-regulated gene required for the endothelial response to phorbol esters.* Mol Cell Biol, 2004. **24**(24): p. 10515-28.
293. Coughlin, J.J., et al., *RasGRP1 and RasGRP3 regulate B cell proliferation by facilitating B cell receptor-Ras signaling.* J Immunol, 2005. **175**(11): p. 7179-84.
294. Wang, Y. and N. Sheibani, *Expression pattern of alternatively spliced PECAM-1 isoforms in hematopoietic cells and platelets.* J Cell Biochem, 2002. **87**(4): p. 424-38.
295. Matsubara, A., et al., *Endomucin, a CD34-like sialomucin, marks hematopoietic stem cells throughout development.* J Exp Med, 2005. **202**(11): p. 1483-92.
296. Verma, A., et al., *Endothelial cell-specific chemotaxis receptor (ecscr) promotes angioblast migration during vasculogenesis and enhances VEGF receptor sensitivity.* Blood.
297. Balemans, W., et al., *Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST).* Hum Mol Genet, 2001. **10**(5): p. 537-43.
298. Tallquist, M.D., W.J. French, and P. Soriano, *Additive effects of PDGF receptor beta signaling pathways in vascular smooth muscle cell development.* PLoS Biol, 2003. **1**(2): p. E52.
299. Gridley, T., *Notch signaling in vascular development and physiology.* Development, 2007. **134**(15): p. 2709-18.
300. Wang, S., et al., *Thrombospondin-1-deficient mice exhibit increased vascular density during retinal vascular development and are less sensitive to hyperoxia-mediated vessel obliteration.* Dev Dyn, 2003. **228**(4): p. 630-42.

301. Koury, M.J., M.C. Bondurant, and J.B. Atkinson, *Erythropoietin control of terminal erythroid differentiation: maintenance of cell viability, production of hemoglobin, and development of the erythrocyte membrane.* Blood Cells, 1987. **13**(1-2): p. 217-26.
302. Kaupke, C.J., S. Kim, and N.D. Vaziri, *Effect of erythrocyte mass on arterial blood pressure in dialysis patients receiving maintenance erythropoietin therapy.* J Am Soc Nephrol, 1994. **4**(11): p. 1874-8.
303. Qiu, C., et al., *Globin switches in yolk sac-like primitive and fetal-like definitive red blood cells produced from human embryonic stem cells.* Blood, 2008. **111**(4): p. 2400-8.
304. Kikushige, Y., et al., *Human Flt3 is expressed at the hematopoietic stem cell and the granulocyte/macrophage progenitor stages to maintain cell survival.* J Immunol, 2008. **180**(11): p. 7358-67.
305. Dicker, F., et al., *Trisomy 13 is strongly associated with AML1/RUNX1 mutations and increased FLT3 expression in acute myeloid leukemia.* Blood, 2007. **110**(4): p. 1308-16.
306. Hieronymus, T., et al., *The transcription factor repertoire of Flt3+ hematopoietic stem cells.* Cells Tissues Organs, 2008. **188**(1-2): p. 103-15.
307. Racioppi, L. and A.R. Means, *Calcium/calmodulin-dependent kinase IV in immune and inflammatory responses: novel routes for an ancient traveller.* Trends Immunol, 2008. **29**(12): p. 600-7.
308. Feske, S., *Calcium signalling in lymphocyte activation and disease.* Nat Rev Immunol, 2007. **7**(9): p. 690-702.
309. Tober, J., et al., *The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis.* Blood, 2007. **109**(4): p. 1433-41.
310. Adelman, D.M., E. Maltepe, and M.C. Simon, *Multilineage embryonic hematopoiesis requires hypoxic ARNT activity.* Genes Dev, 1999. **13**(19): p. 2478-83.
311. Ramirez-Bergeron, D.L., et al., *HIF-dependent hematopoietic factors regulate the development of the embryonic vasculature.* Dev Cell, 2006. **11**(1): p. 81-92.
312. Goodnough, L.T., B. Skikne, and C. Brugnara, *Erythropoietin, iron, and erythropoiesis.* Blood, 2000. **96**(3): p. 823-33.
313. Kesselring, F., K. Spicher, and H. Porzig, *Changes in G protein pattern and in G protein-dependent signaling during erythropoietin- and dimethylsulfoxide-induced differentiation of murine erythroleukemia cells.* Blood, 1994. **84**(12): p. 4088-98.
314. Boudot, C., et al., *Involvement of the Src kinase Lyn in phospholipase C-gamma 2 phosphorylation and phosphatidylinositol 3-kinase activation in Epo signalling.* Biochem Biophys Res Commun, 2003. **300**(2): p. 437-42.
315. Damen, J.E., et al., *The role of erythropoietin receptor tyrosine phosphorylation in erythropoietin-induced proliferation.* Leukemia, 1997. **11 Suppl 3**: p. 423-5.
316. He, T.C., et al., *Erythropoietin-induced recruitment of Shc via a receptor phosphotyrosine-independent, Jak2-associated pathway.* J Biol Chem, 1995. **270**(19): p. 11055-61.
317. Schmidt, U., et al., *Btk is required for an efficient response to erythropoietin and for SCF-controlled protection against TRAIL in erythroid progenitors.* J Exp Med, 2004. **199**(6): p. 785-95.
318. van den Akker, E., et al., *Tyrosine kinase receptor RON functions downstream of the erythropoietin receptor to induce expansion of erythroid progenitors.* Blood, 2004. **103**(12): p. 4457-65.
319. Ghaffari, S., et al., *AKT induces erythroid-cell maturation of JAK2-deficient fetal liver progenitor cells and is required for Epo regulation of erythroid-cell differentiation.* Blood, 2006. **107**(5): p. 1888-91.
320. Hecht, F., et al., *Predominance of hemoglobin Gower 1 in early human embryonic development.* Science, 1966. **152**(718): p. 91-2.
321. Eisenberg, E. and E.Y. Levanon, *Human housekeeping genes are compact.* Trends Genet, 2003. **19**(7): p. 362-5.
322. Gibson, U.E., C.A. Heid, and P.M. Williams, *A novel method for real time quantitative RT-PCR.* Genome Res, 1996. **6**(10): p. 995-1001.
323. Barber, R.D., et al., *GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues.* Physiol Genomics, 2005. **21**(3): p. 389-95.
324. Higa, S., et al., *Gene organization and sequence of the region containing the ribosomal protein genes RPL13A and RPS11 in the human genome and conserved features in the mouse genome.* Gene, 1999. **240**(2): p. 371-7.

325. Battegay, E.J., et al., *PDGF-BB modulates endothelial proliferation and angiogenesis in vitro via PDGF beta-receptors.* J Cell Biol, 1994. **125**(4): p. 917-28.
326. Beitz, J.G., et al., *Human microvascular endothelial cells express receptors for platelet-derived growth factor.* Proc Natl Acad Sci U S A, 1991. **88**(5): p. 2021-5.
327. Jiang, H., et al., *Downregulation of CREB-binding protein expression inhibits thrombin-induced proliferation of endothelial cells: possible relevance to PDGF-B.* Cell Biol Int.
328. Zania, P., et al., *Thrombin mediates mitogenesis and survival of human endothelial cells through distinct mechanisms.* Am J Physiol Cell Physiol, 2008. **294**(5): p. C1215-26.
329. Olivot, J.M., et al., *Thrombomodulin prolongs thrombin-induced extracellular signal-regulated kinase phosphorylation and nuclear retention in endothelial cells.* Circ Res, 2001. **88**(7): p. 681-7.
330. Minami, T., et al., *Thrombin and phenotypic modulation of the endothelium.* Arterioscler Thromb Vasc Biol, 2004. **24**(1): p. 41-53.
331. Janmaat, M.L., et al., *Erythropoietin accelerates smooth muscle cell-rich vascular lesion formation in mice through endothelial cell activation involving enhanced PDGF-BB release.* Blood. **115**(7): p. 1453-60.
332. Ashley, R.A., et al., *Erythropoietin stimulates vasculogenesis in neonatal rat mesenteric microvascular endothelial cells.* Pediatr Res, 2002. **51**(4): p. 472-8.
333. Kyba, M., et al., *Enhanced hematopoietic differentiation of embryonic stem cells conditionally expressing Stat5.* Proc Natl Acad Sci U S A, 2003. **100 Suppl 1**: p. 11904-10.
334. Shelburne, C.P., et al., *Stat5 expression is critical for mast cell development and survival.* Blood, 2003. **102**(4): p. 1290-7.
335. Grebien, F., et al., *Stat5 activation enables erythropoiesis in the absence of EpoR and Jak2.* Blood, 2008. **111**(9): p. 4511-22.
336. Bunting, K.D., *STAT5 signaling in normal and pathologic hematopoiesis.* Front Biosci, 2007. **12**: p. 2807-20.
337. Kerenyi, M.A., et al., *Stat5 regulates cellular iron uptake of erythroid cells via IRP-2 and TfR-1.* Blood, 2008. **112**(9): p. 3878-88.
338. Belcher, J.D., et al., *Heme degradation and vascular injury.* Antioxid Redox Signal. **12**(2): p. 233-48.
339. MedicineNet.com. *Definition of Erythropoietin (EPO)*. 2009 [cited 2009 05/11/09]; Available from: http://www.medterms.com/script/main/art.asp?articlekey=7032.
340. Prado-Lopez, S., et al., *Hypoxia Promotes Efficient Differentiation of Human Embryonic Stem Cells to Functional Endothelium.* Stem Cells.
341. Satoh, K., et al., *Important role of endogenous erythropoietin system in recruitment of endothelial progenitor cells in hypoxia-induced pulmonary hypertension in mice.* Circulation, 2006. **113**(11): p. 1442-50.
342. Marzo, F., et al., *Erythropoietin in heart and vessels: focus on transcription and signalling pathways.* J Thromb Thrombolysis, 2008. **26**(3): p. 183-7.
343. Garcia-Cardena, G., et al., *Dynamic activation of endothelial nitric oxide synthase by Hsp90.* Nature, 1998. **392**(6678): p. 821-4.
344. Adamo, L., et al., *Biomechanical forces promote embryonic haematopoiesis.* Nature, 2009. **459**(7250): p. 1131-5.
345. Yamazaki, J. and K. Kitamura, *Cell-to-cell communication via nitric oxide modulation of oscillatory Cl(-) currents in rat intact cerebral arterioles.* J Physiol, 2001. **536**(Pt 1): p. 67-78.
346. Holan, V., et al., *Nitric oxide as a regulatory and effector molecule in the immune system.* Mol Immunol, 2002. **38**(12-13): p. 989-95.
347. Shiode, N., et al., *Flow-mediated vasodilation of human epicardial coronary arteries: effect of inhibition of nitric oxide synthesis.* J Am Coll Cardiol, 1996. **27**(2): p. 304-10.
348. Krasnov, P., et al., *Neuronal nitric oxide synthase contributes to the regulation of hematopoiesis.* Mol Med, 2008. **14**(3-4): p. 141-9.
349. Balemans, W., et al., *Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease.* J Med Genet, 2002. **39**(2): p. 91-7.
350. van Bezooijen, R.L., et al., *SOST/sclerostin, an osteocyte-derived negative regulator of bone formation.* Cytokine Growth Factor Rev, 2005. **16**(3): p. 319-27.
351. Kamiya, N., et al., *BMP signaling negatively regulates bone mass through sclerostin by inhibiting the canonical Wnt pathway.* Development, 2008. **135**(22): p. 3801-11.
352. Semenov, M., K. Tamai, and X. He, *SOST is a ligand for LRP5/LRP6 and a Wnt signaling inhibitor.* J Biol Chem, 2005. **280**(29): p. 26770-5.

353. Weidauer, S.E., et al., *NMR structure of the Wnt modulator protein Sclerostin.* Biochem Biophys Res Commun, 2009. **380**(1): p. 160-5.
354. Ellies, D.L., et al., *Bone density ligand, Sclerostin, directly interacts with LRP5 but not LRP5G171V to modulate Wnt activity.* J Bone Miner Res, 2006. **21**(11): p. 1738-49.
355. Croce, J.C. and D.R. McClay, *Evolution of the Wnt pathways.* Methods Mol Biol, 2008. **469**: p. 3-18.
356. Schweizer, L. and H. Varmus, *Wnt/Wingless signaling through beta-catenin requires the function of both LRP/Arrow and frizzled classes of receptors.* BMC Cell Biol, 2003. **4**: p. 4.
357. Bennett, C.N., et al., *Regulation of osteoblastogenesis and bone mass by Wnt10b.* Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3324-9.
358. Gering, M., et al., *Lmo2 and Scl/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1.* Development, 2003. **130**(25): p. 6187-99.
359. Visvader, J.E., Y. Fujiwara, and S.H. Orkin, *Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development.* Genes Dev, 1998. **12**(4): p. 473-9.
360. Cohen-Kaminsky, S., et al., *Chromatin immunoselection defines a TAL-1 target gene.* EMBO J, 1998. **17**(17): p. 5151-60.
361. Begley, C.G., et al., *Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript.* Proc Natl Acad Sci U S A, 1989. **86**(6): p. 2031-5.
362. Aplan, P.D., et al., *Involvement of the putative hematopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia.* Blood, 1992. **79**(5): p. 1327-33.
363. Begley, C.G., et al., *The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 10128-32.
364. Hall, M.A., et al., *The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 992-7.
365. Elefanty, A.G., et al., *Characterization of hematopoietic progenitor cells that express the transcription factor SCL, using a lacZ "knock-in" strategy.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11897-902.
366. Krause, M., et al., *CeMyoD accumulation defines the body wall muscle cell fate during C. elegans embryogenesis.* Cell, 1990. **63**(5): p. 907-19.
367. Garrell, J. and S. Campuzano, *The helix-loop-helix domain: a common motif for bristles, muscles and sex.* Bioessays, 1991. **13**(10): p. 493-8.
368. Yin, C., et al., *Hand2 regulates extracellular matrix remodeling essential for gut-looping morphogenesis in zebrafish.* Dev Cell. **18**(6): p. 973-84.
369. Thomas, T., et al., *The bHLH factors, dHAND and eHAND, specify pulmonary and systemic cardiac ventricles independent of left-right sidedness.* Dev Biol, 1998. **196**(2): p. 228-36.
370. Wang, X., et al., *Identification of a novel function of TWIST, a bHLH protein, in the development of acquired taxol resistance in human cancer cells.* Oncogene, 2004. **23**(2): p. 474-82.
371. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins.* Cell, 1989. **56**(5): p. 777-83.
372. EMBL-EBI. *ClustalW2*. 2010 [cited 2010 22/09/10]; Available from: http://www.ebi.ac.uk/Tools/clustalw2/index.html.
373. Portal, P.T.P.M. *PMP | Query Result: SCL (human)*. 2008 [cited 2010 20/09/10]; Available from: http://www.proteinmodelportal.org/query/uniprot/P17542.
374. Portal, P.T.P.M. *PMP | Query Result: SCL (mouse)*. 2010 [cited 2010 18/09/10]; Available from: http://www.proteinmodelportal.org/?pid=modelDetail&pmpuid=1000002850365&range_from=1&range_to=329&ac=P22091&zid=async.
375. Longo, A., G.P. Guanga, and R.B. Rose, *Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition.* Biochemistry, 2008. **47**(1): p. 218-29.
376. Hsu, H.L., et al., *Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins.* Mol Cell Biol, 1994. **14**(2): p. 1256-65.
377. Jin, H., et al., *The 5' zebrafish scl promoter targets transcription to the brain, spinal cord, and hematopoietic and endothelial progenitors.* Dev Dyn, 2006. **235**(1): p. 60-7.
378. Terme, J.M., et al., *TGF-beta induces degradation of TAL1/SCL by the ubiquitin-proteasome pathway through AKT-mediated phosphorylation.* Blood, 2009. **113**(26): p. 6695-8.

379. Liu, F., et al., *Fli1 acts at the top of the transcriptional network driving blood and endothelial development.* Curr Biol, 2008. **18**(16): p. 1234-40.
380. Bockamp, E.O., et al., *Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1.* J Biol Chem, 1998. **273**(44): p. 29032-42.
381. Wadman, I.A., et al., *The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins.* EMBO J, 1997. **16**(11): p. 3145-57.
382. Warren, A.J., et al., *The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development.* Cell, 1994. **78**(1): p. 45-57.
383. Mead, P.E., et al., *Primitive erythropoiesis in the Xenopus embryo: the synergistic role of LMO-2, SCL and GATA-binding proteins.* Development, 2001. **128**(12): p. 2301-8.
384. Liao, E.C., et al., *SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish.* Genes Dev, 1998. **12**(5): p. 621-6.
385. Xia, X., et al., *Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner.* Stem Cells Dev, 2007. **16**(1): p. 167-76.
386. Yu, X., et al., *Lentiviral vectors with two independent internal promoters transfer high-level expression of multiple transgenes to human hematopoietic stem-progenitor cells.* Mol Ther, 2003. **7**(6): p. 827-38.
387. Sakurai, K., et al., *Efficient integration of transgenes into a defined locus in human embryonic stem cells.* Nucleic Acids Res. **38**(7): p. e96.
388. Kurosawa, H., *Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells.* J Biosci Bioeng, 2007. **103**(5): p. 389-98.
389. Kurita, R., et al., *Tal1/Scl gene transduction using a lentiviral vector stimulates highly efficient hematopoietic cell differentiation from common marmoset (Callithrix jacchus) embryonic stem cells.* Stem Cells, 2006. **24**(9): p. 2014-22.
390. Song, S.H., C. Hou, and A. Dean, *A positive role for NLI/Ldb1 in long-range beta-globin locus control region function.* Mol Cell, 2007. **28**(5): p. 810-22.
391. Nemati, H., G. Bahrami, and Z. Rahimi, *Rapid separation of human globin chains in normal and thalassemia patients by RP-HPLC.* Mol Biol Rep.
392. Scotland, H. *Scots close to medicine's Holy Grail ... a true blood substitute*. 2010 [cited 2010 26/09/2010]; Available from: http://www.heraldscotland.com/news/health/scots-close-to-medicine-s-holy-grail-a-true-blood-substitute-1.1013172.
393. Wired. *Scientists Create Blood From Stem Cells*. 2008 [cited 2010 26/09/2010]; Available from: http://www.wired.com/wiredscience/2008/08/universal-blood/.
394. Mead, P.E., et al., *SCL specifies hematopoietic mesoderm in Xenopus embryos.* Development, 1998. **125**(14): p. 2611-20.
395. Elwood, N.J., et al., *Enhanced megakaryocyte and erythroid development from normal human CD34(+) cells: consequence of enforced expression of SCL.* Blood, 1998. **91**(10): p. 3756-65.
396. Tanigawa, T., et al., *The SCL gene product is regulated by and differentially regulates cytokine responses during myeloid leukemic cell differentiation.* Proc Natl Acad Sci U S A, 1993. **90**(16): p. 7864-8.
397. Detmer, K. and A.N. Walker, *Bone morphogenetic proteins act synergistically with haematopoietic cytokines in the differentiation of haematopoietic progenitors.* Cytokine, 2002. **17**(1): p. 36-42.
398. Lu, S.J., et al., *Hemangioblasts from human embryonic stem cells generate multilayered blood vessels with functional smooth muscle cells.* Regen Med, 2009. **4**(1): p. 37-47.
399. Weng, W., E.W. Sukowati, and G. Sheng, *On hemangioblasts in chicken.* PLoS One, 2007. **2**(11): p. e1228.
400. Choi, J.H., et al., *In vitro development of a hemangioblast from a human embryonic stem cell, SNUhES#3.* Life Sci, 2009. **85**(1-2): p. 39-45.
401. Ema, M. and J. Rossant, *Cell fate decisions in early blood vessel formation.* Trends Cardiovasc Med, 2003. **13**(6): p. 254-9.
402. Wang, J., et al., *In vitro hematopoietic differentiation of human embryonic stem cells induced by co-culture with human bone marrow stromal cells and low dose cytokines.* Cell Biol Int, 2005. **29**(8): p. 654-61.
403. Miyagi, T., et al., *Flk1+ cells derived from mouse embryonic stem cells reconstitute hematopoiesis in vivo in SCID mice.* Exp Hematol, 2002. **30**(12): p. 1444-53.