SYSTEMS ENGINEERING OF STEM CELL FATE

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

RAVALI T RAJU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisers: Wei-Shou Hu and Catherine Verfaillie

August 2015

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisers Dr. Wei-Shou Hu and Dr. Catherine Verfaillie for their invaluable guidance, support and encouragement through all the years of my graduate career. It has been an honor and privilege to work under the guidance of two of most renowned and respected researchers in the fields of bioprocess engineering and stem cell biology. They have always promoted a collaborative environment to share ideas and information, which has been instrumental in shaping my research. I will be forever grateful to my advisers.

I would like to extend thanks to members of my thesis committee, Dr. Robert Tranquillo and Dr. Samira Azarin in the Department of Chemical Engineering and Materials Science and Dr. Walter Low in the Department of Neurosurgery for their patience and valuable feedback.

I would like to extend my gratitude to present and former members of Hu and Verfaillie research groups: David Chau, Dong Seong Cho, Haiyun Pei, Hsu-Yuan Fu, Kyoungho Lee, Liang Zhao, Arpan Bandyopadhyay, Kartik Subramanian, Yonsil Park, Jason Owens, Shikha Sharma, Andrew Yongky, Bhanu Chandra Mulukutla, Nitya Jacob, Marlene Castro, Siguang Sui, Huong Le, Anushree Chatterjee, Kathryn Johnson, Nandita Vishwanathan, Karthik Jayapal, Joon Chong Yee, Anne Kantardjieff, Katie Wlaschin, Gargi Seth and Ziomara G. It has been a pleasure to be a part of such a dynamic and diverse group. All of you have greatly enriched my graduate school experience and I have made many great friends along the way. I would also like to thank members of the Stem Cell Institute for their support all these years: Susan Keirstead, Jonathan Slack, James Dutton, Nobuaki Kikyo, Tim O'Brien, Beth Ziemba, Genya Gekker, Lauri Andersen and Jesse Browers.

I am grateful to have had so many good friends in Minnesota who have been pillars of support through this period: Krishna, Ashok, Harika, Prathyusha, Shiva, Akhila, Ganesh, Sowjanya, Manohar, Vivek, Sundeep, Srikar, Vijay, Kalyan, Sujana, Sagar, Vijaya, Shailaja, Praneetha, Divija, Senthil, Sindhuja, Umesh, Chandana, Vinayak, Shradha, Shikha and Pavani. This journey would not have been the same without your friendship.

Finally and most importantly, I would like to thank my wonderful family for their love and undying belief in me. To my husband Nari, thank you for being the foundation of my strength, my critic and my best friend. I cannot imagine having undertaken this journey without you as my companion. To my beautiful daughter Nihira, you have brought indescribable joy into our lives and have inspired me to become a better person. To my brother Rahul, thank you for continuous support. Lastly, and most importantly, to my parents, thank you for letting me pursue my dreams without questioning them whatsoever. If not for your love and all the sacrifices you made, I would not be here today. I owe all my successes and the life that we have built today to the values you instilled in me during my upbringing, your unconditional love and never ending support.

DEDICATION

To my parents Raju and Ramani

ABSTRACT

Recent advances in the derivation of functional cells from pluripotent stem cells have raised hope for cell therapy to treat liver ailments. They have enhanced the prospects of developing reliable *in vitro* models for liver diseases and drug toxicity screening. A differentiation protocol mimicking key signaling cues of embryonic development was developed to direct stem cells (ES) towards the hepatic fate and express key hepatic markers and functions. While these results are encouraging, most directed differentiations from stem cells to the target cell types are hampered by lack of functional maturity, cell heterogeneity and low cell yields limiting their translation to the clinic. These cells are therefore refereed to hepatocyte-like cells (HLCs).

An integrative strategy was employed including both experimental techniques as well as a systems-based analysis towards enhancing the product quality and yields of HLCs. Functional maturity was enhanced by initiating three dimensional spheroid formation upon differentiation. Enrichment of hepatic cells using selective medium conditions was performed to obtain higher fraction of cells with the desired properties. Cell expansion was incorporated during differentiation to improve cell yields. Several additional strategies have been used to increase hepatocyte maturity in literature including co-culture and transfection with transcription factors. These methods including ours have shown improvement, however a universal gap to maturation is still present when compared to primary hepatocytes.

Comparison of transcriptome data of differentiation to embryonic liver development can elucidate the genetic roadblocks preventing ES cells from reaching the functional maturity of their tissue counterparts. Transcriptome data was compiled from various depositories for mouse fetal liver development (from E8.5 to post-natal). Transcriptome data was obtained during the time course of our human hepatic differentiation protocol and was augmented with human *in vitro* hepatic differentiation data in the public depository. Interestingly, majority of the HLCs are similar irrespective of the cell source and protocol. The entire cohort of HLCs clustered separately from the primary hepatocytes and adult liver indicating an inherent roadblock to maturation. The transcriptome data of human ES hepatic differentiations was then integrated with mouse liver development using a unique approach. This allowed us to identify the corresponding development stage at which the *in vitro* stem cell differentiation is blocked. The analysis uncovered a pivotal gene set with contrasting profiles in ES differentiation and mouse liver development that merit combinatorial genetic intervention to enhance maturation of ES derived hepatocytes. Thus, one can envision the availability of stem cell based liver therapies in the not so distant future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Introduction

The liver is the largest internal organ in an adult organism. It performs many important functions that sustain the organism's vitality. Liver metabolizes nutrients from ingested food and regulates glucose levels by converting excess glucose to glycogen for storage and releasing it when the blood glucose level is low. It detoxifies xenobiotics and harmful metabolites, and synthesizes many proteins in the blood [7]. Most of these functions are provided by the parenchymal cell type, hepatocytes, that comprise approximately 70% of the adult liver mass.

Liver failure may arise from many causes, including cirrhosis, viral infections and drug overdoses. Typically, the liver has a tremendous regenerative capacity to repair itself. After partial hepatectomy to remove two-thirds of the liver surgically, it is capable of regaining its original mass over time. However, liver disorders can compromise its inherent regenerative capacity and result in complete liver failure leading to death. Although treatment of the symptoms can alleviate the severity of liver failure, organ transplantation is the only curative treatment. However, a severe shortage of donors has limited the access of liver transplants for many patients. As of 2012, there are approximately 17,000 people on the waitlist for liver transplantation in the United States alone, while only half the number of transplantations were performed annually because of the shortage of donor organs (United Network for Organ Sharing: http://optn.transplant.hrsa.gov).

Extracorporeal liver devices have been explored as a treatment to sustain patients until successful liver regeneration, or until a donor organ becomes available. These extracorporeal devices comprise of hepatocytes from a variety of cell source (porcine, human, etc.) as well as mechanical components to provide temporary assistance [8]. The mechanical components of the device employ filtration, adsorption or dialysis to remove small molecular weight toxic metabolites from the patient's blood, while hepatic cells provide the bio-transformative and biosynthetic functions [8,9]. Other approaches of liver failure treatments include transplantation of dissociated hepatocytes from organs and

implantation of tissue engineered liver analogues to augment liver's regenerative capacity for liver recovery [10-12].

For applications involving liver cells such as extracorporeal devices, cell transplantation and tissue engineering, primary human hepatocytes have been the preferred cell source because of its low risk of immunogenicity. The use of isolated liver cells can potentially expand the pool of donor organs, as even organs unsuitable for transplantation may be suitable for use in hepatocytes transplant. However, difficulties in expanding and maintaining primary hepatocytes in culture remain a major hurdle in this field. Even with expanded pools of donor organs as the source of hepatocytes, the need still exceeds the availability of hepatocytes. Furthermore, functional capabilities decrease rapidly during *in vitro* culture [10]. In addition to maintaining our desired cell type, we must also address the need for large quantities of primary cells are needed for the treatment of even a single patient.

Hepatocytes isolated from other species, primarily porcine, may provide an alternative source, however, these cells also suffer from rapid decrease of functional activities when cultured *in vitro* similar to primary human hepatocytes. Moreover, the differences in their drug metabolism and other hepatic functions with human hepatocytes, along with potential immunogenic concerns, render these xenogeneic hepatocytes less than desirable compared to human sources [9].

For future medical applications of liver cells, including cell therapy and extracorporeal liver assist devices, *in vitro* cultivation is most likely to be employed to expand the supply of human cells. These expanded cell population can then be guided to differentiate to the desired cell type for specific applications. In the past few years, stem cell research has made significant advances; stem cells and progenitors cells can now be isolated from various sources, and expanded and differentiated towards the liver lineage. This has brightened the prospect of generating large numbers of functional hepatocytes for applications in hepatic cell transplantation, extracorporeal liver-assist devices and liver tissue engineering.

1.2 **Research Objectives**

The aim of this study is to develop a platform for the robust generation of stem cell derived hepatocytes for clinical applications. The key objectives towards this goal are:

Objective 1: To improve the process technology for stem cell differentiations to hepatocytes, focusing on increasing cell yields and homogeneity to enable clinical transition

Objective 2: To perform systems-based informatics analysis of transcriptome data, comparing stem cell differentiations across the state of the art with embryonic liver development to elucidate universal roadblocks in hepatic differentiation.

1.3 **Scope of Thesis**

A detailed literature review is presented in Chapter 2 adapted from published review articles [5,6,13Raju, 2013 #21]

The first part consists of an introduction to stem cells and a description of their clinical applications. A detailed description of *in vivo* embryonic liver development is provided followed by different strategies used in generating hepatocyte-like cells. The last part focuses on how the knowledge of conventional bioprocess development for protein production can be adapted for translation of stem cell research into a robust bioprocess. This knowhow will be essential to develop strategies for stem cell processes to satisfy clinical demands in the near future.

Chapter 3 contains information about the materials and methods used in this study. Chapter 4 describes how product yields can be improved by utilizing endodermal cell expansion, where the product refers to stem cell derived hepatocytes. Chapter 5 focuses on improving product quality through increased cell homogeneity using surface marker and selective medium-based techniques.

Chapter 6 and 7 represent a comprehensive systems analysis of *in vitro* differentiation and *in vivo* liver developments. Details on sample collection, data integration across platforms and species, analysis for identification of genes that are responsible for the universal gap to maturation and their regulators are included. Finally, in Chapter 8, a summary of findings and their relevance is presented, and potential topics for future investigation are discussed.

2. BACKGROUND

2.1 Introduction to Stem Cells

Stem cells have the ability to propagate in a self-renewing manner, while possessing the potential to differentiate into more specialized cell types in response to molecular cues [14,15]. These special intrinsic characteristics have engendered much interest in using them not only for clinical applications such as regenerative therapy but also for facilitating research by providing better *in vitro* models of diseases for developing novel small molecules of biologics drugs, and drug toxicity testing.

The stem cell nomenclature is often derived from the developmental stage at which they are sampled or from the tissue from which they are isolated. Embryonic stem cells (ESCs) are derived from a specific stage of the developing mammalian embryo [14]. These cells have the potential to differentiate into all three germ layers. Adult stem cells, on the other hand, are derived from post-natal organisms, and are often sub-classified on the basis of the tissue of isolation. These cells are believed to be responsible for tissue homeostasis or repair/regeneration following an injury, and have been found in many tissues such as bone marrow [16], digestive tract [17], heart [18] and kidney [19] amongst others. Hematopoietic stem cells (HSCs) continue to be the best characterized and most clinically relevant adult stem cells to this date [20]. These latter cells are able to differentiate into a variety of specialized hematopoietic cells such as lymphocytes, nature killer cells, and megakaryocytes [21]. Notably, hematopoietic stem cells do not represent the only adult stem cell type that has been isolated from the bone marrow. Mesenchymal stem cells (MSCs), endothelial progenitor cells and other types of progenitor cells also have been isolated. Moreover, multiple neural precursor cells have been isolated from the

adult nervous systems, such as neural stem cells, neuronal progenitor and glial progenitor cells [22].

Stem cells originated from different sources have vastly different properties, the most fundamental of which being the ability to differentiate into various types of specialized cells. Particularly, embryonic stem cells, given their ability to differentiate into cell types characteristic of all three germ layers, namely, mesoderm, endoderm and endoderm, are pluripotent [14],Martin [15] as seen in Figure 2-1.

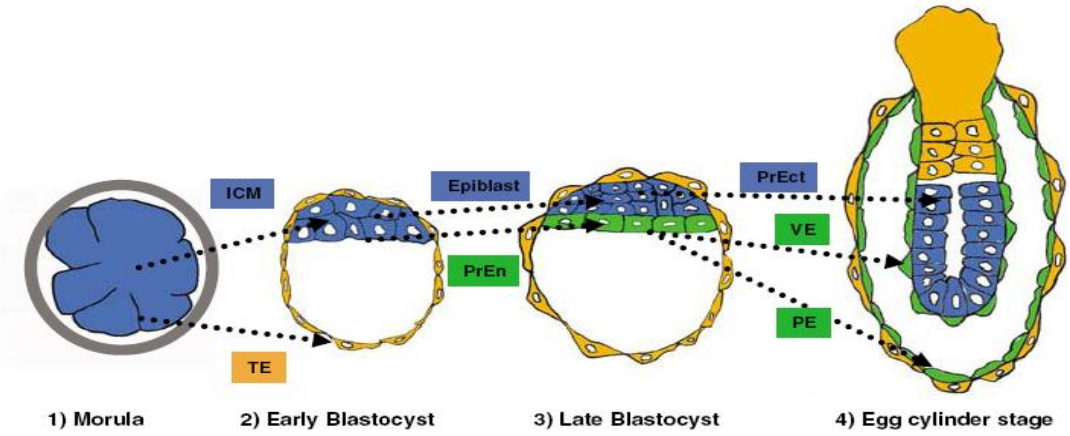


Figure 2-1: Developmental stages in the mouse embryo. The development of the different layers from the inner cell mass (ICM) of the early blastocyst and the subsequent formation of the pluripotent primitive ectoderm (PrEct) is seen The PrEct contains pluripotent cells that differentiate to all three germ layers [23].

2.2 Unique properties of stem cells

2.2.1 Self-renewal

Stem cells are not prone to cellular aging like somatic cells after a certain number of cell divisions (Hayflick limit). This phenomenon is due to the presence of the enzyme telomerase, which prevents the shortening of the telomeres with cell division. The self-renewal potential of embryonic stem cells derived from the blastocyst is different from that of adult stem cells derived from adult tissues. Understanding the reasons for this behavior of the respective stem cells and the factors regulating this process is necessary when expanding large numbers of undifferentiated cells for long term clinical applications [23]

2.2.2 Potency

Potency is defined by the ability of the stem cells to differentiate towards specific cell lineages. A pluripotent cell can generate all the three germ layers: the ectoderm, mesoderm and endoderm while a multipotent cell can differentiate to more restricted cell lineages. Similarly, a unipotent cell, in spite of its self-renewal capacity, can differentiate into only one specific cell type. A typical example of pluripotent cells are embryonic stem cells. Adult stem cells are defined as multipotent as they can differentiate to organ specific cell types like neural stem cells or hematopoietic cells. Unipotent cells reside in adult tissues, undergo self-renewal and differentiate into a specific cell type during injury or wear and tear. Understanding the control mechanisms that maintain these different types of stem cells in a proliferative yet undifferentiated state are critical especially when the large scale expansion of these cells is desired for clinical trials.

Stem cells can remain uncommitted to any lineage for long periods of time. Several transcription factors play an important role in maintaining the undifferentiated state of a stem cell along with chromatin-modifying enzymes, regulatory RNA molecules, and signal-transduction pathways [3]. The most common TF's implicated in this role are the POU-domain transcription factor Oct4, the homeodomain transcription factor Nanog and the high mobility group protein Sox2 [23].These TF's enhance their own expression by binding to their promoters via auto-regulatory loops. They maintain pluripotency by activating related genes involved as well as silencing the genes that promote differentiation. Gene silencing occurs with the recruitment of Polycomb group proteins which induce epigenetic modifications such as histone methylation that condense the chromatin structure making it inaccessible for modifications [3] as seen in

Stem cells can proliferate for extended periods of time but when provided with the right cues they can differentiate into specific cell lineages. The Polycomb machinery is accordingly modified and cell lineage specific genes are activated while the remaining genes are silenced. It is still unclear how these Polycomb group proteins are recruited and how they regulate the cells transformation from a pluripotent state to a differentiated state.

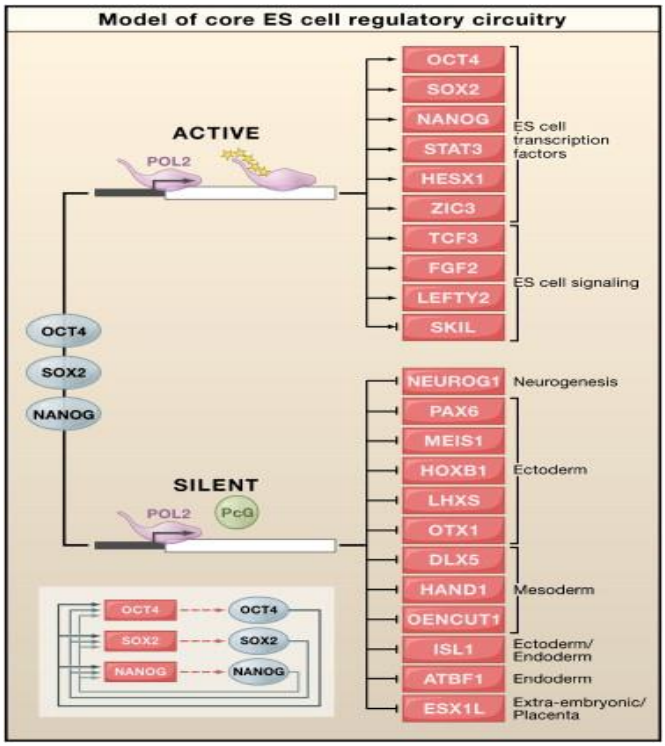


Figure 2-2: Circuit for maintaining pluripotency of ES cells [3]

2.3 Stem Cell Applications

2.3.1 Stem Cells in Regenerative Therapy

As briefly discussed above, the properties of human embryonic stem cells (ESCs) have allowed their use in various clinical applications such as regenerative therapy and transplantation [14]. Their potential is being evaluated through clinical trials to treat diseases in which the function of impaired tissues or organs can be restored by the replenishment of desired cell types. The first clinical trial involving human embryonic stem cells sanctioned by the US FDA in 2009 used a 42-day long differentiation protocol to obtain oligodendrocyte progenitor cells (OPCs) [24]. These human ESC-derived OPCs, which were developed by Geron Corporation for the treatment of spinal cord injury, were found to be 95% positive for oligodendroglial markers such as RIP and GalC [24,25]. In another clinical trial, human ESC-derived retinal pigment epithelium (RPE) cells were evaluated for the treatment of Stargardts's macular dystrophy (SMD) and Age-

related Macular Degeneration (AMD) based on the long term rescue observed in preclinical studies [26-29]

Embryonic stem cells give rise to teratomas consisting of cell types of all three germ layers in immunocompromised mice [30]. The potential of teratoma formation poses risk when differentiated cells derived from hESCs are used clinically. Therefore, the presence of even a small quantity of undifferentiated cells in an hESC transplant is a cause for great concern as exemplified by Geron's Phase I clinical trial application to the FDA in 2010 (http://cell-therapies.geron.com/grnopc1). A very robust differentiation protocol with a strong quality control and quality assurance procedures must be present to assure the absence of hESCs in the final therapeutic cell dose. Notably, since the final product consists of live cells and thus cannot be sterilized, ideally all manufacturing operations are conducted in a closed system to minimize, as much as possible, the risk of contaminations with infectious agents with all primary raw materials also being subjected to extensive tests to ensure the absence of infectious agents. In contrast, the use of mesenchymal stem cells (MSCs) generates a lower safety concern on the tumorigenic potential given their much more limited differentiation properties and the years of practice of bone marrow transplantation. As a result, the therapeutic potential of MSCs is currently explored through various clinical trial, comprising Phase I, Phase II, and Phase III trials, by several companies including Mesoblast, Angioblast, TCA Cellular Therapy or Osiris (http://www.mesoblast.com/clinical-trials/overview), (http://www.tcacellulartherapy.com/fda_clinical_trials.html), (http://www.osiris.com/clinical.php). Notably, the differentiation potential of MSCs is not limited to the mesodermal lineage since these cells can also differentiate into neuroectodermic and endodermic lineages [31-33]. Importantly, MSCs can be readily expanded *in vitro*, thus facilitating their industrial scale manufacturing [34]. Given their immune privilege status, MSCs are used allogeneically in clinical trials for a large variety of indications comprising myocardial and limb ischemia [35], myocardial infarction [36], spinal cord injury [37], Crohn's disease [38], and bone marrow regeneration [39],[40].

2.3.2 Liver support systems

An extracorporeal liver support device may be solely mechanical relying on filtration or adsorption mechanisms to remove toxins in the blood, or a cell-based device designed to perform bio-transformative and homeostasis sustaining functions [9]. A cell based extracorporeal liver support, or bio-artificial liver (BAL) aims to not only remove toxic substances, but also carry out bio-transformative and homeostasis maintaining functions. Depending on the membrane employed and the flux across the membrane, the device may also provide different levels of other excretory and synthetic liver functions, such as urea and albumin synthesis. Some BALs are used in conjunction with an adsorption column (charcoal or ion exchanger). All employ an oxygenator to increase the oxygen content of the blood stream before it enters the cell chamber or bioreactor. As the circulating blood is the only source of oxygen for cells in the bioreactor, such a practice is essential for sustaining cell viability and function. All bioartificial liver devices invariably employ a membrane separating patient's blood or plasma from cells in the device. Such membrane separation of blood stream and cell bioreactor minimizes the cross-over of patient's complement into the cell chamber and the introduction of excreted molecules or lysis products from the cells to the patient. All BAL devices employ hollow fiber bioreactors for use as the cell chamber. [41]

The main challenges BAL devices face include the source and quantity of cells, and means to sustain their viability and functions. An adult liver consists of about 10^{11} hepatocytes, or about 100 g gross wet weight of cells. Most BAL designs thus aim to supply about 10^{10} cells for each treatment of a few hours to over ten hours. With the scarcity of human liver for hepatocyte isolation, researchers have resorted to continuous cell lines derived from human hepatomas such as HepG2-C3A or porcine primary hepatocytes as the source. The devices that underwent large-scale clinical trials include the HepatAssist [42] which used cryopreserved porcine hepatocytes as a cell source and the Extracorporeal Liver Assist Device (ELAD)[43] that employs the HepG2-C3A cell line. The advances over the last few decades in bioartificial liver support along with detailed review of the above devices has been presented in [8,9,44]

Recently, the liver of immunocompromised Frg/Rag2 mutant mice with fumarylacetoacetate hydrolase deficiency (Fah), causing liver death upon withdrawal of

2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC), have been used to achieve a 150-fold proliferation of human hepatocytes. Both cryopreserved and primary sources of hepatocytes were able to engraft and exhibit Fah enzyme activity [45] Similar efforts are being made in generating mutant Fah-deficient pigs for large scale human hepatocyte proliferation which may prove to be a suitable for use in drug testing and in BAL devices [46].

The cells employed in BAL devices all lack the capability of further proliferation in culture and require surface attachment for survival and function. To sustain their viability during BAL applications porcine hepatocytes have been cultivated on microcarriers (Cytodex 3), which are microbeads of about 200 μm diameter with external surface suitable for cell attachment. Collagen gel entrapment of porcine hepatocytes has also been used to provide cell anchorage [47]. Another approach is to allow hepatocytes to self agglomerate and form spheroids by using either a stirred bioreactor or a rocking pan [48]. These spheroids develop organized and polarized structures resembling hepatocytes in liver tissue, and have a prolonged viability and functional activities [49].

A design challenge in BAL is the very high loading cell density in the bioreactor. The bioreactor volume is constrained by the total volume in the blood circulating loop external to the patient. The maximum size of the cell chamber is limited; to maximize the biological activities of the BAL, hepatocyte loading in the cell chamber should also be maximized while not causing oxygen diffusion limitations and decreased in cell viability. The use of microcarriers or alginate beads in the cell chamber may result in the loss of cell loading volume. In the past two decades clinical trials have been performed with a number of BAL systems. Improvements in the patient status have been reported; however, overall clinical betterment has not been fully demonstrated.

BAL technologies are critically dependent on reliable supply of hepatocytes with consistent quality and in sufficient quantities. The hepatoma lines such as HepG2-C3A exhibit altered bio-transformative and metabolic profiles, such as cytochrome P450 and ureagenesis activities [9]. Even primary hepatocytes have substantially lower transcript levels of key phase I and phase II biotransformation enzymes when compared to adult

liver. The lower activities, coupled with the limited amount of cells that can be loaded for patient treatment may have contributed to the elusiveness of better clinical outcomes.

2.3.3 Disease model and drug metabolism testing

Stem cell derived hepatocytes have great potential as *in vitro* models for studying liver disease mechanisms and drug metabolism. The primary tools to date for such studies are animals and primary human hepatocytes; the former are not ideal model for human disease while the latter is constrained by cell availability and representation of genetic diversity. As experimentation on animal models faces increasing ethical challenges, hepatocytes generated from cultured stem cells are becoming even more attractive. Furthermore, with iPSC and iHep, the investigation can be performed using derived hepatocytes of different genetic backgrounds.

A notable example of a cell-based liver model is the replication of hepatitis C using hepatocyte-like cells derived from stem cells [50-53]. The model should facilitate the screening of inhibitors of HepC virus replication pathways. Given the influence of host genotypes on hepatitis C virus entry, assembly and replication, cell based models representing diverse genetic background will be a welcoming development. Another example is the use of hepatocyte-like cells obtained from iPSCs derived from patients with familial hypercholesterolemia. These cells with a mutation in the low density lipoprotein receptor serve as a model of the deficiency in LDL-cholesterol uptake [54].

Such *in vitro* applications as a liver analogue for disease studies and drug screenings require the reproduction of biological functions or even the structure. Like other *in vitro* applications the maturity of the derived hepatocytes will still need to be enhanced substantially.

2.4 Embryonic Liver Development- the guide for *in vitro* culture processes

In this section, we will describe the development of mouse liver, as an example of mammalian development, being cognizant that the development in mouse and man differs in certain aspects.

In early embryo development, the blastocyst consists of an inner cell mass and an outer layer of trophoblast cells. As the primitive blastocysts become polarized and exposed to

a number of signaling pathway cues, they will give rise to the inner cell mass [55,56]. During this developmental stage, embryonic stem cells can be isolated from the inner cell mass which can give rise to all three germ layers. The inner cell mass will further differentiate to two specialized cell type, hypoblast and epiblast cells [57]. Hypoblast cells will give rise to extraembryonic tissues, while epiblast cells will make up all the tissues in an adult by first differentiating to ectoderm and mesoendoderm [58]. During this time, the formation of the primitive streak will set in place the bilateral symmetry and anterior-posterior axis indicating the start of gastrulation [59]. This event marks the beginning of the delineation between the three germ layers, ectoderm, endoderm, and mesoderm, leading up to organogenesis shown in Figure 2-3

The differentiation of the intermediate stage, mesoendoderm, is driven primarily by *Nodal* signaling, a member of the transforming growth factor (TGF- β) family, as demonstrated in an explant model [60-62]. The *Nodal* protein acts as a morphogen; high levels promote endoderm formation and low levels promote mesoderm formation. In the developing mouse embryo, NODAL is produced at the anterior region of the primitive streak where it can exert its effect through a number of downstream transcription factors, including FOXA2, SOX17, GATA4-6, MIXL1 and EOMESODERMIN [63]. The expression level of the genes regulated by these transcription factors delineates the difference between endoderm and mesoderm. The importance of *Nodal* signaling for endoderm commitment was demonstrated in multiple transplantation studies that showed ectopic regions expressing *Nodal* signaling can induce cells to express endoderm markers and differentiate further into endodermal derivatives [64,65].

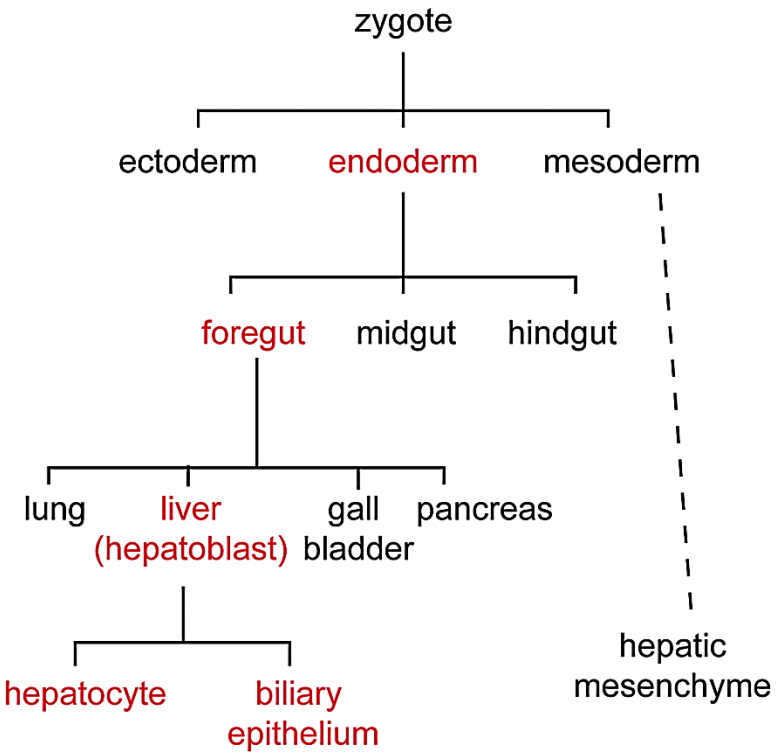


Figure 2-3: Delineation of Liver Development [4]

The epithelial layer of endoderm, in close contact with a thick layer of mesenchymal cells called the septum transversum, will give rise to the digestive and respiratory organs, including the liver. The processes by which cells undergo a massive transformation into a multi-layered group of cells from the blastula are regulated by several signaling pathways. The most widely studied pathway, FGF signaling, causes cells to undergo an epithelial to mesenchymal (EMT) transition by decreasing the amount of cell-cell adhesion [66,67]. The decrease allows cells to expand and spread out to form new layers. The initial stage of gastrulation involves the invagination of the epithelium which results in the cell movement to subdivide the gut tube into foregut, midgut and hind gut regions. In the ventral region of foregut, the cardiogenic mesenchymal cells adjacent to the endoderm secrete several fibroblast growth factors (FGFs), in conjunction with bone morphogenic proteins (BMP)(also members of TGF-superfamily) produced by septum transversum mesenchyme (STM) cells, to induce hepatic specification [68-72]. Fast

proliferating hepatoblasts then emerge and commit this segment of endoderm to develop into the liver bud.

Hepatic endoderm cells or hepatoblasts are bi-potential and can undergo differentiation to hepatocytes or biliary epithelial cells that line the lumen of the intrahepatic bile ducts. Hepatoblasts are capable of proliferating extensively and invading the surrounding septum transversum. Endothelial cells then interact with these hepatoblasts by providing specific growth factors needed for hepatoblasts maintenance and proliferation [73]. Between E10-15 in mice, the formation of the liver bud undergoes rapid growth and vascularization. By day E14-15, the liver bud is a highly vascularized tissue. During this stage of liver development, the STM continues to provide BMP, and additionally expresses hepatocyte growth factor (HGF), while the hepatoblasts express c-Met, the HGF receptor [74]. HGF acts as a suppressor for the apoptosis of hepatoblasts by promoting hepatoblasts proliferation through Wnt3a [75]. It has been speculated that both HGF and BMP provide growth signals, perhaps through parallel pathways. In addition to HGF, Oncostatin is released by the hematopoietic cells which promotes hepatocyte differentiation and maturation by the JAK/Stat3 signaling pathway through activation of the gp130 receptor [76]. The importance of this finding will be illustrated in sections below when we discuss the selective growth factors for *in vitro* differentiation.

The functional cells slowly undergo maturation and the biliary network is formed as the liver attains the appropriate tissue architecture and functional capability [4,7,77]. An overview of liver development is shown in Figure 2-4.The detailed mechanisms behind liver bud development and subsequent maturation of the hepatocytes have been investigated further using microarray analysis on samples taken through various timepoints of mouse embryonic development [78].

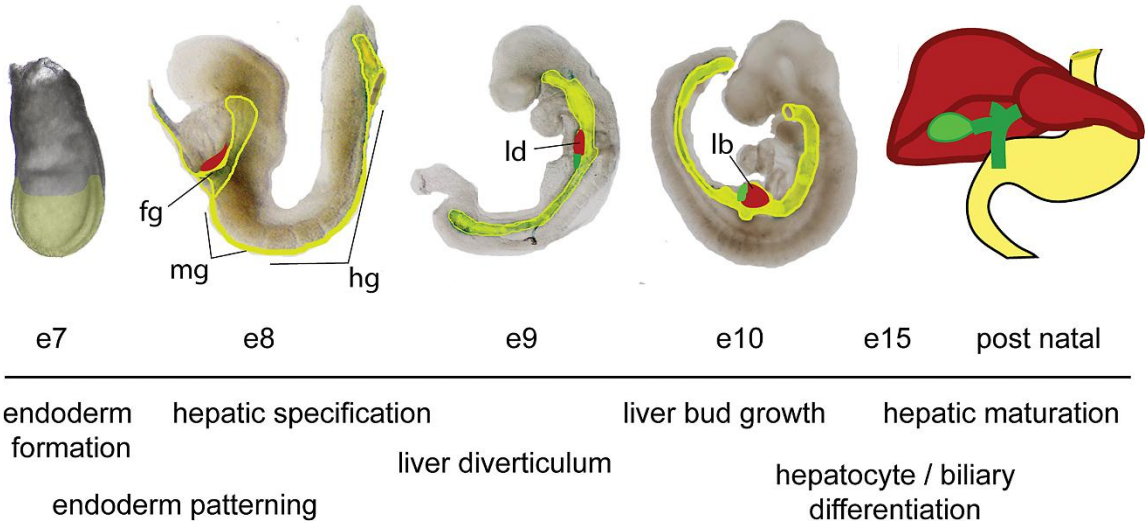


Figure 2-4: Overview of Mouse Liver Development[1]

2.5 Different Types of Stem Cells

2.5.1 Pluripotent Stem Cells

ESCs were first derived by Evans and Kaufman from mouse embryos by culturing mouse blastocysts on layer of STO fibroblasts [30]. The 'pluripotential' cells, derived from the outgrowth of the inner cell mass, not only differentiated *in vitro* into cell types representing multiple tissues, but also formed teratocarcinomas upon subcutaneous injection into syngenic male mice [30]. Human embryonic stem cells were derived several years later from human blastocysts by Thomson et al [14]. The seminal observation of this latter work was that these human cells are similar to their mouse counterparts in many ways: (1) as high nucleus to cytoplasm ratio, (2) expression of surface antigen SSEA-1, and (3) differentiation into the three germ lineages *in vivo* and *in vitro*.

Adult stem cells are generally more restricted in their potency [16,18,79] in line with their suggested role of maintaining homeostasis in the native tissue or organ and initiating proliferation or differentiation only when repairing the tissue or otherwise necessary [80]. It has recently been shown that the potency of stem cells as well as somatic cells is subject to change through a process called 'cellular reprogramming' [81]. Expression of a

combination of exogenous key transcription factors, along with modification of culture conditions, enables the conversion of somatic cells to pluripotent stem cells. Shinya Yamanaka et al. have shown that adult somatic cells from mice can be reprogrammed to a pluripotent, stem cell-like fate by exogenous induction of four transcription factors: Oct4, Sox2, Klf4 and c-Myc [81]. In these experiments, 24 candidate genes that are expressed at high levels in mouse ESCs were screened, leading to the discovery that the expression of these four genes, often referred as OSKM or Yamanaka factors, is sufficient to kick-start a cellular program that remodels the epigenome of the somatic cells such that they become what are now known as induced pluripotent stem cells or iPSCs [81]. These cells are morphologically similar to ESCs, display highly similar gene expression profile, and they differentiate *in vivo*, as well as *in vitro*, into the three germ layers [81-83]. The derivation of iPSCs has opened the possibility of deriving autologous pluripotent stem cells from adult somatic cells. Such autologous, pluripotent cells can in turn be differentiated into specific cell types for transplantation, possibly with minimal risk of immune response. To demonstrate this therapeutic concept, Hanna et al. have derived iPSCs from the fibroblasts of humanized sickle cell anemia model of mouse, corrected their genetic defect by gene targeting, and rescued the mice by using iPSC-derived hematopoietic progenitors [84].

With the prospect of clinical application of stem cells, there is thus a heightened interest in developing technology that would permit their large scale bioprocessing at the lowest cost possible and with the highest quality possible. A number of large-scale efforts are underway for bringing about this bench to the bedside transition.

2.5.2 Fetal Liver Stem Cells

Progenitor cells derived from the fetal liver may be a promising source of hepatocytes for liver cell based therapy. In embryonic liver development, hepatoblasts, a common liver lineage committed cell type, can give rise to both hepatocytes and cholangiocytes. These hepatoblasts exhibit a very much larger proliferative potential compared to primary hepatocytes because of its less differentiated state. Furthermore, they are less prone to rapid de-differentiation in culture than primary hepatocytes and are more suitable for transplantation [85]. Thus there has been a long sustained interest in their isolation and *in*

vitro cultivation. However, because these cells are isolated from fetal livers, their study has mostly been performed in animal models.

Hepatic progenitor cells have been isolated from fetal rodent livers around E11 to 14.5 by flow cytometry or magnetic activated sorting of dissociated liver cells based on the expression of one or a combination of surface markers. Early cell sorting studies used the absence of CD45 and TER119 along with low c-kit expression to sort for these types of early progenitor cells [86,87]. Later, Delta-like protein 1 (DLK1), which begins its expression in the mouse liver bud around E10.5, was used as a surface marker to enrich for hepatoblasts from rat [88] and mouse fetal livers [89,90]. E-cadherin, a cell adhesion protein [91] and epithelial cell adhesion molecule (EPCAM), a cell surface glycoprotein [92] have also been used to isolate hepatoblasts. These fetal liver derived progenitor cells also express cytokeratin 19 (CK19) and albumin. In culture, these cells also possess the capacity of proliferation with the supplementation of growth factors such as HGF and EGF, and are capable of differentiating to both hepatic lineage and cholangenic lineage in culture [91].

These progenitor cells have also been isolated from fetal human liver. The Fausto laboratory isolated clonal progenitor cells from clusters of small cells arising from long term primary cultures established from fetal livers between the gestational stage of 74 and 108 days [93]. These cells exhibited many markers reported in rodent hepatoblasts, including CK19, EPCAM, C-kit. However, maintained on feeder layers of irradiated NIH 3T3 cells, they appear to be pre-hepatoblasts as they express neither albumin nor α-fetal protein. They also lacked the transcripts of a number of liver transcription factors (HNF1α, HNF3α and HNF4α) and are capable of differentiating to not only hepatic lineage but also mesenchymal lineage. Thus, they claimed that the characteristics of these cells made them possibly closer to the more primitive mesendoderm stage of development.

Using EPCAM and ICAM as markers, hepatic progenitors or hepatic stem cells were isolated from both human fetal and postnatal (neonatal, pediatric and adult) liver [94,95]. In the absence of a supporting feeder layer, $EPCAM^{+}NCAM^{+}AFP^{-}ICAM1^{-}$ cells became predominant. These cells were termed human hepatic stem cells (hHpSC). When cultured

on STO feeders (a mouse embryonic fibroblast line) cells, hHpSC will differentiate into $EPCAM^{+}NCAM^{-}AFP^{+}ICAM1^{+}$ hepatoblasts [94]. The proliferative capacity of these hepatoblast cells were demonstrated when the authors showed the maintained phenotype for over 150 doublings in culture.

2.5.3 Adult liver stem cells

Hepatocytes in an adult liver are capable of being highly regenerative when the liver experiences acute organ failure. Upon acute liver injury or partial hepatectomy, they replicate to generate the lost cell mass. Through serial transplantation of cells into the liver of metabolically deficient transgenic mice, hepatocytes isolated from adult mouse liver have been shown to proliferate up to 70 doublings and continue to contribute to the liver mass [96]. However, hepatocyte replication is not the only mechanism by which the liver can regenerate. In situations of chronic insult or severe liver injury, the regenerative capability of adult hepatocytes is overwhelmed and the resident hepatic stem cells will become the key contributor to the repair of the damaged hepatic tissue in a process known as ductal reaction [85,95]. These proliferative cells are often referred to as oval cells due to their shape in rodent livers.

Thus, both hepatocytes and stem/progenitor cells in adult liver are capable of proliferation and reconstitution of lost liver mass. Such capability can potentially be harnessed for *in vitro* culture and clinical applications. However, as briefly mentioned before, hepatocytes in culture lose their high proliferative capacity. In contrast, many have focused our ability to isolate and culture liver stem cells or progenitor cells in hopes of utilizing these specialized cell types towards clinical applications.

Study of liver stem/progenitor cells often employs partial hepatectomy or treatment with drugs like acetylaminofluoren in rodents to enrich the cell population in the liver [97]. By BrdU labeling of proliferative cells in regenerating livers of rodents after induced acetaminophen injury, the region that stem cells reside and their niches were identified [98]. Four differently labeled stem cell populations were identified, namely periductal mononuclear cells and peribiliary hepatocytes along with cells in the canal of Hering and intralobular bile ducts [98]. These findings suggest that there may be more than one cell type that gives rise to the regenerative process. Instead, multiple cell populations may

possess this ability and can respond to their specific niche cues differently allowing for their participation in the liver repair process.

The availability of genomic data and the enhanced capability to identify genes expressed differentially under various differentiation states has helped identify markers suitable for cell isolation. The progenitor/stem cell markers identified in rodents has been used to enable for selective isolation of their counter parts in human liver. For example, EPCAM has been a key marker for enriching both fetal and adult liver stem/progenitor cells in rodents and human [94,99]. However, through the detection of high level expression of aldehyde dehydrogenase using a fluorescent substrate, adult liver stem cells from both mice and human liver were isolated without using any chemical stimuli to enrich their population [100]. This methodology could possibly provide a readily robust method of isolating a relatively homogenous population of hepatic progenitor cells.

A recent review by Turner et al. provides a more in-depth look at the current understanding of the localization of liver stem cells in the liver lobule and their characteristics [95]. They comprehensively describe the progression of liver stem cells during the development of the liver. It is postulated that the hepatic stem cells reside in the periportal region of the liver, specifically the canal of Hering. As these cells proliferate, "older" cells migrate along the cell plate toward the central vein and grow more committed and mature; first becoming bipotential, then becoming proliferative hepatocytes, and mature hepatocytes. As they move toward the central vein, they also become larger in size. Further details regarding the origin and repair mechanisms are described elsewhere [85,95,97].

Figure 2-5 summarizes the current stem cell/progenitor cells populations which have been isolated from stem cell hepatic differentiations, fetal liver, and adult liver. Multipotent cells from human fetal liver reported by the Fausto lab [93] are more primitive than the bipotential hepatic stem cells characterized by the Reid lab owing to the former's capability to differentiate to mesenchymal cells.

The hepatic stem cells have been isolated from both fetal liver and adult liver; whereas the mesendodermal cells have only been reported from human fetal liver. Both precede the hepatoblast; although both share many markers of hepatoblasts including EPCAM, E-

cadherin, CK8, CK18, and CK19, both are also AFP and albumin negative. A distinction between hepatic stem cells and hepatoblast cells is their NCAM/ICAM-1 expression; ++/- and -/++ for hepatic stem cell and hepatoblast respectively.

Hepatic stem cells isolated from human adult liver are about 7-10 µm in diameter [95,100]; somewhat smaller than hepatoblasts in both adult and fetal liver (10-12 µm). They exhibit a high nucleus to cytoplasm ratio. Hepatoblasts are abundant in fetal liver (~80%), but decreases after birth to less than 0.01% in adult liver. In contrast, hepatic stem cells remain relatively stable comprising of about 0.5 to 1.5% of the liver cell mass throughout [101].

Hepatic growth factors HGF, EGF are often included in the medium for hepatoblasts. In addition to growth factors, surface properties also affect the outcome of cell fate. Schmelzer et al. reported that both hepatic stem cells and hepatoblasts grew out from EPCAM sorted cell population when plated on plastic surface, but only hepatic stem cell colonies emerge when plated on STO feeder layer [94]. Turner et al report that angioblast feeder layers or collagen III aid in proliferation of hepatic stem cells while stromal

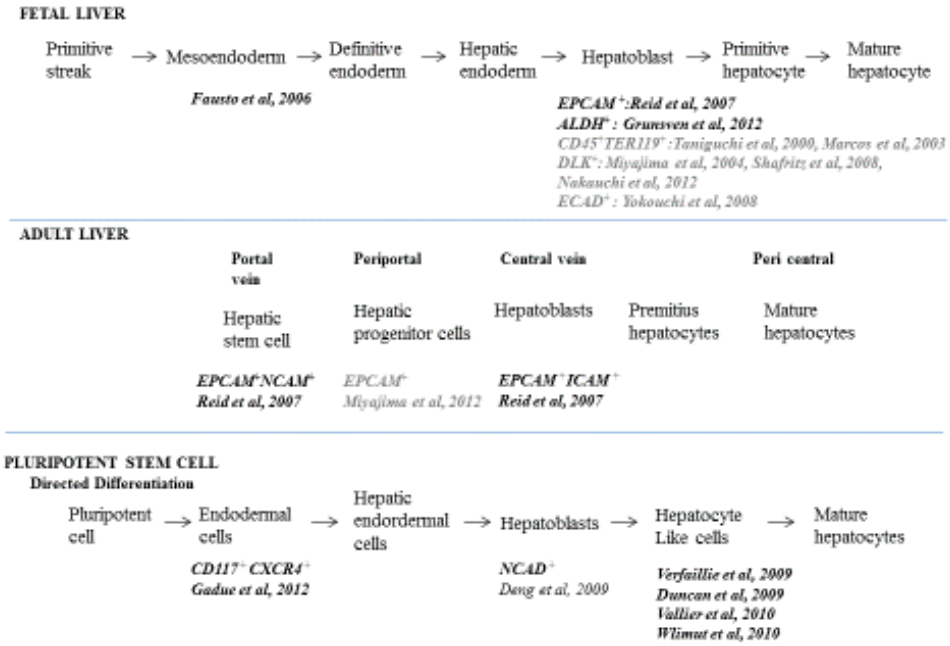


Figure 2-5: Isolation of Progenitor Cells in the Hepatic Lineage [6]

feeder cells or collagen IV aid in hepatoblasts proliferation [95] .

2.6 Differentiation to hepatic lineage from stem cells

Due to the liver's high regenerative capacity, there has been a long history of research on liver stem/progenitor cells and on their *in vitro* expansion. Those attempts aim to expand cells already committed to a developmental path. In contrast, another set of efforts seek to direct uncommitted embryonic stem cells, to the hepatic lineage. Embryonic stem cells have unlimited self-renewal and differentiation capacity thereby resulting in a potentially unlimited supply of hepatocytes. The recent emergence of iPSCs only amplified efforts towards generating functional hepatocytes from pluripotent stem cells.

2.6.1 Differentiation of hepatic progenitor cells isolated from liver

Hepatoblasts and liver stem cells isolated from fetal and adult livers have been shown to differentiate to hepatocyte-like cells in culture. A combination of growth factors, HGF, EGF, Oncostatin and Dexamethasone administered over one week was used to induce hepatic differentiation for mouse hepatoblasts [100]. Hepatic progenitor cells isolated from humans can be expanded and differentiated to hepatocyte-like cells with many liver functions including urea synthesis, albumin secretion and CYP activity. Decellularized liver biomatrix was used as an ECM for human differentiations, thereby reducing the need of commonly used growth factors HGF and EGF [100,102].

2.6.2 Hepatic Differentiation of ES and iPS cells using soluble growth factors

The development of liver *in vivo* entails the specification to definitive endoderm (about E8.5 in mouse), followed by commitment to hepatoblast and formation of liver bud, and finally fully differentiated hepatocytes and the emergence of bile ducts. The progression of the development is guided by a number of inductive signals dynamically described briefly above. The early differentiation to mesoendoderm and subsequent distinction of mesenchyme and endoderm is driven primarily by *Nodal,* BMPs and Activin signaling [103-105]. Further signaling from the FGF and BMP family, specifically BMP4, FGF2, FGF4 induce differentiation to hepatoblasts. After liver bud formation and expansion, the inductive signals of HGF and Oncostatin stimulate the hepatoblasts to differentiate towards hepatocytes.

Most hepatocyte differentiation protocols start with Activin and Wnt3a for a 3-5 day period to induce definitive endoderm commitment [103,106-108]. It has been shown in mice that Wnt3 signaling can maintain *Nodal* expression and Activin signaling can replace the *Nodal/Crypto signal needed for endodermal formation* [109,110]. FGF and BMP play a significant inductive role in promoting endodermal progenitor maintenance and expansion and will subsequently be used in the second stage of directed differentiation [68,71,111,112]. Further treatment with a mixture of FGFs facilitates the commitment to hepatic fate or the equivalent of the hepatoblast state in liver development based on the *in vitro* findings of Sekhon et al [113]. The last stage often entails the use of oncostatin or follistatin and HGF mixed along dexamethasone. It was previously reported that oncostatin-treated stem cells would lead to an increased activity level of STAT3 [114]. The use of follistatin or oncostatin is used to favor hepatic differentiation over cholangiocytes, whereas HGF is used universally to mimic the hepatic environment [115]. HGF, on the other hand, was shown to contribute to liver development in a

STAT3-independent manner [114].Depending on the protocol, dexamethasone is often used as a mature hepatic inducer [116].

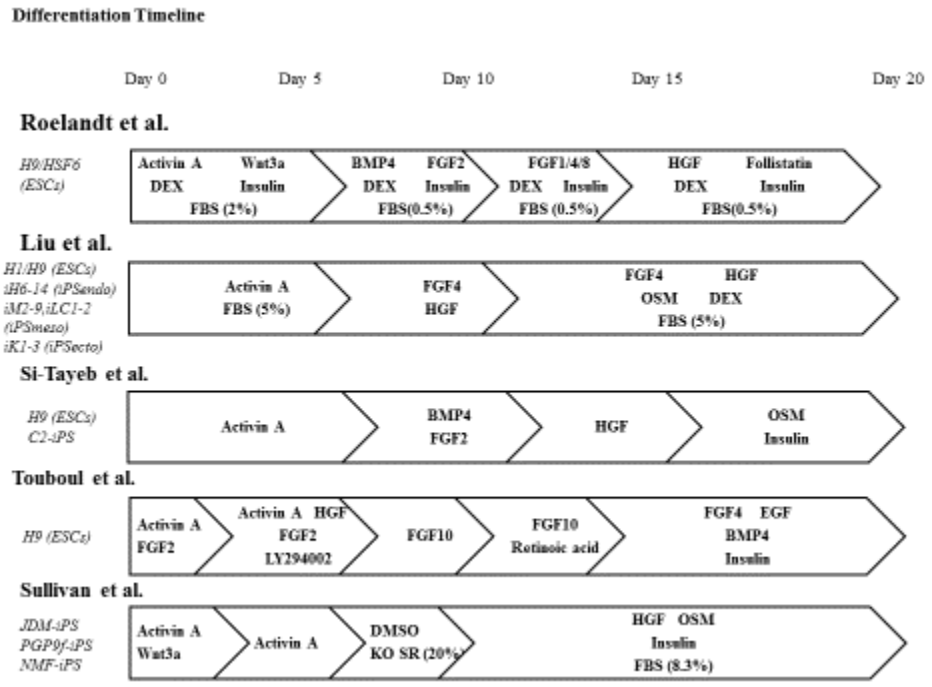


Figure 2-6: Comparison of Various Differentiation Protocols [6]

Most efforts in guiding the differentiation of PSCs to hepatic lineage utilize multiple soluble growth factors to mimic the temporal dynamics of cues in hepatogenesis *in vivo*. However, it should be noted that the development of the liver and any other tissue *in vivo* is a continuous process, whereas our efforts to replicate these processes tends to be in discrete stages. Various protocols for guiding the differentiation of stem cells to hepatic lineage are segmented into separate differentiation stages. Thus, each protocol may implement different numbers of stages, with varying duration and medium composition. However, most protocols share common motifs across the duration of their differentiation which is depicted in Figure 2-6 that illustrates several representative protocols for differentiation to hepatic lineage *in vitro*.

2.6.3 Transdifferentiation to liver lineage from other tissue cells or stem cells

In addition to the PSCs, a number of extrahepatic stem cells isolated from adult tissues, including adult mesenchymal stem cells (MSC) derived from human bone marrow [117,118] and adipose tissues [119,120], were shown to have the capacity to differentiate to hepatic lineage. A protocol calls for treating MSCs with demethylating agent 5'-azacytidine and then culturing them in hepatocyte maintenance medium containing HGF and EGF [118]. After 3 weeks the differentiated cell exhibited hepatic markers including albumin, CYP1A1, CYP3A4. Engraftment was shown upon transplantation into immune-compromised mice which underwent partial hepatectomy. A similar protocol for adipose tissue-derived MSC underwent hepatic differentiation without the demethylation step [120], however, comparison of the effect of demethylation has not been available. The differentiation status of hepatocyte-like cells derived from MSC has not been characterized as well as the hepatocyte-like cells derived from PSCs. Nevertheless the results indicate that the future source of hepatocyte can still potentially be expanded to other adult stem cells in addition to the conventional PSCs.

2.6.4 Directing to liver cell fate through gene transfection

A number of laboratories have demonstrated the differentiation of iPSCs to hepatocyte-like cells using the protocols developed for ES cells [121-123]. It appears that the protocols developed for embryonic stem cells are mostly applicable to iPSCs. This opens the possibility of personalized hepatocyte-like cells for drug screening and even for therapy. However, deriving iPSC is a lengthy process. Recently the feasibility of directly inducing fibroblastic cells to hepatic lineage through gene transfection mediated reprogramming was demonstrated in mice models. A combination of HNF4α with FOXA1, FOXA2 or FOXA3 was used to transfect and reprogram adult mouse fibroblasts to hepatocytes-like cells which are called iHEP cells [124]. The efficiency was about 1 in 1000 of initially transfected embryonic fibroblasts. Interestingly the resulting iHep cells possess a very high proliferative potential, whereas such proliferative potential has not been reported for hepatocyte-like cells derived from from pluripotent cells. In another

study the over- expression of *Gata4*, *Hnf1α* and *Foxa3* along with $p19^{Arf}$ inactivation enabled mouse fibroblasts to become iHep cells [125]. In this case inactivation of $p19^{Arf}$ suppresses senescence of fibroblasts to enhance the success of reprogramming. The reprogrammed iHep cells expressed key hepatic genes and other liver functions and are successful engrafted in animal models. The obvious question then is why do the resulting iHep cells have a higher proliferative potential compared to the hepatocyte-like cells derived from pluripotent cells. The differences and similarities between the different types of hepatocyte-like cells derived from stem cells and primary hepatic stem cells have yet to be studied extensively. However, the approach will likely see many *in vitro* applications if it can be shown to work in reprogramming human cells.

2.6.5 Expansion and differentiation of *in vitro* derived progenitor cells

All reported directed differentiation of ES or iPS cells typically span a 20-25 day period and give rise to a mixture of differentiated cells that include a high percentage of hepatocyte-like cells at different degrees of maturity as assessed by the various liver specific markers. In the transition of pluripotent state to definitive endoderm stage, typically a significant degree of cell death is seen. Some cell expansion is typically seen in the endoderm stage and in the bipotential hepatic progenitor state. At the end of directed differentiation the differentiated cells typically are nearly confluent on the culture surface [126].

In embryo development, cell expansion accompanies endoderm development, hepatic lineage commitment as well as in the hepatoblast stage. Directed hepatic differentiation of pluripotent stem cells passes through the corresponding differentiation stages of *in vitro* equivalents of the endoderm and hepatoblast stages. In these stages they are also likely to be capable of proliferation although their proliferative potential has not been quantified. Zhao et al. isolated hepatic progenitor cells from hepatic differentiation of ES cells, using FACS sorting of N-cadherin positive cells [127]. These colony forming cells were cultured on mouse embryonic stromal feeders using serum free DMEM based medium which promoted their expansion. These cells, like their counter parts isolated

from fetal and adult livers, were bipotential and could be passaged several times in culture.

Recently endoderm progenitor cells were isolated based on their expression of CXCR4 and CD117 from cultures of directed human ES or iPS cells undergoing directed differentiation to the hepatic lineage [111]. These cells are apparently in an earlier differentiation stage than the hepatoblast stage and are capable of differentiating to both pancreatic and liver lineage. They possess a very large proliferative potential and can be maintained in medium containing BMP4, FGF, VEGF and EGF. These endoderm progenitor cells are at an earlier developmental state than the hepatic stem cells and hepatoblasts; but are not as early as the cell line reported earlier by Fausto's group that is capable of differentiating to mesenchyme in addition to endoderm.

2.7 **Properties of the *in vitro* differentiated hepatocyte-like cells**

The hepatocyte like cells derived from directed differentiation of pluripotent stem cells all have distinct epithelial morphology, express liver specific genes including albumin, alpha-antitrypsin, and a number of cytochrome P450 enzymes. A direct comparison of different protocols is not easy because different combinations of assays were used in each report and the functionality is often evaluated at different levels, transcripts or protein or functional activities. Typically the maturity of the differentiated cells is evaluated by transcript and/or protein levels of marker genes, and assessed by representative characteristic markers of different aspects of liver functions, including drug metabolism, bile secretion, gluconeogenesis, urea synthesis, and protein synthetic functions (serum albumin, alpha antitrypsin). As the genomic tools become widely employed and data on transcriptome, miRNA and epigenetic state of cells at different differentiation states become available, we should gain a better understanding of cell differentiation under different protocols. The hepatocyte-like cells obtained from directed differentiation of pluripotent cells are not mature hepatocytes. While they express higher level of transcripts of liver specific gene, the levels generally fall short of that of primary hepatocytes, especially for cytochrome P450 enzymes. As hepatocytes mature, many enzymes in metabolic pathways switch from the dominant infant isoform to an adult

isoform. Gluconeogenesis is one key metabolic pathway that is observed primarily in the liver. Using microarray analysis, we can quantitatively observe that the transcript-level of a fetal-specific isozyme, ADH1A, was significantly higher than its adult isozymes, ADH1C (data not published). Detailed transcriptome analysis then suggests that our current methodology of deriving hepatocytes from stem cells is more representative of the fetal state than the adult tissue.

The capability of in-vitro cultured hepatocyte-like cells to progressing to mature hepatocytes have been demonstrated in animal models as illustrated in several studies [122,128]. However, the cues necessary to achieve that *in vitro* have yet to be unveiled.

To increase the percentage of cells differentiated to hepatic lineage and enhance their maturity, transient forced gene expression of hepatic transcription factors was used. Lentiviral tranfection of FOXA2, HNF4α and C-EBPα in adult liver derived progenitor cells increased hepatic maturity and functional capabilities of albumin secretion and glycogen storage [129]. Similarly, by infecting differentiating ES cells with SOX17 at Stage 1, HEX at Stage 2 and HNF4 at stage 3 of hepatic differentiation and further incubation for 11 days, hepatic cells expressing higher transcript levels of all major genes examined was observed [130]. In a follow up study, adenoviral vector with Foxa2 was used to infect differentiating pluripotent stem cells at the definitive endoderm conversion, hepatic specification and hepatic expansion stages, followed by a combination of adenoviral vector-FOXA2 and HNF1 infection to further increase maturation [131]. The cytochrome P450 activities were markedly increased, albeit still lower than those in primary hepatocytes and the level desired for toxicity testing. Nevertheless, the work represents encouraging progress for *in vitro* generation of mature hepatocytes by directed differentiation.

2.8 Cell Culture Processes

2.8.1 Therapeutic proteins VS. Stem Cells

The optimal cell culture process used for manufacturing stem cells needs to be designed with the end use in mind. As previously discussed, the fundamental property of stem cells to be able to differentiate into various specialized cell types and lineages enables their

use in treating diseases as well as drug discovery and toxicity testing. The latter applications have gained even more attention with the discovery of iPSCs, as differentiated cells representing a diverse genetic background can now be generated from a large pool of patients. To this end, initiatives are launched to assemble banks of iPS-derived cells for drug discovery, as exemplified by the pre-competitive initiative for iPS-cells banking launched by a group of pharmaceutical companies called Stembancc (http://www.stembancc.org/) or by the University of Kyoto (http://www.cira.kyoto-u.ac.jp/e/). For stem cell applications in drug discovery or toxicity testing, the quantity of cells that are required may not be large. Microfluidic devices with automated cell or medium manipulation capability are well suited to allow for process miniaturization and large scale parallel and combinatorial testing [132,133]. Nevertheless, the generation of stem cells as off-the-shelf reagents in larger quantities may require a different approach.

In contrast to drug discovery and toxicity testing applications, stem cell based therapy at the clinical trial stage and beyond requires robust bioprocess technology in order to generate products of consistent quality at sufficient quantities. Cell culture processes have already been used for the production of therapeutic proteins for more than thirty years; this segment of the biotechnology industry currently represents tons of proteins with a commercial value of more than US $50 billion per annum worldwide [134]. The processes, technological know-how, and learning gained through these years of practice can now serve as a guide to develop an optimal process technology for stem cell applications. To benefit from those experiences and translate them to the stem cell field, it is important to first identify the similarities and the differences that exist between traditional cell culture processes for therapeutic protein production and emerging stem cell processes.

The production cell lines for therapeutic protein production were first derived from transfecting the host cells with a vector containing the selective marker and the transgenes [135]. The transfected cells invariably undergo single cell cloning and screened and tested for productivity and product quality before they are further expanded to prepare master cell banks (MCBs) [135]. Prior to cell banking, further tests on microbial contamination, cell line stability are performed to ensure product quality and

consistency. It is not unusual that by the time a master cell bank is prepared those cells have undergone nearly forty population doubling [135]. Frozen vials from MCB are then used for further expansion to generate working cell banks (WCBs) from which cells for the manufacturing runs are obtained [135]. It is typical that the maximum passage number when reaching the final production scale is defined for a manufacturing process [135]. But over the lifetime of a product a total of 60-80 population doublings since single cell cloning would have occurred [135]. The cell banking protocols for stem cells are likely to be fairly variable depending on the type of stem cells used. While iPSCs are virtually all single cell cloned at their isolation stage, MSCs and others are used as unpurified population. Nevertheless with the exception of some autologous processes cell banking will be practiced and extensive quality control, including the control of number of population doublings or passages, will be imposed as practiced in current stem cell banking [136].

A typical manufacturing process for therapeutic protein production is initiated using cells obtained by thawing a frozen vial from a WCB. Subsequent cell expansion is performed in a series of bioreactors with increasing scale, before reaching production. For most recombinant proteins, the final production culture is operated under fed-batch mode with a scale ranging from 1,000 liters to 12,000 liters [30]. Fed-batch productions are initiated by inoculating with a seed culture derived from the WCB filled to 60-70% of the bioreactor capacity; nutrients typically fed in steps to bring the culture volume to a full capacity and avoid too high concentrations of nutrients at the beginning of the process. The production process may last for 12-15 days, reaching a final cell concentration of about 10^{10} cells/L and up to 10g/L of protein product [137]. Due to the nutrient feeding and base addition to neutralize lactate produced by proliferating cells, the final osmolality value may reach 400 mOsm or higher, far beyond 280-300 mOSm that is optimal for cell growth [137]. The protein product is recovered through a series of unit operations, which include cell removal, followed by a number of chromatography and membrane filtration steps. Residual cellular DNA is removed through an anion exchange adsorption process and possible retroviral contamination is eliminated through viral inactivation steps such

as low pH incubation. The final product is obtained after buffer exchange and sterile filtration [137,138].

In the cell expansion bioreactors for therapeutic protein production, the cell concentration may reach 2-4 x 10^9 cells/L; while in the final production reactor, a higher concentration of 10^{10} cells/L is commonly observed [139], [138]. A bioreactor of 1,000 L thus may contain 10^{13} cells.

The number of cells required per dose of live cell therapeutics is relatively small in comparison. In Phase I of Geron's clinical trial for spinal cord injury, 2 million oligodendrocyte progenitor cells were transplanted at the site of injury (http://cell-therapies.geron.com/grnopc1). Cell transplantation for liver disease treatment will likely require at least 10^9 cells or about one tenth of liver mass [140]. A similar number of beta cell equivalents is estimated to be required per patient for therapy of type I diabetes [141,142]. The Phase I clinical trials of acute myocardial infarction using Prochymal, bone marrow-derived human mesenchymal stem cells, by Osiris Therapeutics, Inc., used a dose of 300 million cells per patient. Moreover, in Phase 2 of clinical trials using Prochymal for the treatment of acute graft versus host disease (GvHD), 2-8 million cells were administered per kg of patient body weight (the equivalent of about 1-5 x 10^8 cells per patient) (http://www.clinicaltrials.gov/ct2/show/NCT00482092). Similarly, in Phase 3 clinical trials of Prochymal for treatment-resistant moderate-to-severe Crohn's Disease, 600 million (low dose) to 1200 million (high dose) cells were intravenously delivered in four infusions (http://www.osiris.com/clinical_prochymal_piii_NDAGVHD.php). The number of cells required for therapeutic applications thus varies from about 10^8 to 10^9. Compared to the production process for therapeutic proteins, the operating scale is likely to be smaller even if thousands of doses are produced in each run. However, the complexity of the process is likely to surpass that for protein production.

Traditional flask culture, or its variations such as parallel flat plates, can routinely generate 10^{10} cells per production batch [139]. However, for economics of scale and thus to reduce the final cost of goods, a large number of doses should be produced in each batch. This will likely result in bringing the number of cell produced per batch to almost 10^{12} cells, a level beyond the reach laboratory apparatuses. Larger scale production using

a bioreactor will also bring down the cost associated with assessing product quality, which is a major cost item in the manufacturing of the biologics product [139]. Beyond a 3-dimensional environment that most closely mimics the natural cell niches, other advantages of using bioreactors include improved control over process parameters such as pH, dissolved oxygen or nutrient levels, and. Scalability. Even for smaller scale applications, as observed in industrial microbiology, cultivation in a well-constructed bioreactor enhances process robustness and product quality consistency [139].

Commercial stem cell processes will increasingly bear a close resemblance to processes that have been developed for the production of recombinant protein, with the exception that batches will remain smaller in scale that those typically conducted to manufacture recombinant enzymes [143]. This is especially true at the cell expansion stage. On the other hand, the final steps of the production processes differ between industrial microbiology processes and cell therapeutic manufacturing processes, since in the former the aim is to achieve a final product titer while in the latter aim is to generate a live cell therapeutic at its optimal physiological state. Notably, the optimal physiological state of the live cells thus produced is defined by their ultimate application and does not necessarily relate to maximal cellular fitness. Beyond the field of stem cell therapeutics, an example of that concept can be found in the production of attenuated live vaccines. Those two very distinct process objectives thus impose a very different design space for defining process parameters. While it is a common practice for a recombinant protein process to reach a high osmolality due to the addition of base for neutralizing the pH and nutrient supplements; such culture conditions are likely to cause extensive cell damage and low viability and as a result are best avoided for stem cell manufacturing processes.

It is relevant here to invoke a historical perspective on the evolution of cell culture processes during the past quarter century and relate these lessons learned to stem cell bioprocessing. At the dawn of cell culture production of therapeutic proteins, it was a common practice to provide cells with optimal chemical and physical conditions for growth. It was not until the years 2000 and particularly the advent of post-genomic technologies and systems biology techniques that the focus was shifted to finding the conditions that give rise to the highest productivities but not necessarily providing cells

with the most optimal conditions for cell growth [144]. Although we envision the employment of stem cells in bioreactors with culture conditions optimal for growth or differentiation, it is not unimaginable that a decade from now stem cell culture processes may look entirely different. For example, rather than operating at optimal

conditions for cell growth or differentiation, the processing may be tailored to suit the purpose of producing cells with not only high quality and consistency but also superior efficacy for the specific treatment considered.

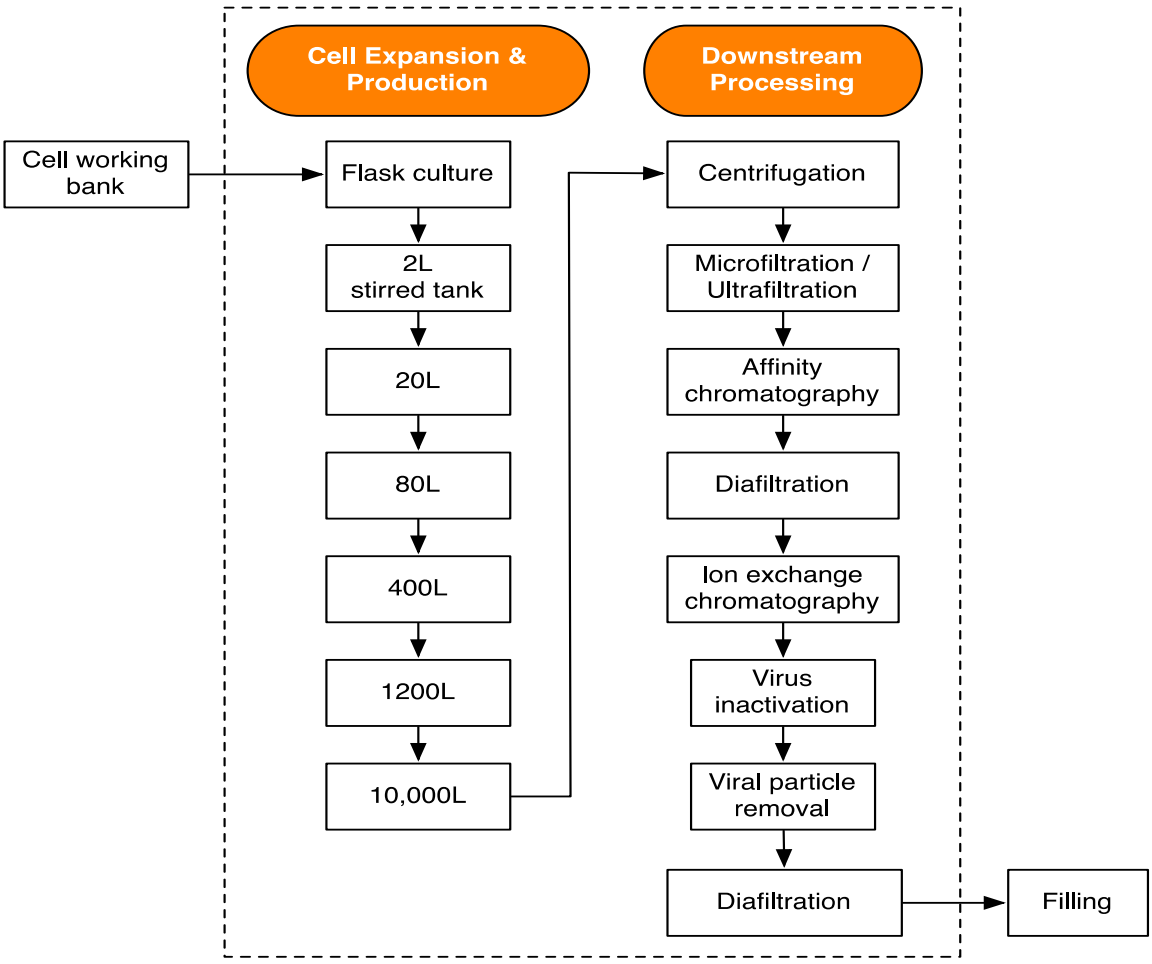


Figure 2-7: Workflow of typical a cell culture process for recombinant protein production [5]

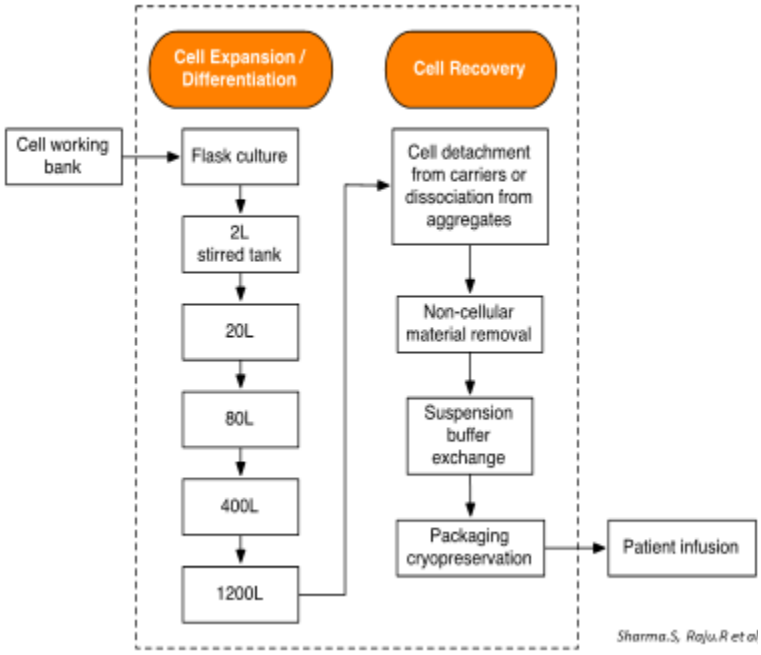


Figure 2-8: Scenario of a stem cell process[5]

2.8.2 Autologous and allogeneic applications from a bioprocessing perspective

To generate clinical grade stem cells products, the manufacturing process to produce stem cells will need to meet clinical requirements of cell quantity and quality. A major factor that will affect process decisions, and particularly obviously that of scale, is whether the cells will be used for autologous (using patient's own cells) or allogeneic (using donor's cells) applications.

2.8.2.1 Allogeneic cell process

A major advantage of employing allogeneic cells is that of establishing cell banks to derive industrially robust cell manufacturing systems for producing off-the-shelf cell preparations. Adult cells isolated from tissues, cell clones established from hESC or cells derived by reprogramming somatic cells using the iPSC protocol, all exhibit a wide range of intrinsic properties. For allogeneic applications, the "best" cells can be isolated for a particular application and characterized for their differentiation potential, risk of genetic abnormality or instability, and absence of adventitious contaminants as well as viability.

Qualified cells can in turn be used to establish cell banks for production use over the entire product life cycle. This is an approach that has been implemented by companies developing MSC products, including among other Mesoblast, Athersys, or Pluristem [145]. The consistency in cell source during different runs, ensured by the use of MCBs and WCBs, also facilitates the development of a robust process to produce cells of consistent quality such that all patients are adequately treated. Moreover, allogeneic applications enable large-scale operations for the manufacturing of a large number of doses for each batch; this results in economy of scale with the additional flexibility that the same manufacturing suite can be used for producing different stem cells products by operating in production campaigns, as it the common practice in industrial microbiology plants.

2.8.2.2 Autologous cell processes

Autologous treatment using the patient's own cells avoids risks associated with immune rejection as a result of graft vs. host reactions; notably, these treatments are also developed for MSCs despite these cells being recognized as immune privileged. With the prospect of generating iPSCs from the patient's somatic cells, patient specific therapy seems increasingly plausible thus bringing the concept of personalized medicine to its most individual level [146]. The first step of an autologous stem cell production processes is that of cell isolation from either the patients' own tissues or by the reprogramming of the patient's somatic cells to generate iPS cells. What is more, cell manufacturing may extend beyond *ex vivo* expansion to include for example engraftment of the resulting cells in a tissue-engineering construct and sometimes further cultivation for additional expansion or differentiation. The quality and quantity of isolated cells and their response to culture conditions may vary from patient to patient. Developing and implementing a robust process that minimizes the variability of process outputs and the clinical outcome thus represents a major challenge. Processes employing autologous adult stem cells and iPSCs likely will involve extensive cell manipulations and may *ipso facto* be subject to a higher risk of microbial contamination. It is therefore critical to define the ranges of variability of each key production parameter, to establish strict change control procedures, and implement a model for predicting process output. Here

again, the best practices of industrial microbiology can be adapted with only minor changes to stem cell manufacturing.

In pharmaceutical manufacturing, the production processes of different products are typically completely segregated from one another by either space or time to eliminate risks of cross contamination. Products simultaneously manufactured in the same plant are completely separated in space with its own supplies and staff [137]. For autologous stem cell processing, cells from each patient must be processed separately by time schedule, or by space. Such a production scheme may seem accomplishable in clinical trials or in the experimental stages, but will certainly pose additional operational constraints in the long term as the new technology reaches a larger population base. Processes for autologous applications will by nature be conducted at a scale that is much smaller than those used for allogeneic cells. Consequently, rather than economies of scale, the economics of autologous manufacturing stem cell facilities will be driven by economies of learning and economies of scope as a large patient base is reached and thus a large number of parallel lines of cultivation trains is deployed, with each train designated for only one patient at a time. Notably, this latter mode of manufacturing is more akin to a service business model such as that of a specialized clinic or hospital, than the traditional large-scale production of mass-market off-the-shelf products that large pharmaceutical firms have followed to date. One can envision that manufacturing cost considerations may drive autologous stem cell processes to become fully automated with robotic operations and artificial intelligence based personalized process optimization just like the erythropoietin process of roller bottle production was automated by robotics nearly three decades ago [147]. It is not too early to begin contemplating efficient designs for the large scale manufacturing of stem cells or their products.

2.8.3 Open system vs. bioreactors

Therapeutic applications based on hESCs or iPS cells require not only an *ex vivo* cell expansion step but also differentiation to the desired cell lineages. As discussed in the preceding paragraphs, the *ex vivo* expansion stages will most likely bear resemblance to the typical cell cultivation for recombinant protein production. Recombinant protein

production processes involve thawing frozen vials of cells and expanding them in a serial manner from a seed culture and progressively increasing the scale to thousands of liters [137]. At the end of the fermentation process, cells are separated from the spent growth medium and are discarded while the spent medium is further processed to isolate recombinant products. The use of flasks in a production process, with the exception of viral vaccine production, is limited to the steps immediately following initial thawing from a cell bank. In any of the remaining scale up cultivation steps, cells are in the completely enclosed environment of bioreactors. The opening of a flask destroys this barrier and generates *de facto* the risk of introducing contaminants or even adventitious agents. Therefore, even if some operations are performed in laboratory vessels, it is worthwhile to explore the possibility of a bioreactor system, or at least to establish a secure barrier between cells and their surroundings, and ensure that culture integrity can be maintained even during transfer of cultures between different reactors.

2.9 Stem cell niche and optimal culture conditions

To develop optimal culture conditions, it is useful to understand the niche or the native microenvironment in which the cells reside. Stem cells have extensive interactions with their underlying matrix as well neighboring cells [148]. Those complex interactions provide cues for their renewal, maintenance, differentiation or apoptosis [148]. Understanding the nature of these stem cell niches will facilitate the development of processes from cell renewal and expansion. Some of the well studied stem cell niches include hematopoietic, skin, intestinal and muscle stem cells [149]. However, these niches are all rather complex. A plethora of intricate regulatory mechanisms, including cell-cell interactions and various spatial and temporal cues coach the cells to remain in a quiescent state, to transform into an activated state, or to differentiate. Even stem cells in a niche may be at different states as is governed by changes in specific set of regulatory cues. In fact most stem cells *in vivo* are likely to be at a resting state instead of proliferating; they begin to proliferate or differentiate only upon stimuli of the appropriate cues. Whereas stem cells in culture are continuously self-renewing except when under directed differentiation. While striving to understand stem cell niches, it also

important to keep in mind the dynamic nature of those niches in modulating stem cell state.

Hematopoietic stem cells (HSC) residing in the bone marrow or blood can isolated from clinical samples to provide invaluable treatments of hematological and immune disorders as reviewed and tabulated by Burt et al [150] These cells have the potential to generate all the different types of cells of the hematopoietic and the immune system [151]. In the past two decades, significant advances have been accomplished for their isolation and subsequent transplantation [150]. Notably, to this end the knowledge gained regarding their specific niches in which they naturally reside is being used to guide and implement novel stem cell-based treatment strategies [151]. HSCs are associated with two different cell types within the bone marrow: osteoblastic cells and vascular cells [149]. These supporting cells interact with the HSCs either through soluble factors or through direct cell-cell interactions; remarkably, some of these mechanisms can be replicated in the laboratory of manufacturing suite to promote HSC expansion *in vitro* as reviewed in [152]. Particularly, this can be attempted through manipulating the Notch signaling pathway, since on the one hand HSCs express Notch receptors and on the other hand the supporting cell types in the bone marrow express Jagged 1, a Notch ligand [152]. Notably, soluble forms of Jagged 1 have been employed to expand HSCs isolated from bone marrow [152]. Additionally, osteoblasts secrete proteins such as angiopoetin and thrombopeotin that interact with their receptors present on the surface of HSCs to maintain them at a quiescent state; the vascular cells provide signals for HSC activation to generate mature differentiated cells to repair daily wear and tear [149].

From these various observations, it is clear-cut that the stem cell microenvironment plays an essential role in determining the appropriate culture conditions for stem cell culture and expansion. The niche presents a specialized platform for cross-talk between the stem cells and their microenvironment, involving compounded effects of cell-cell signaling,

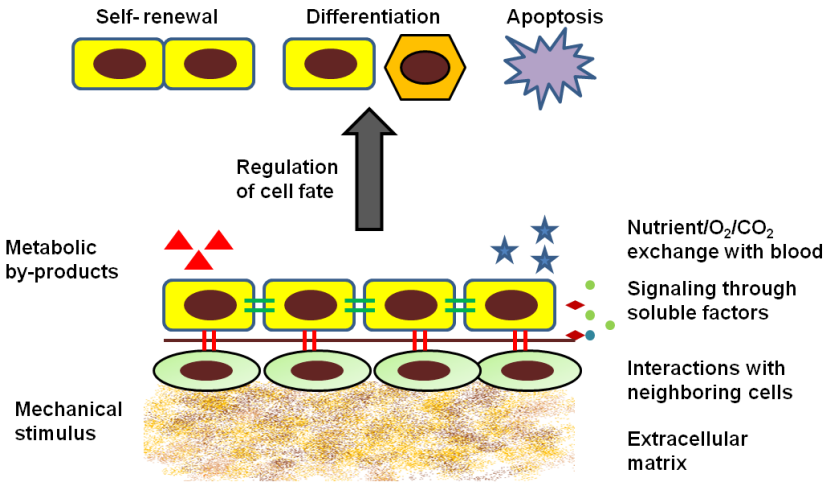


Figure 2-9: Components of a Stem Cell Niche [2]

soluble factors and cell-ECM interactions as shown in Figure 2-9.

Notably, advances in high-throughput technologies enable large scale screening of novel supporting factors, or small molecules with similar effects, and have shown great potential in deciphering these combinatorial interactions, whether they include growth factors, a variety of ECM molecules or mixtures of soluble signals. For example, such screens were employed to dissect the influence of ECM proteins on hepatocyte specification from ESCs. Thirty-two different combinations of ECM proteins were evaluated using a promoter-reporter system for fetal liver gene and a fold change of 140 was observed between the least and most optimal conditions [153]. With the use of such high-throughput platforms, novel regulatory features can be uncovered and applied towards optimization of a well-defined environment for the maintenance of stem cells, even outside their native niches.

In contrast to adult stem cells, which have their equivalents *in vivo*, embryonic stem cells are derived from the blastocysts, a transitory state that occurs only during embryonic

development [14], the "niche" concept may be very difficult to apply for these cells since the relatively stable ESCs observed in cultures do not exist *in vivo*. The first isolation and derivation of human embryonic stem cells was accomplished by culturing them on a supporting layer of mouse embryonic fibroblasts (MEF), which have been rendered non-proliferative through irradiation treatment or exposure to chemicals [14]. The mechanism of action of these so-called feeder cells is to continue to support the undifferentiated state of ESCs by providing a microenvironment suited for self-renewal [154].

Remarkably, human ESCs were demonstrated to propagate under MEF-free conditions when cultured in the presence of mouse fibroblasts conditioned medium and on Matrigel, a complex mixture of extracellular proteins derived from mouse Engelbroth-Holm-Swarm sarcoma cells [155]. This observation allowed for a systematic analysis of the factors present in the conditioned medium in an effort to determine the key components responsible for the remarkable maintenance of stemness of ESCs thus achieved. Of the several factors studied, basic fibroblast growth factor (bFGF) is one component that has been implicated to play a major role through modifications in the BMP signaling pathway: as a result large concentrations of bFGF have been employed to successfully establish reproducible feeder-free cultures [155-158].

The inclusion of animal-derived components in the culture medium is ideally to be avoided, not only due to lot-to-lot variations, but also due to the risk of cross contamination with xenogenic components; and is of great concern for clinical applications. The transmittance of infectious virus or mad cow disease has been associated with the use of animal serum or serum-based proteins and has fueled the move towards more defined culture systems [159,160]. A systematic analysis of the complex feeder based methods has already provided useful insights for the development of well-defined, xeno-free ESC media such as mTESR and Essential-8-medium, which includes bFGF along with other important growth factors [156,158]. As it is evident in the use of MEF in ES culture, key components of the niche are often associated with auxiliary cells present in culture. The approach for medium development is thus to reproduce the niche environment with paracrine factors and all of the key ECMs provided by those auxiliary cells.

Even though these factors are invariably difficult to characterize and quantify, efforts are being made to elucidate the supportive components of the extracellular matrix secreted by the feeder cells in an attempt to move towards feeder-independent cultures. Matrigel, which as previously described is commonly used in ESC culture, is derived from mouse sarcoma cells; this creates de facto additional risks that are not compatible with the translation of these protocols to therapeutics manufacturing. Recent studies have focused on developing a defined biomatrix suitable for human ESC growth. For example, Ludwig et al [158] have demonstrated the use of a defined matrix composed of four different ECM proteins in conjunction with mTESR, thereby providing a defined system amenable for the scale up of ESC culture. Other examples of defined systems have been reported in recent years including Synthemax surface (Corning), Laminin-511 (Biolamina). Moreover, a defined surface incorporating a fusion protein composed of the extracellular domain of E-cadherin coupled with the IgG Fc domain has been demonstrated to maintain the self-renewal properties of ESCs and potentially replace Matrigel [161].

Therefore, several studies are being conducted in an attempt to recapitulate all the different aspects of the niche; this involve either the direct analysis of adult stem cell niches or deductive analyses based on ESC derivation and feeder-based techniques. The use of advanced high-throughput screening methods enables to determine the complex and intricate signaling pathways between stem cells and their native microenvironments. In turn, these studies can contribute towards the development of a completely defined, animal source free system that can be scaled up with relative ease, thus enabling the successful translation of laboratory protocols to clinical applications.

2.9.1 Media for stem cell culture processes

The composition of the culture medium is amongst the most critical factors for determining the outcome of a particular process. What is more, this variable is represents the most important manufacturing cost burden when large production scales are used (Chapter 4). As discussed, insights into cell niches provide guidance for designing an industrially performing medium. In general, stem cell culture media can be broadly considered of consisting of three components that must be considered to design a new

expansion medium: the basal, medium consisting of small molecular weight biological molecules (glucose, amino acids, lipids, vitamins) and inorganic (bulk salts, trace elements) compounds, protein factors (e.g, insulin, transferrin), and complex additives (e.g., serum, conditioned medium) [139]. Additionally, chemical components, especially ECM components such as fibronectin or collagen are provided in different forms, such as thin coating under specific conditions or as gels to support cell adhesion and growth [147].

The basal medium and defined protein components used for cell production process are generally the same as those devised for laboratory use. However, the inclusion of complex components, such as serum, serum substitutes, makes processes susceptible to parameter fluctuations that may result impact quality of the final product [158]. Therefore, switching to chemically defined medium and eliminating complex components represents an ideal solution that the industry will increasingly turn to [158]. Examples of serum free media developed for expanding ex vivo adult stem cells include StemPro MSC medium (Life technologies, Carlsbad, CA) for mesenchymal stem cells and the chemically defined X-VIVO medium (Lonza, Walkersville, MD) for hematopoietic cells. A number of chemically defined medium have also been developed for hESC culture including mTESR [158], HEScGro [162]. Small molecules, which confer specific biological activities, especially in inhibiting or activating signaling pathways involved in stem cell maintenance or differentiation, are increasingly being used during cultivation. For example, EHNA[erythro-9-(2-hydroxy-3-nonyl) adenine] is a small molecule used to support the maintenance of pluripotency markers in various hESC lines [163].

However, the difficulty in devising a robust defined medium for manufacturing should not to be underestimated. In general, a manufacturing process is even more susceptible to environmental fluctuations than laboratory-scale processes. Moreover, most chemically defined media are less resilient to counter environmental fluctuations than those supplemented with complex additives especially serum. Serum is thought to play a role of scavenger in buffering the effects of chemical assault that may arise from unforeseen factors, such as residual detergents or bleaching metal ions from corroded reactor parts.

Lessons learned from traditional cell culture processing can clearly be applied to stem cell bioprocessing at the expansion stage, bearing in mind that the final product will be the cells themselves and not biologics produced by them. Particularly, these cell products must be of high viability and quality. Traditional cell culture products undergo extensive downstream processing to attain a very high degree of purity after cell cultivation. These purification steps cannot be applied to the isolation of stem cell products. Therefore, it is critical at the medium design stage to minimize the quantity of any components that will absolutely need to be removed from the final cell product. Furthermore, extreme precautions must be taken to ensure that other contaminating molecules are not introduced during cultivation.

2.10 **Reactor Considerations in Stem Cell Bioprocesses**

In stem cell processes, robust systems with relatively large operating parameter windows and well defined change control procedures will be needed to meet the demand of clinical applications. In many cases, this will entail the employment of stirred tank systems ranging from tens to thousands of liters based on the needs of the process. A stirred tank bioreactor consists of a vessel fitted with impellers to provide a homogenous environment. Notably, it allows for the continuous monitoring and regulation of various chemical and physical parameters [5,138,139,164]. Cell sampling for monitoring the quality of the stem cells produced and their differentiation status can thus be readily performed.

Stem cells require surface adhesion for growth. This is in contrast to cells used in

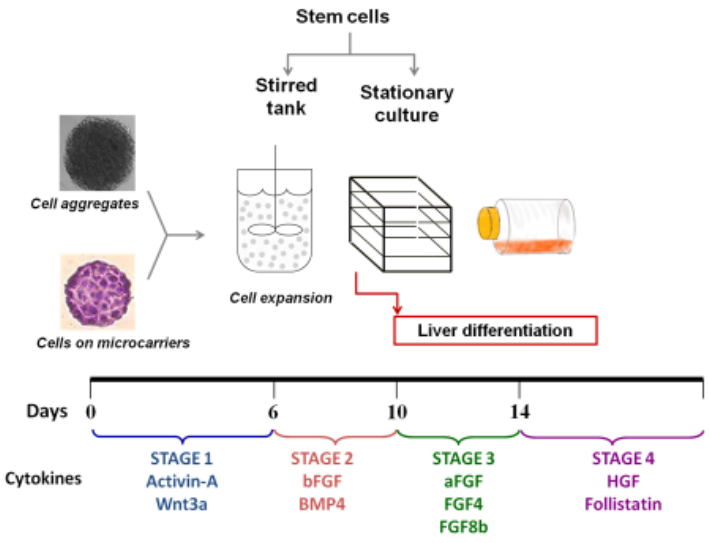


Figure 2-10: Different Strategies for Large Scale Cell Expansion and Differentiation [2]

recombinant protein production most of which have been adapted to grow in suspension. To cultivate anchorage dependent stem cells in a stirred tank, suitable surfaces must be provided. This is achieved by using microcarriers. Alternately, some stem cells may grow as aggregates, essentially by clumping to one another [165]. The different strategies for hESC expansion and subsequent differentiation to hepatocytes are illustrated in Figure 2-10.

Embryonic stem cells can be cultured as aggregates called embryoid bodies (EB) allowing for their cultivation in stirred bioreactors [165]. These aggregates can be formed from a suspension of single cell using different strategies, including forced aggregation methods where cells are suspended in low attachment culture dishes or hanging drop methods where cells are suspended in a drop format from the lids of tissue culture dishes [166], [167]. EB are formed from human ESCs and undergo a 15-fold expansion in a stirred tank system compared to a 4-fold expansion in a static culture over a period of 21 days [165]. These EBs maintain their potency and can be differentiated into hematopoietic progenitor cells [165]. In another study, homogenous EBs were formed, expanded in a stirred tank bioreactor and differentiated to cardiomyocytes [168]. Furthermore, recent reports of ESC single cell suspension survival with the Rho-kinase inhibitor [169] have led to the design of protocols yielding more homogeneous ESC cultures. This has been demonstrated in the formation of EBs that can be formed from single cell suspension of ESCs in the presence of this inhibitor with the resulting cells undergoing large scale expansion while retaining their properties when cultured in spinner flasks in the defined mTESR medium [170].

However, the aggregates, if allowed to overgrow, can result in necrotic cell clumps owing to the deprivation of oxygen. Necrotic centers developed in the human glioma spheroids with diameters of 600-800 μm and oxygen tension continued to increase with size [171]. In addition, since many applications require the use of dispersed cells rather than aggregates of millimeter size, dissociating the cells from these EBs after differentiation poses a challenge as the procedure may involve enzymatic treatment, which could damage cell integrity and function.

An alternative route to the EB technology is the use of microcarriers. Microcarriers are beads of about 150-200 micron in diameter with a surface compatible for cell adhesion. Microcarriers can be either microporous or solid; they are made from a variety of materials such as dextran (Cytodex, GE Healthcare, Waukesha, WI), gelatin (Cultispher, Percell Biolytica AB, Astorp, Sweden) or polystyrene (Solohill, Sigma, St Louis, MO). In addition, their surfaces can be modified through coating with extracellular matrix proteins like collagen, peptides or different types of charged residues to facilitate cell attachment [172,173]. The main advantage that microcarriers offer is that they offer high cell culture surface to reactor volume ratios [139].

Human ESCs can be cultured on matrigel-coated polystyrene beads and expanded by 30-40 fold within 8 days [174]. Cells grown using this process typically differentiate to the definitive endoderm with 80% efficiency [174]. A similar system involving human ESCs cultured on microcarriers was shown to be useful to expand and differentiate ESCs into cardiomyocytes[175]. Rat multipotent stem cells seeded on Cytodex-1 microcarriers can be expanded over 80-fold in a spinner flask system and differentiated towards the hepatic lineage [176]. Several other similar studies have been reviewed [164].

In the scale up of stirred tank bioreactors for stem cell culture, important lessons can be learnt from the accomplishments of the past 3 decades in traditional cell culture processing. Robust protocols for oxygen and pH control, fluid transfer, mixing control, mechanical stability and aseptic operations are all in place [139]. This prior knowledge can provide a huge database of relevant information for the design of similar systems for suspension-based stem cell bioprocesses.

2.11 **Concluding remarks**

The availability of an expandable source of human hepatocytes will have profound implications in the treatment of liver failure and congenital liver diseases. Sources of hepatocytes include hepatic progenitors from fetal or adult liver, differentiation of pluripotent stem cells or mesenchymal stem cells and direct reprogramming to fibroblasts. Hepatocytes like cells obtained with current differentiation protocols are still far away from their primary counterparts in their functional maturity. While these cells

may not be suitable for clinical transplantation currently, this may not limit their use in bio-artificial liver devices or as *in vitro* models for drug toxicity and diseases. This has been demonstrated in proof of concept studies in establishing liver disease models. Finally, prior expertise in cellular bioprocessing technology can be harnessed for the development of a bioprocess catering to the large scale production of stem cell derived hepatocytes. Thus, one can envision the translation of stem cell research to the clinic in the not too distant future.

3. MATERIALS AND METHODS

3.1 Human Embryonic Stem Cell Culture

The human embryonic stem cells (hESC) line H9 was cultured using medium consisting of 80% Dulbecco's modified Eagle medium (DMEM) (Gibco/BRL), 20% Knockout Serum Replacement (Gibco), 2mM glutamine, 0.1 mM nonessential amino acids, 0.1 mM β-mercaptoethanol and basic fibroblast growth factor (bFGF, R&D) (10 ng/ml). These cells were cultured on irradiated mouse embryonic fibroblasts (MEFs) obtained from E13-E14 CF-1 mice (Charles River Laboratories, Wilmington, MA), at 37°C in 10% CO_2. The cells were passaged on a regular basis every 2-3 days using 0.1% (w/v) Collagenase Type IV (Gibco) in DMEM.

3.2 Hepatocyte Differentiation

H9 cells were cultivated onto 12 well plates coated with 2% matrigel (BD biosciences) in mTESR medium until 50% confluency was reached. Differentiation was induced with addition of the differentiation medium. This medium consisted of a 60/40 (v/v) ratio of low glucose Dulbecco's Modified Eagle media (DMEM) (Gibco, USA) and MCDB-201 (Sigma). This was supplemented with 0.026 μg/ml ascorbic acid 3-phosphate (Sigma), linoleic acid bovine serum albumin (LA-BSA, Sigma), insulin-transferrin-selenium (ITS, Sigma) (2.5 μg/ml insulin, 1.38 μg/ml transferrin, 1.25 ng/ml sodium selenite), 0.4 μg/ml dexamethasone (Sigma), 4.3 μg/ml β-mercaptoethanol (Hyclone), 100 IU/ml penicillin and 100 μg/ml streptomycin (Gibco). Fetal bovine serum (2% (v/v) was added into the differentiation medium for the first six days and 0.5 % (v/v) for the remaining period.

Growth stages and cytokines specific for each differentiation stage were supplemented accordingly. Activin A (100 ng/ml) and Wnt3a (50 ng/ml) were used to initiate definitive endoderm formation. FGF2 (10 ng/ml) and BMP4 (50 ng/ml) were used to initiate hepatic specification. FGF8b (25 ng/ml), FGF1 (50 ng/ml) and FGF4 (10 ng/ml) were added to induce formation of hepatic progenitor cells formation and finally HGF (20 ng/ml) and Follistatin (100 ng/ml) were added for hepatocyte maturation.

3.3 Endodermal Cell Expansion and Differentiation

For endodermal cell expansion, cells were harvested as small clusters at the end of six days of the first stage of differentiation using 0.1% (w/v) Collagenase Type IV (Gibco) in DMEM. These cells were passaged onto matrigel coated plates in Stage 2 medium containing bFGF (10 ng/ml) and BMP4 (50 ng/ml). The cells were passaged every three days up to a four passages. To initiate differentiation, medium was switched to Stage 3 medium for 4 days and Stage 4 medium as described previously in Section 3.2.

3.4 Culture of Hepatic progenitor cells in Kubota's medium

Kubota's medium consisted of 500 ml RPMI 1640 (GIBCO # 11875-093), 0.5 g BSA (Sigma # A8806 fatty acid free), 270 mg Niacinamide (Sigma # N0636), 5 µg/ml Insulin (Sigma # I5500), 10 µg/ml Transferrin/Fe (Ado, bovine) (Sigma # T1283), 5 ml l-Glutamine 200 mM (2 mM, GIBCO # 25030-081), 5.0 ml Antibiotics (GIBCO # 15240-062, AAS), 10E-8 M Hydrocortisone (Sigma # H0888), 1.75 ml b-Mercaptoethanol (5E-5 M, Sigma # M6250), 10E-10 M Selenium (Aldrich # 22,982-7), 10E-10 M Zinc sulfate heptahydrate (10E-10 M, Specpure # JMC156) , 10 µg/ml High density lipoprotein from human plasma (Sigma #L8039-solution, Sigma #L1567-lyophilized powder)and 38 µl of the free fatty acids mixture described below. The free fatty acids were reconstituted as 1M solutions of Palmitoleic acid (MW 254.4, Sigma # P9417) , Oleic acid (MW 282.5, Sigma # O1008), Linoleic acid (free acid, not sodium salt), Linolenic acid (MW 278.4, Sigma # L2376) and a 151 mM solution of stearic acid (MW 284.5, Sigma # S4751). These were then combined using 31.0 µl of 1.0 M palmitic acid, 2.8 µl of 1.0 M palmitoleic acid, 76.9 µl of 151 mM stearic acid, 13.4 µl of 1.0 M oleic acid, 35.6 µl of 1.0 M linolenic acid, 5.6 µl of 1.0 M linolenic acid and 834.7 µl of ethanol.

Day 14 cells were harvested as small clusters and plated in Kubota's medium or Stage 4 medium on different extracellular matrices, matrigel (BD, #354230) or collagen 1 (BD, #354236)

3.5 Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

.

Total RNA was obtained from cell lysates using RNAeasy microkit (Qiagen) and cDNA was synthesized using the Superscript III reverse transcriptase kit (Invitrogen) according to manufacturer's instructions. The cDNA samples were mixed with SYBR Green Mix PCR reaction buffer (Applied Biosystems) and primers (5 μM working stocks). The qRT PCR reaction was carried out using the following program : 50°C for 2 min, 95°C for 10 min, and 40 cycles at 95°C for 15 sec and 60°C for 1 min followed by a dissociation protocol to obtain a melting curve, on a Realplex mastercycler (Eppendorf). Transcript abundance levels were measured relative to a housekeeping gene GAPDH and calculated as ΔCt (Ct(gene of interest)-Ct(GAPDH)). For comparison across samples, the transcript intensities were normalized to hESCs or any other reference sample and expressed as $\log_2$(Transcript levels relative to hESCs) and calculated as ΔΔCt or ΔCt (hESC)-ΔCt(day of sample).

3.6 Immunohistochemistry

The cells were fixed with 4% paraformaldehyde at room temperature for 20 min. The cells were blocked with PBS containing 0.2% Triton-X-100 (PBST) with 3% donkey serum at room temperature for 1 hr and incubated with primary antibodies AFP (Dako, 1:1000), SOX17 (R&D, 1:20), FOXA2 (Abcam, 1:1000) and DAPI (Life Technologies, 1:500) overnight at 4 ˚C. The cells were then incubated with secondary antibodies anti-mouse IgG1 A488 labeled (Molecular Probes, 1:500 dilution) and anti-rabbit IgG A488 (Molecular Probes, 1:500 dilution) for 30 minutes at room temperature prior to visualization. The negative controls were cells incubated with only the secondary antibody and relevant isotype control and were processed simultaneously.

3.7 Flow cytometry

The differentiating cells were washed with PBS and treated with 0.1% (w/v) collagenase type IV (Invitrogen) for 10 minutes at 37°C. The cells were scraped and harvested as small clumps. These cells were collected in a 15 ml tube and centrifuged. The cell pellet

was resuspended in a 0.05% (w/v) solution of trypsin supplemented at 2% (v/v) chicken serum at 37°C for an additional 5-10 min. A single cell suspension was then obtained by passing the cell suspension through a 40 um nylon filter. These cells were fixed with 4% paraformaldehyde for 15 minutes and washed twice with PBS. The cells were treated with a blocking solution consisting of SAP buffer (PBS with 0.1% (w/v) saponin and 0.05% (w/v) sodium azide) supplemented with 5% donkey serum for 1 hour. This was followed by incubation with primary antibodies used listed in Table 2. Isotype control antibody was used as negative control. The secondary antibody was anti-rabbit IgG A488 or A546 (Molecular Probes, 1:500 dilution). Antibody dilutions were optimized using HepG2 and hESCs as positive and negative controls. For cell surface markers, the blocking and antibody incubation was performed using PBS with 5 % serum instead of SAP buffer.

3.8 Mass cytometry

Lyophilized antibodies for FOXA2, SOX17, GATA4, GATA6, AFP, ALB, DLK1, A1AT and HNF4α proteins were obtained from R&D systems. About 200 µg of antibodies were conjugated with a selected panel of heavy metal isotopes using the MaxPar® antibody labeling kit from Fluidigm according to the manufacturer's instructions with slight modifications. The antibodies were eluted in 50 µl PBS instead of W buffer provided in the kit and this was found to improve antibody recovery. The conjugated antibodies were diluted to 0.5 mg/ml in antibody stabilizer solution (Candor Biosciences, 131050) and stored at 4 C until use. Metal conjugated CXCR4 and CD44 were directly obtained from DVS Sciences.

Cells were dissociated into single cells by treatment with 0.1% collagenase in DMEM and trypsin as previously described [177]. About 500,000 cells per time point were fixed using 10% formalin for 20 minutes at room temperature. The cells were washed with PBS, centrifuged and each cell pellet were suspended in 5 µL Human TruStain FcX™ (BioLegend, 422302) and 95 µL PBS at room temperature for 10 minutes for blocking. The cells were first incubated in a cocktail of metal conjugated antibodies targeting surface markers in 100 µl of PBS for 30 minutes at room temperature. Cells were washed

twice with PBS and incubated with a second cocktail of antibodies targeting intracellular proteins suspended in 100 µl SAP buffer (PBS with 0.1% (w/v) saponin (Sigma, 47036) and 0.05% (w/v) sodium azide (Sigma, 438456) for 30 minutes at room temperature. After washing the cells twice with SAP buffer, cells were incubated with MaxPar® Intercalator-Ir 125 µM (DVS Sciences, 201192A) at a dilution of 1:1000 in 1 ml of SAP buffer overnight at 4°C. Cells were washed twice, suspended in 500 µl water and passed through cell strainer and were run on the CyTOF2 instrument (DVS Sciences). During titrations, an antibody cocktail was made at 8 µg/ml and serially diluted twofold to achieve a wide range of concentrations. The negative control used was hESCs. It was critical to run a negative control for every titration as well as experiment in case of changes in antibody activity. A list of working concentrations are listed in the Supplementary table. Data was analyzed using the Cytobank software and visualized in Spotfire (Tibco).

3.9 **Functional Assays to Characterize HLCS**

Albumin secretion was measured using an ELISA kit (Starters Kit Bethyl E101 and Bethyl E80-129) following the manufacturer's instructions. The amount of albumin present in fresh medium was subtracted from albumin concentration in the samples to quantify the amount secreted by the cells over 24 hours. Cell numbers were estimated and albumin secretion is reported as pg/ cell/day.

Urea was measured using QuantiChrom Urea Assay Kit (BioAssay Systems) according to manufacturer's instructions. The concentration of urea was determined using a standard provided with the kit and reported on a per cell basis for the differentiated samples.

Periodic-Acid-Schiff staining (Sigma-Aldrich) was performed according to the manufacturer's protocol. The cells were immersed in Periodic acid solution for 5 minutes, in Schiff's reagent for 15 minutes at room temperature and were washed with water prior to visualization.

3.10 **Transcriptome Analysis**

Total RNA was extracted from human liver differentiation (H9) samples at various time points of endodermal cell expansion and differentiation using the RNeasy Mini kit (Qiagen). Samples were hybridized to the Illumina HT12 bead array v3 (Illumina Inc).
Data was processed using the *lumi* package in R [178]. Transcriptome data from 34,000 probes representing about 20,000 genes was obtained. Manually curated gene sets from (http://www.broadinstitute.org/gsea/index.jsp) containing genes belonging to different developmental stages such as pluripotent, endodermal and hepatic progenitor genes were used to visualize trends in the data. To gain a global perspective on the gene behavior, hierarchical clustering and principal component analysis (PCA) were performed in R. Pearson correlation analysis among the different time points of the original and modified protocols was performed to identify genes exhibiting opposite trends. Spotfire (Tibco), Database for Annotation, Visualization and Integrated Discovery (DAVID) and Ingenuity Pathway Analysis (IPA) (Qiagen) were software used for data visualization and functional analysis. Gene Set Analysis (GSEA), a technique used to interrogate gene sets that were statistically different between the EN and EN2 cell states was also implemented using the software from the Broad Institute (http://www.broadinstitute.org/gsea/index.jsp) [179,180]

Table 3-1: List of Human primers for qRT-PCR

Gene	Forward sequence	Reverse sequence
AAT	GTCAAGGACACCGAGGAAGA	TATTTCATCAGCAGCACCCA
AFP	AAATGCGTTTCTCGTTGCTT	GCCACAGGCCAATAGTTTGT
ALB	TGGCACAATGAAGTGGGTAA	CTGAGCAAAGGCAATCAACA
CXCR4	AACTTCAGTTTGTTGGCTGC	GAAACAGGGTTCCTTCATGG
CYP3A4	AAGTCGCCTCGAAGATACACA	AAGGAGAGAACACTGCTCGTG
CYP3A7	TGCTTTGTCCTTCCGTAAGGG	CAGCATAGGCTGTTGACAGTC
CYP7A1	CTGAGGCTTTCCAGTGCCT	AGGTAGTCTTTGTCTTCCCGT
EOMES	AACAACACCCAGATGATAGTC	TCATAGTTGTCTCTGAAGCCT
FOXA2	ATTGCTGGTCGTTTGTTGTG	TACGTGTTCATGCCGTTCAT
G6PC	GTGTCCGTGATCGCAGACC	GACGAGGTTGAGCCAGTCTC
GAPDH	GAGTCAACGGATTTGGTCGT	GACAAGCTTCCCGTTCTCAG
GSC	TCTCAACCAGCTGCACTGTC	CCAGACCTCCACTTTCTCCTC

KRT18	TGATGACACCAATATCACACGAC	TACCTCCACGGTCAACCCA
MGST	TTTGGAACACCATTCCAGAC	TTTCAAGGTCATTCAGGTGG
OCT4	CTTCGCAAGCCCTCATTTC	CCTTGGAAGCTTAGCCAGGT
PEPCK1	AAGAAGTGCTTTGCTCTCAG	CCTTAAATGACCTTGTGCGT
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
TTR	ATCCAAGTGTCCTCTGATGGT	GCCAAGTGCCTTCCAGTAAGA

Table 3-2: Antibodies for flow cytometry and immunohistochemistry

Name	Company	Catalog no.	Dilution	Type
DLK1	Livtech	DI-6	1:20	Primary
AFP	Dako	A0008	1:1000	Primary
ALB	Dako	A0001	1:1000	Primary
FOXA2	Abcam	ab40874	1:1000	Primary
SOX17	Abcam	ab84990	1:20	Primary
Goat IgG	Jackson Immunoresearch	005-000-003	Same as primary	Isotype
Rabbit IgG	Jackson Immunoresearch	011-000-003	Same as primary	Isotype
Donkey anti-goat Alexa 555 (red)	Invitrogen	A21432	1:500	Secondary
Goat anti-rabbit Alexa 555 (red)	Invitrogen	A21429	1:500	Secondary
Goat anti-rabbit Alexa 488 (green)	Invitrogen	A11008	1:500	Secondary
anti-mouse IgG1 A488	Molecular Probes	A-21121	1:500	Secondary
Hoechst (nuclear staining, blue)	Sigma	33258	1:500	Secondary

Table 3-3: Antibodies for mass cytometry

Antigen	Vendor	Catalog number	Metal Conjugated	Concentration
SOX17	R&D	AF1924	Sm154	4 µg/ml
FOXA2	R&D	AF2400	Gd156	4 µg/ml
AFP	R&D	AF1369	Nd143	4 µg/ml
DLK1	R&D	MAB1144	Tm169	4 µg/ml
ALB	R&D	MAB1455	Er166	4 µg/ml
A1AT	R&D	AF1268	Tb159	4 µg/ml
GATA6	R&D	AF1700	Eu151	4 µg/ml
Gata4	R&D	AF2606	Dy162	4 µg/ml
CXCR4	DVS	3175001B	Lu175	2 µl

CD44	DVS	3150018B	Nd150	2 µl

4. ENDODERM CELL EXPANSION TO FACILITATE HIGHER YIELDS OF STEM CELL DERIVED HEPATOCYTES

4.1 Introduction

The derivation of hepatocytes from embryonic and induced pluripotent stem cells, have the potential of replacing primary hepatocytes in cellular therapies for treatment of liver failure. Pluripotent stem cells, due to their self-renewal abilities, can provide a continuous source of hepatocytes upon differentiation, thereby circumventing issues of donor shortages, which are prevalent in applications relying on cells from primary sources. Stem cell derived hepatocytes also have applications in bio-artificial liver devices, disease modeling and drug toxicity screening.[6,11,181].

Understanding the molecular signals guiding liver development has aided in the development of *in vitro* strategies to differentiate stem cells towards the hepatic lineage by several research groups [122,123,177,182,183]. The resulting hepatocyte-like cells (HLCs) at the end of differentiation begin to express desired hepatocyte transcripts and functions and have been extensively characterized in previous reports [78,182]. With the promising advances in differentiating stem cells to hepatocytes, one major challenge to overcome prior to clinical translation is the current low cell yields [5]. Any therapeutic application will require 10^9-10^{10} cells per treatment. Production of large quantities of cells to meet clinical demands currently entails expanding large quantities of pluripotent stem cells followed by differentiating them. The conventional method of expanding and differentiating stem cells remain labor intensive and prone to batch variation. The process from thawing to cell expansion and differentiation is also time consuming. Expansion of an intermediate cell state during stem cell differentiation will circumvent this issue by reducing the reliability on large quantities of the initial starting pool of stem cells. Furthermore, it is well known that cells undergo extensive proliferation throughout the liver development process. With the appropriate signaling cues mentioned before, it may be possible to replicate this phenomenon in *in-vitro* differentiations.

To address these issues, some reports have been published on the proliferation of intermediate cells during stem cell differentiation. Hepatic progenitor cells were enriched

using the surface markers EPCAM or N-cadherin and expanded on stromal feeders [127,184,185]. Recently, a self-renewing endodermal cell line was reported where the cells were sorted based on expression of CXCR4 and CD117 and expanded on mouse embryonic feeders [111]. In the current study, we demonstrated the expansion of an early stage of endodermal intermediates without the need for cell sorting or relying on feeder cells to achieve higher yields of HLCs after subsequent differentiation of these endodermal intermediates.

4.2 Results

4.2.1 Expansion of endodermal cells

In this study, we demonstrated that during the course of hepatic differentiation from human embryonic stem cells, an intermediate endodermal cell could be expanded and continued to differentiate to a HLC phenotype. The hESCs were differentiated to definitive endoderm in Stage 1 medium containing Activin (100ng/ml) and Wnt3a (50ng/ml) to reach cell densities of $2.5x10^5$ cells/cm^2 in six days. The cells were detached by collagenase treatment and passaged, at 8 $x10^4$ cells/cm^2 onto Matrigel coated plates in Stage 2 medium containing the growth factors FGF2 (10ng/ml) and BMP4 (50ng/ml). The morphology of the cells immediately after plating and after three days in Stage 2 conditions are shown in Figure 4-3. After 3 days of expansion, cells were passaged again under the cues of FGF2 and BMP4, which are known to provide proliferative cues during embryonic liver development [71]. The cell population expanded approximately 8 fold after two passages as shown in Figure 4-2 (n=4). The passaged populations will herein be referred to as EN1 and EN2, respective of their passage number. Continued passaging could result in cell expansion up to 15 fold, however the growth rate began to slow down and fibroblastic looking cells began to overtake the native population. In contrast, the cells only expand two times from the endodermal to the hepatic stages during a regular differentiation protocol as seen in Figure 4-4.

4.2.2 Expression of hepatic genes and proteins in expanded endodermal cells

The transcript levels of gene markers for hepatic differentiation were assayed by qRT-PCR results are reported as fold changes with respect to undifferentiated hESCs

(n=4). Octamer-binding transcription factor 4 (OCT4), a master regulator of the pluripotency network in hESCs [186], decreased about 1000 fold during endodermal expansion as expected (Figure 4-5). This may be the result of preferential attachment of epithelial fated cells over undifferentiated hESCs upon cell passaging.

The endodermal transcription factors, goosecoid homeobox (GSC), a key marker for differentiating definitive and visceral endoderm that drives extraembryonic cell fate [187], and CXCR4, a surface marker co-expressed with GSC [187], were both highly expressed in the EN population and decreased by 1000 fold and 10 fold, respectively in EN1 and EN2 (Figure 4-5). The expression profile suggests the transitioning of the endodermal-committed cells towards a hepatoblast lineage. This is further confirmed by a 1000 fold increase of alpha-fetoprotein (AFP) and albumin (ALB) over the same period [188].

Transcript levels were also compared with cells from two different differentiation stages using the original protocol; 1) day 10 (HPC) where cells were exposed to BMP4 and FGF2 signaling for four days, 2) day 14 (HPC1), after 4 days of further treatment with Stage 3 growth factors. Fold changes of transcript levels for EN, EN1, EN2, HPC and HPC1 cells were reported with respect to hESCs (

Figure 4-6). The gene expression of samples from each protocol is represented as two continuous curves, with respect to the corresponding day of differentiation. The expression of OCT4, and endoderm markers GSC, CXCR4 decrease while hepatic AFP and ALB increase to higher extent in EN1 and EN2 cells when compared to both HPC and HPC1 cells. This indicates that the EN2 cells appear more differentiated than the corresponding cells (HPC1) from the conventional protocol. Notably, this increase in hepatic characteristics is simultaneously occurring with an 8-fold increase in cell number indicating that a selective proliferation of the desired cell type is occurring instead of proliferation of contaminating cell types. The transition from the endodermal stage to hepatic stage is also seen by immunostaining (Figure 4-7). Forkhead Box A2 (FOXA2) and SRY (Sex Determining Region Y)-Box 17 (SOX17), key transcription factors in the establishment of definitive endoderm are prominent in the EN stage but not the hepatic

marker, AFP. In contrast, a decrease in the expression of FOXA2 and SOX17 and increase in AFP expression is seen in EN1 and EN2.

We next examined whether this endoderm-hepatic transition was restricted to a subpopulation or if it occurs in the entire population. Mass cytometry was employed to examine co-expression of a panel of endodermal and hepatic markers. This method allows for multiple markers of cell fate transition to be assayed at single cell resolution [189]. Time of flight analysis was performed on cells labeled with heavy metal conjugated antibodies. The co-expression of CXCR4, FOXA2, SOX17, ALB, DLK1, AFP, ALB, AIAT and CD44 proteins was evaluated in differentiating cells. Only four markers are shown in each graph, thus multiple figures are used to present the progression of different markers in the different stages (Figure 4-8, Figure 4-9). In Figure 2c, SOX17, CXCR4 and ALB are represented on the three axes while CD44 is represented as the fourth dimension, with red indicating high and green indicating low protein expression. CD44, which plays an important role in HGF signaling through its receptor c-MET expressed in hepatocytes [190]. The data shown represents about 150,000 cells of the endodermal cells in different stages of expansion. In Figure 2D, DLK1, CXCR4 and ALB are plotted on the three axes with CD44 represented through color. Similar to the gene expression results, cells display reduction of SOX17, CXCR4 proteins and gain in DLK1, ALB, AFP, CD44 proteins with endodermal expansion. Also, this clarifies that majority of the cells co-differentiate from an endodermal to a hepatic fate i.e. all the cells simultaneously acquire hepatic and reduce endodermal expression. This reaffirms that the gain of hepatic phenotype occurs uniformly and is not a random event occurring with endodermal cell expansion.

4.2.3 Differentiation potential and functional activity of expanded endodermal cells

The differentiation capacity of the expanded cells (EN1, EN2) to the hepatic lineage was examined by switching the expansion medium to stage 3 medium and continuing to stage 4 medium. Transcript levels for Phosphoenolpyruvate Carboxykinase (PEPCK), a mature hepatic marker that participates in gluconeogenesis [191] was measured along with alpha-1 antitrypsin (AAT), an important hepatic glycoprotein

functioning as a serine protease inhibitor [192], AFP, a blood plasma protein produced by the liver during fetal development and ALB, a blood plasma protein produced by the adult liver. The resulting HLCs from expanded endodermal cells (EN2-HLCs) expressed the hepatic transcripts for PEPCK, and AAT at levels similar to HLCs derived using the original protocol (Figure 4-10). However, the expression for the AFP and ALB transcripts was higher for EN2-HLCs compared to HLCs, which could be a result of differentiating a more homogenous population of expanded endodermal cells.

Cytochrome P450 enzymes play a key role in drug detoxification in the liver and are often used as markers of hepatocyte maturity. Drug exposure often induces CYP450 enzymes levels resulting in accelerated transformation of the drug to allow for rapid clearance [193]. To evaluate if the induction could be recapitulated *in-vitro*, CYP2C8/2C9/2A6 transcript levels in HLCs were monitored after exposure to rifampicin [194]. The levels were reported as $\log_2$ expression relative to uninduced cells as shown in Figure 4-11. CYP2C8 and CYP2A6 transcripts were induced similarly in HLCs and EN2-HLCs, upon rifampicin stimulation while CYP2C9 was induced to a lower extent in EN2-HLCs.

The rate of albumin secretion and urea synthesis of the EN2-HLC cells were compared to HLCs generated from the original protocol (Figure 3c). EN2-HLCs secreted albumin at a higher rate (4.2 pg/cell/day) than HLCs (1.6 pg/cell/day). Both are still within the same range as our previous reports [177]. Urea production was comparable in both HLCs and EN2-HLCs at approximately 15 pg/cell/day. Glycogen synthesis (red) was also seen using PAS staining in both HLC and EN2-HLC cells (Figure 4-13). The results demonstrate that HLCs and EN2-HLCs have similar functional and phenotypic characteristics.

4.2.4 Comparative Transcriptome analysis of cells from original and modified protocols

4.2.4.1 Analysis of cell lineage subsets to confirm hepatic trends with endodermal expansion

Transcriptome analysus was performed on the cells at different stages of expansion and differentiation. First of all, characteristic gene markers for pluripotent, endodermal, hepatoblast and hepatic cells were used to evaluate the hepatic transition of

the expanded endodermal cells and their differentiation, in reference to the original protocol. The comparison is presented as a plot with two time curves corresponding to the gene expression dynamics in each of the protocols. The first curve represents the gene dynamics of EN, EN1, EN2 and EN-HLC while the other curve represents EN, HPC, HPC1 and HLCs. The x-axis represents the corresponding day during differentiation. (Figure 4-14). Pluripotent genes NANOG, SOX2, ZIC3, SALL4, and LIN28, which all participate in regulating the self-renewal and pluripotency of hESCs [195] decrease in expression to a greater extent with endodermal expansion as we had observed previously. The endodermal genes EOMES, CXCR4, GSC, and FOXA2 also decreased as the endodermal cells are further passaged. This was accompanied by a greater increase in hepatic progenitor genes such as AFP, DLK, TTR, and KRT8 in addition to the increase in mature hepatic genes including PEPCK or PCK2, transferrin (TF), apolipoproteinA (APOA1) as observed earlier.

4.2.4.2 Analysis of genes participating in key hepatic functions to confirm similarity in HLCs from both protocols

Genes pertaining to important liver functions such as glycolysis/gluconeogenesis, urea cycle, fatty acid metabolism, and detoxification were evaluated for HLCs and EN2-HLCs and were found to be similar. The transcript intensities for HLCs (blue) and EN-HLCs (red) are shown in a bar chart in Figure 4-16. Notably, PEPCK (PCK2), FBP1, ALDOB, ENO1/2/3, PGAM1 are important liver metabolic isozymes that participate in glycolysis or catalyze interconversion of glucose to glycogen. Urea cycle and amino acid metabolism genes including arginase (ARG), carbamoyl phosphate synthetase I (CPS), ornithine aminotransferase (OAT), Aminoacylase 1 (ACY1), Guanidinoacetate N-Methyltransferase (GAMT) were also evaluated. Liver specific cholesterol acyltransferase ACAT2 [196], straight-chain acyl-CoA oxidase (ACOX), carnitine palmitoyltransferase 1a (CPT1a), CYP4A11 participate in fatty acid metabolism [197] while CYP1A1/1B1/2B6/2C8/3A4/3A7 are involved in xenobiotic biotransformation [193]. The transcript levels are very similar for all the comparisons, affirming the

endodermal cells that underwent 8 fold expansion could be successfully differentiated to HLCs similar to the original protocol.

4.2.4.3 Global analysis

To evaluate the transcriptome on a global basis, invariant genes were removed and hierarchical clustering was performed. This showed that the expanded endodermal cells (EN1 and EN2) cluster separately from the EN cells and more closely with cells from later stages of differentiation. The HLCs and EN2-HLCs however cluster together.

From, principal component analysis (PCA) commonly used to reduce the dimensionality of the data, two principal components (PCs) were found to represent 90% of the variance. Therefore, these two PCs were used to visualize the trends in the data while retaining its inherent variability. The samples can be represented by their corresponding PCs and plotted in the two-dimensional PC space, where PC1 and PC2 are represented on the y and x-axis respectively. As seen in Figure 4-15, the expanded endodermal cells follow a similar differentiation trajectory towards hepatocyte-like cells as unexpanded endodermal cells. Furthermore, while the EN1 and EN2 cells are more differentiated than the EN cells, the resulting HLCs are very similar.

Pearson correlation, a measure of linear correlation was computed for differentiation time points of both protocols to identify any genes with contrasting trends. A value of 1 and -1 imply a strongly positive and negative correlation, respectively. The distribution of the Pearson coefficient is shown in Figure 4-17. The genes of interest are those that have a strong negative correlation of -0.8 to -1. As expected, only 12 genes (Table 4-1) follow this trend while most genes are positively correlated indicating they follow the same trends in both protocols.

Finally, genes that were differentially expressed by fivefold between the endpoint HLCs and EN-HLCs are shortlisted in Table 4-3. About 40 genes out of the 20,000 were identified and used to perform functional analysis in DAVID. The only functional cluster that had an enrichment score greater than 2 was response to wounding which included genes SERPINC1, SERPIND1 and MSTI. The other genes belonged to various gene ontology categories such response to extracellular stimulus, ketone metabolic process,

response to stress and regulation of endocytosis. Key hepatic genes were not differentially expressed confirming the similarity of HLCs from both protocols.

4.3 Discussion

The first major differentiation event to occur *in-vivo* involves the formation of the endodermal germ layer. During gastrulation, signaling through Nodal, a member of the TGFβ family, establishes the distinction between endoderm and mesoderm [63,198]. The endoderm is further patterned into the foregut, midgut and hindgut regions [4]. The liver along with the lung, pancreas, and stomach are derived from the foregut section around embryonic day (E) 9 of gestation [4,109]. The Nodal protein activates several specific transcription factors including FOXA2 and SOX17 that help regulate cell fate commitment to different endoderm cell lineages [63]. Signaling cues from the adjacent mesoderm layer are responsible for inducing liver development to enable hepatic specification through the fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) signaling pathways. Hepatic progenitor cells will arise and proliferate significantly to form the liver bud. Continued signaling via growth factors such as hepatocyte growth factor (HGF), FGF, and BMP from the surrounding mesenchyme region will aid in promoting the liver bud development. Soon, bipotential hepatic progenitor cells will arise to form either hepatocytes or cholangiocytes which will line the lumen of the intrahepatic bile ducts. Oncostatin-M (OSM) secreted by the hematopoietic cells along with HGF will be necessary for inducing further maturation towards hepatocytes. These signals will aid in increasing the activity of liver enriched transcription factors, such as C/EBPα and HNF4α which will be necessary in binding to promoters of key genes in liver maturation [4,6,7,109].

Differentiation protocols typically mimic the key signaling cues occurring during embryonic development. However, cellular proliferation, which is integral especially during *in vivo* development, is typically not integrated during *in vitro* differentiations. In this study, we demonstrated that our hepatic differentiation protocol could be modified in order to generate larger quantities of HLCs. This was implemented during the intermediate endodermal stage by simply allowing more surface area for ENs cells to

propagate showing up to at least 8-fold increase in number. The expansion process utilized signaling cues from the FGF and BMP signaling pathway to mimic the cardiogenic mesoderm and septum mesenchyme signaling. These factors play a key role in the transition of the foregut endoderm to the liver bud which is largely dependent on cell migration, proliferation and differentiation [109].

The subsequent passages in the expansion medium were accompanied with a simultaneous 1000 fold increase in expression of hepatic markers and 1000 fold decrease in endodermal and pluripotent markers. The EN2 cells underwent the highest loss in pluripotent and endodermal expression while gaining the most hepatic expression when compared to HPC1 cells, which have been exposed to fibroblast growth factors present in Stage 3 medium. During liver development, extensive crosstalk amongst signaling pathways occurs and therefore cells may embark on varied trajectories to attain the same differentiated state [6]. Therefore, the endodermal expansion coupled with FGF2 and BMP4 signaling may extend the effect of other FGF signaling mechanisms used in achieving HPC1 cells. The continued differentiation that occurs during expansion can help reduce the contamination of undifferentiated stem cells in the subsequent progeny. However, the capability for proliferation is not unlimited as these cells transition into a fibroblastic phenotype and initiate apoptosis, upon continued culture beyond three passages. This growth restriction is expected as cells are not being propagated as progenitor cells, but are allowed to mimic liver development and differentiate through exposure to different signaling cues.

The question of whether this endoderm to hepatic transition was occurring in a subpopulation of cells or if they were a bulk population event was answered using mass cytometry. Mass cytometry was developed as multiparametric single cell analysis tool for applications in cancer, hematopoietic and immune cell biology [189,199,200]. This is the first time, to our knowledge, that this technique has been applied to stem cell directed hepatic differentiations. This technique relies on inductively coupled plasma mass spectrometry to accurately identify the concentration of the heavy metal tags on a single cell and circumvents spectral overlap issues common to flow cytometry [189]. Mass cytometry allowed for multiplexing endodermal and hepatic markers to visualize the cell

fate transition on a per cell basis. This allowed probing into whether genes are activated individually or in a consorted manner and confirmed that hepatic genes are simultaneously activated along with endodermal gene repression for a successful differentiation outcome. Furthermore, this trend was a population-based event and not restricted to a specific sub population (Fig. 2C and D).

Functional activity was then assessed to compare HLCs and EN2-HLCs to conclude that propagation of the intermediate endodermal population did not compromise the quality of hepatic differentiation. HLCs generated from both protocols exhibited similar hepatic transcripts as well as secretory capabilities for albumin and urea. The HLCs generated after incorporating endodermal expansion exhibit similar properties to HLCs and HLCs cultured in the original protocol for 26 days especially as it relates to hepatocyte gene expression and albumin secretion, also reported previously [177]. This indicates that prolonged culture period does not affect cell quality. The end stage HLCs were also similar in expression of characteristic hepatic genes involved in key functions such as urea cycle, glycolysis/gluconeogenesis and xenobiotic biotransformation.

Transcriptome analysis provided further insights in to the behavior of the cells from both protocols. Unsupervised statistical techniques, of hierarchical clustering and PCA, revealed that though the paths along which the differentiation progressed were different, the end goal achieved was the same. Correlation and differential expression analysis demonstrated that none of the key hepatic genes were found to be differentially expressed between both HLCs. Gene set enrichment analysis was used to determine if any key pathways or signaling mechanisms was responsible for the difference in phenotype [179]. Wnt signaling and TGFβ signaling through Nodal which are key in the patterning of the foregut endoderm were upregulated in the EN cells. P53, PPAR signaling, and other characteristic hepatic functions such as complement and coagulation and drug metabolism were upregulated in the EN2 cells Table 4-2. Interestingly, in spite of the advanced differentiated status of the EN2 cells, the EN2-HLC cells are similar to the original HLCs. In most stem cell differentiations, the progeny is not as mature as their primary counterparts, indicating that a roadblock may need to overcome to achieve further maturation. This study reiterates this phenomenon, wherein the initial cues

promote differentiation but only to an extent permissive to create hepatoblasts. A comprehensive comparative analysis between *in-vitro* and *in-vivo* liver development may shed light on further navigating this blockade.

Overall, this modified protocol allows for generation of at least 8 fold higher quantities of HLCs without relying on a higher starting ESC population. The generation of higher quantities of hepatocyte like cells is critical for thorough *in-vitro* characterization and the translation of research protocols to clinical therapies. Typically, end stage HLCs face growth limitations owing to their differentiated state. Some laboratories have reported expansion of progenitor cells at different stages of differentiation [111,127]. However, these methods entail cell sorting based on surface markers [127] , the use of feeders [111] or transfection with SV40 large T antigen [201] to achieve higher cell yields. In Cheng et al., CXCR4+/CD117+ cells were sorted from a bulk population to obtain endodermal progenitor cells that were kept in culture in a complex media on a combination of Matrigel and feeder cells. These may cause unwarranted changes due to xenogeneic factors in affecting cell quality that may hinder clinical translation. The approach of endodermal cell expansion using typical differentiation culture conditions achieves higher cell yields with comparable cell quality to the original protocol.

4.4 Figures

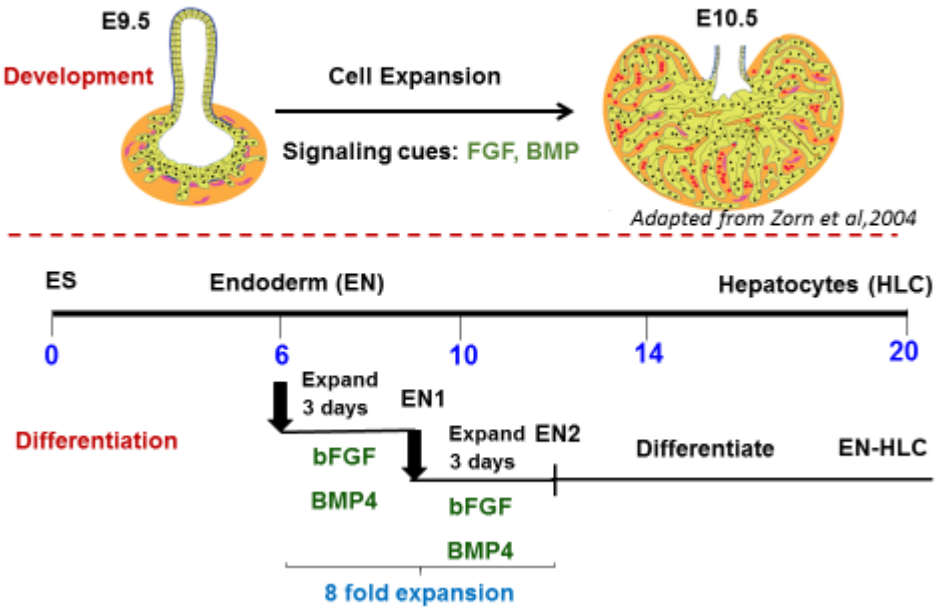


Figure 4-1: Protocol for Expansion of Hepatic Endodermal Cells. A) Cells were differentiated toward the endodermal cells and subsequently expanded up to 8 folds in 6 days using a combination of bFGF and BMP4. The expanded hepatic endodermal cells were then further differentiated towards hepatocyte using the existing differentiation protocol.

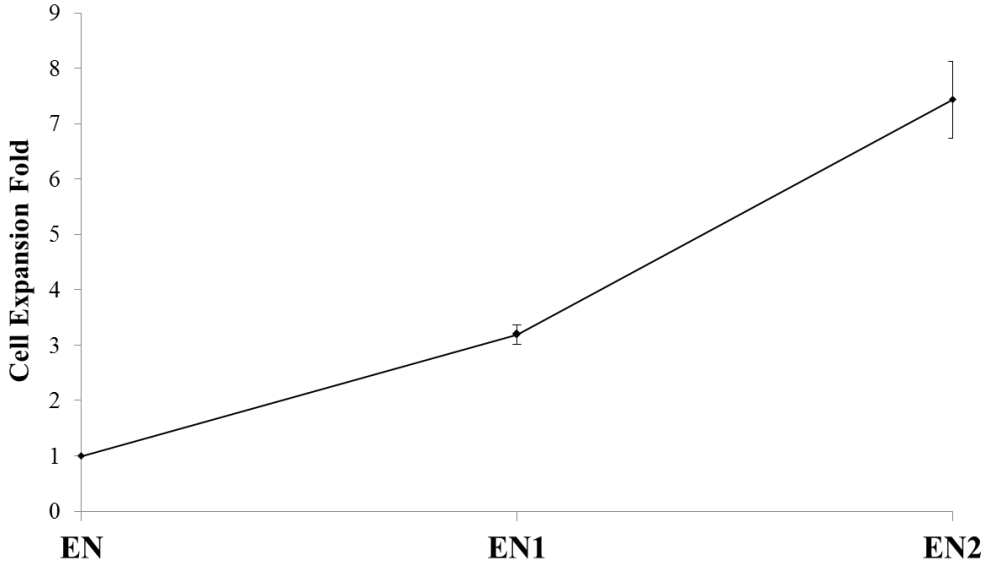


Figure 4-2: Expansion of cells showed a 8-fold increase during the course of 6 days (n=4)

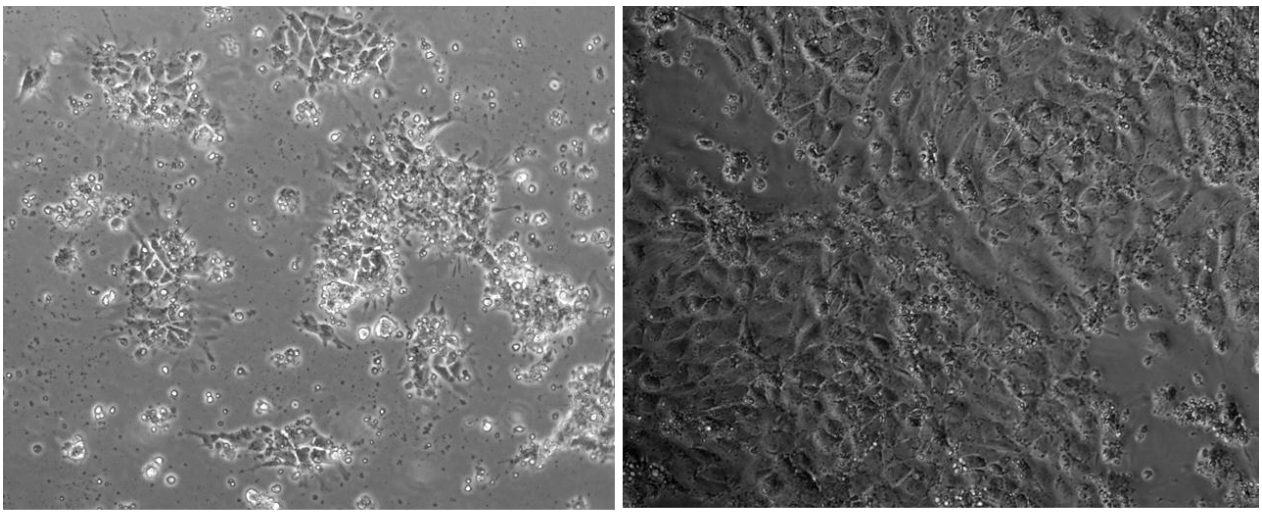


Figure 4-3: Morphology of cells A) immediately after plating and B) after three days in Stage 2 conditions is shown

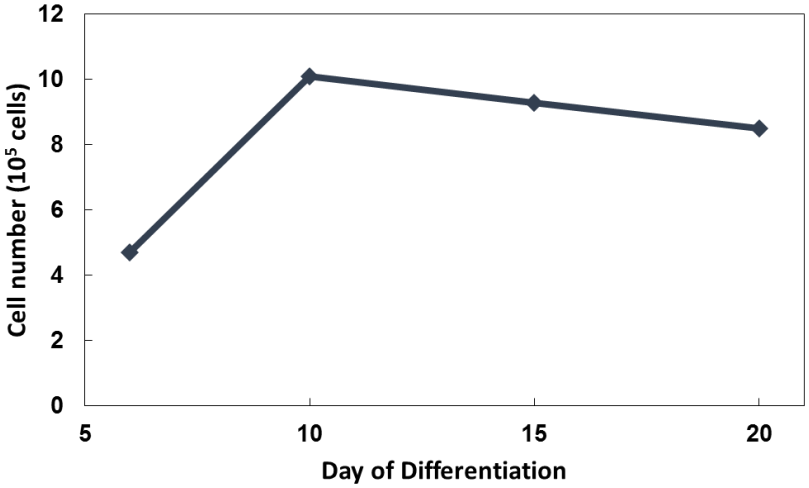


Figure 4-4: The increase in cell number during regular differentiation protocol is shown

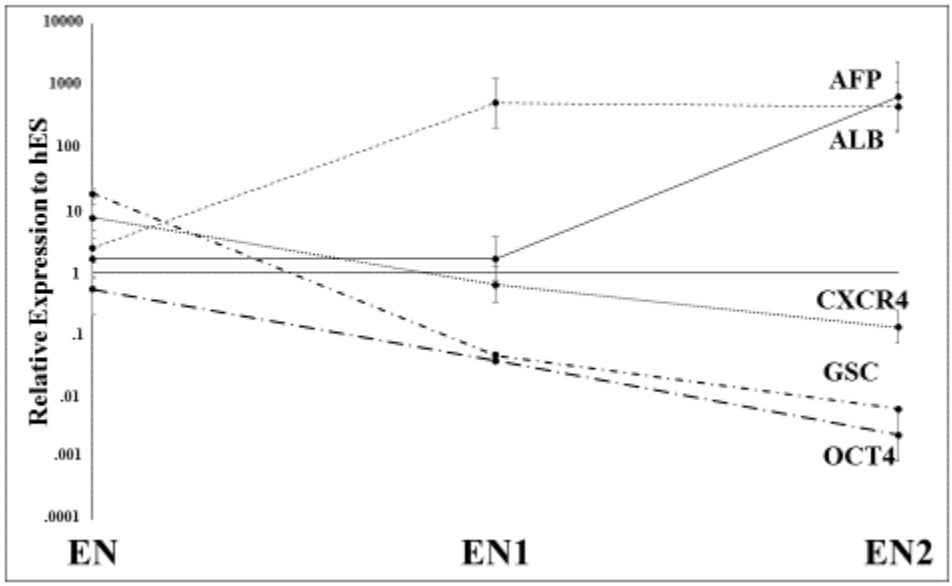
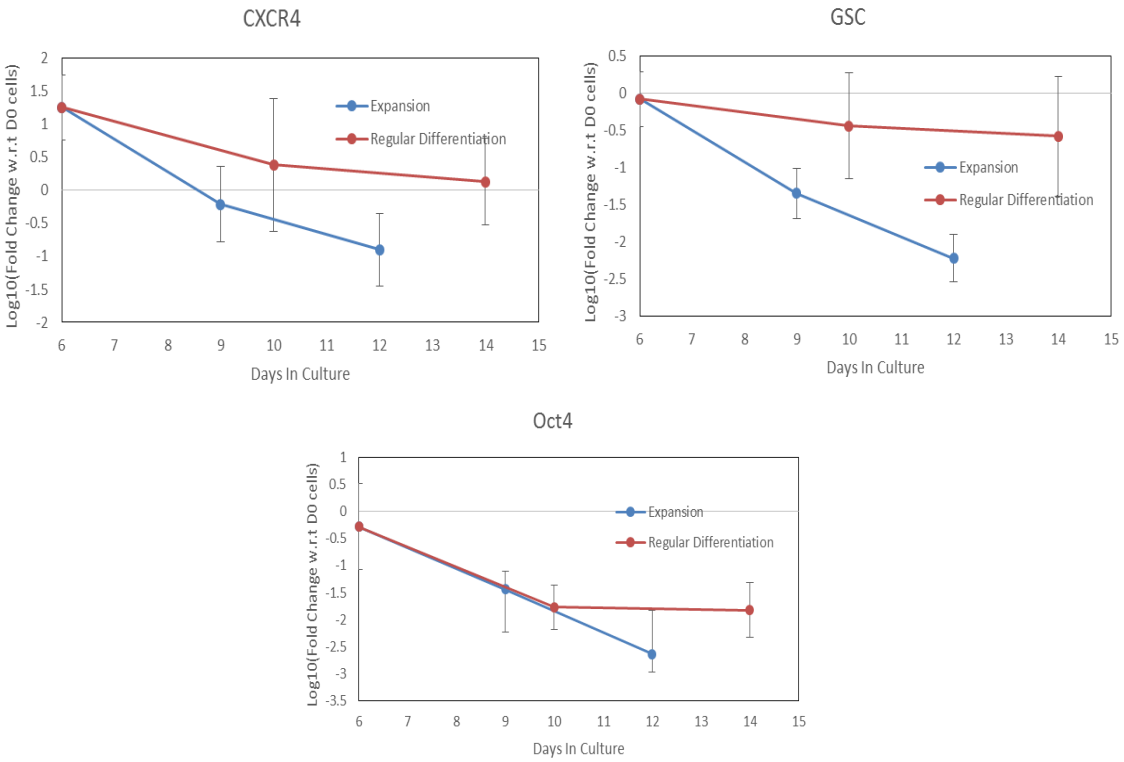


Figure 4-5: Phenotype of endodermal cells undergoing expansion. A) Endodermal cells showed a slight increase in hepatic markers during the expansion process while losing expression of pluripotent and endoderm markers (n=4)


CXCR4
Log10(Fold Change w.r.t D0 cells)
Days In Culture
Expansion
Regular Differentiation
GSC
Log10(Fold Change w.r.t D0 cells)
Days In Culture
Expansion
Regular Differentiation
Oct4
Log10(Fold Change w.r.t D0 cells)
Days In Culture
Expansion
Regular Differentiation

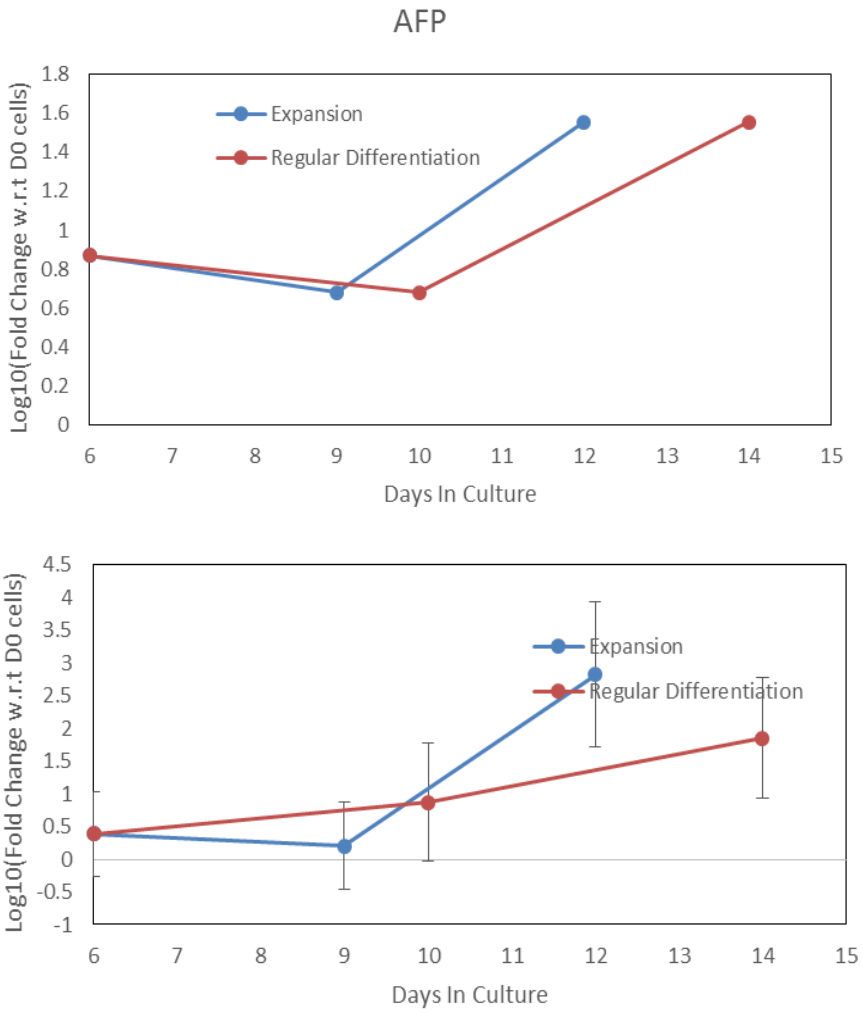


Figure 4-6: Comparison of gene expression in the original and endoderm expansion protocol are shown with respect to the corresponding day of differentiation

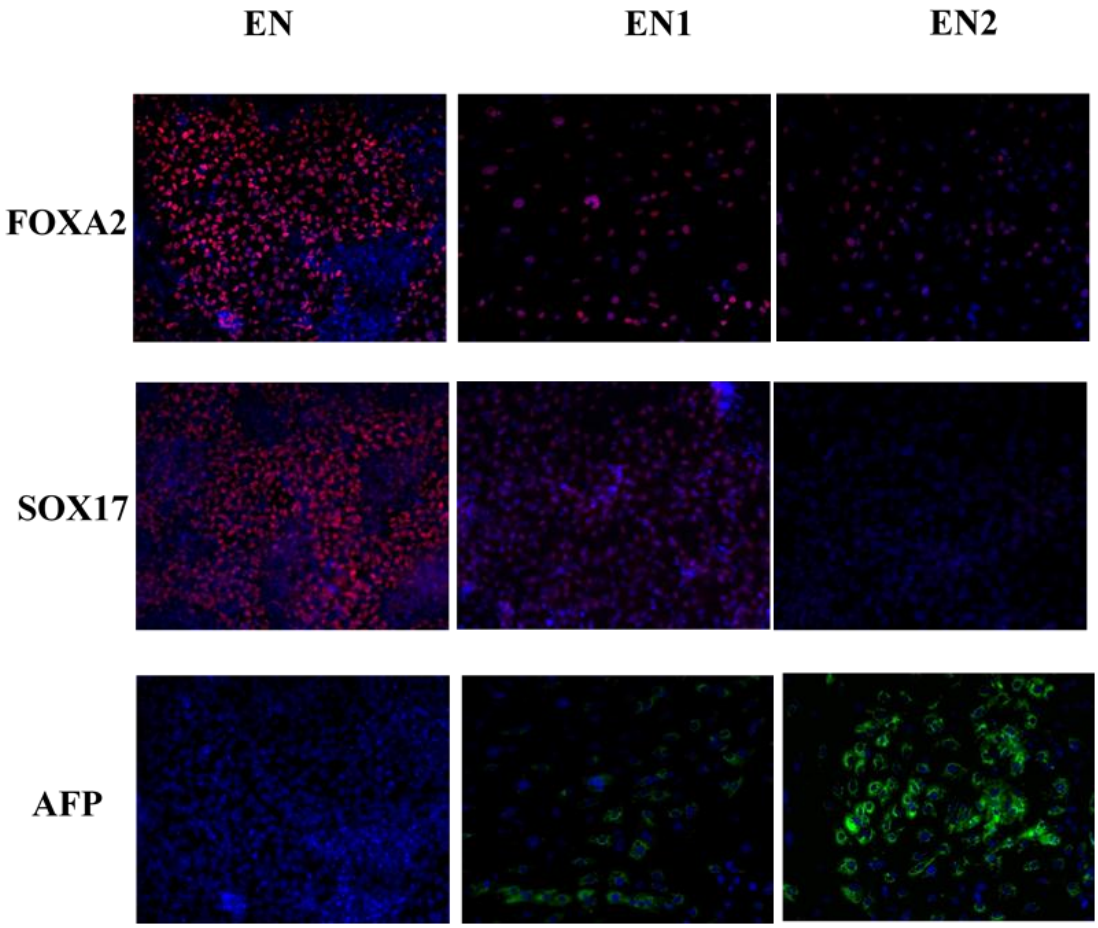


Figure 4-7: Immunostaining reveal similar behavior of endodermal cells undergoing differentiation. Cells expressed endodermal markers (FOXA2 and SOX17) during endoderm stages but begin to lose expression as it underwent through expansion. Instead, cells begin to gain expression in AFP while maintaining FOXA2 illustrating a more hepatic endodermal phenotype

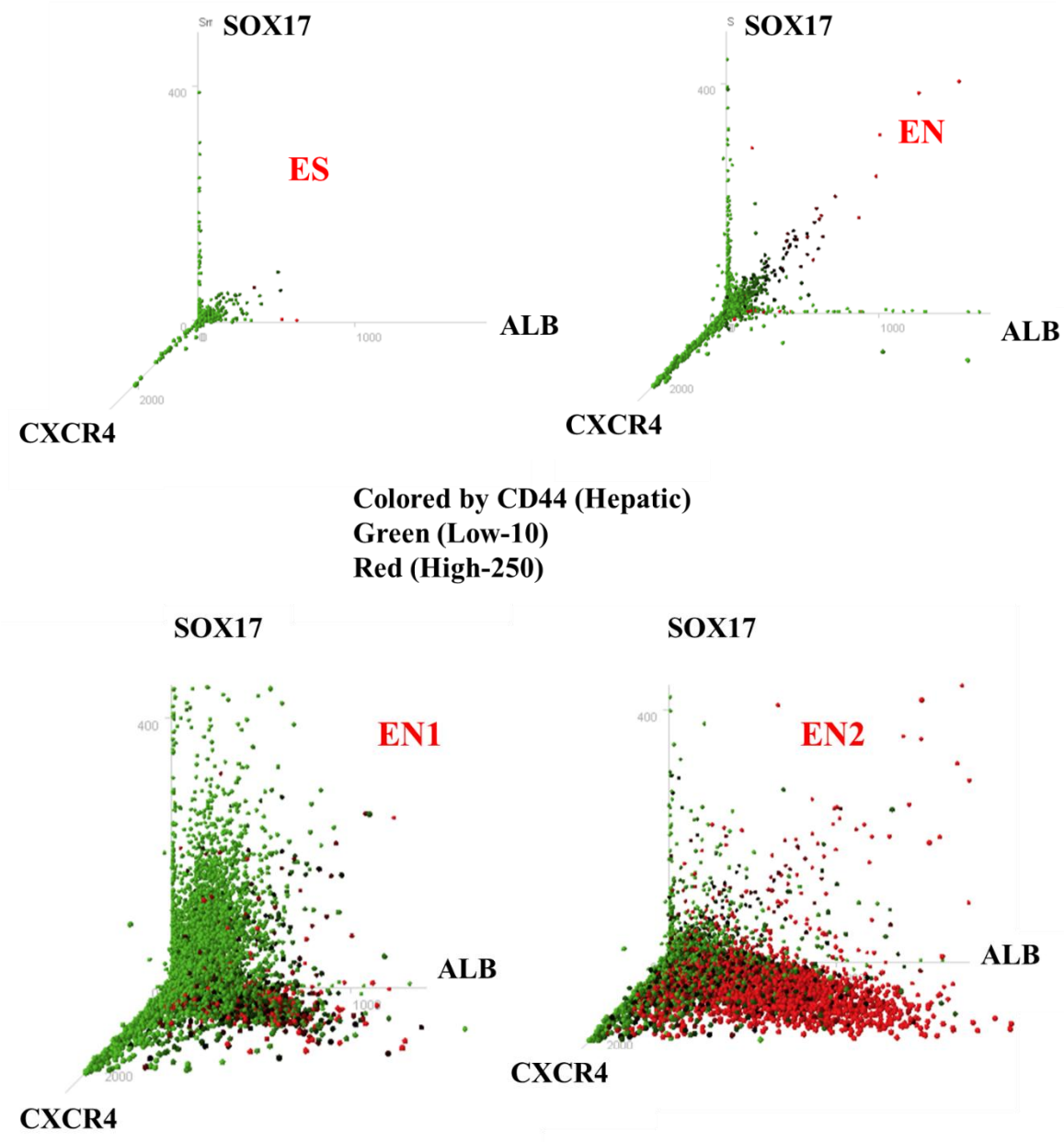


Figure 4-8: Mass Cytometry data for hESCs,EN,EN1,EN2 cells is shown where the x, y and z axes represent ALB, SOX17 and CXCR4 and the color represents CD44 expression. Green indicates low and red indicates high expression. hESCs are shown as the negative control. The endodermal genes decrease in most cells with increase in hepatic gene expression in EN1 and EN2 cells when compared to the EN cells.

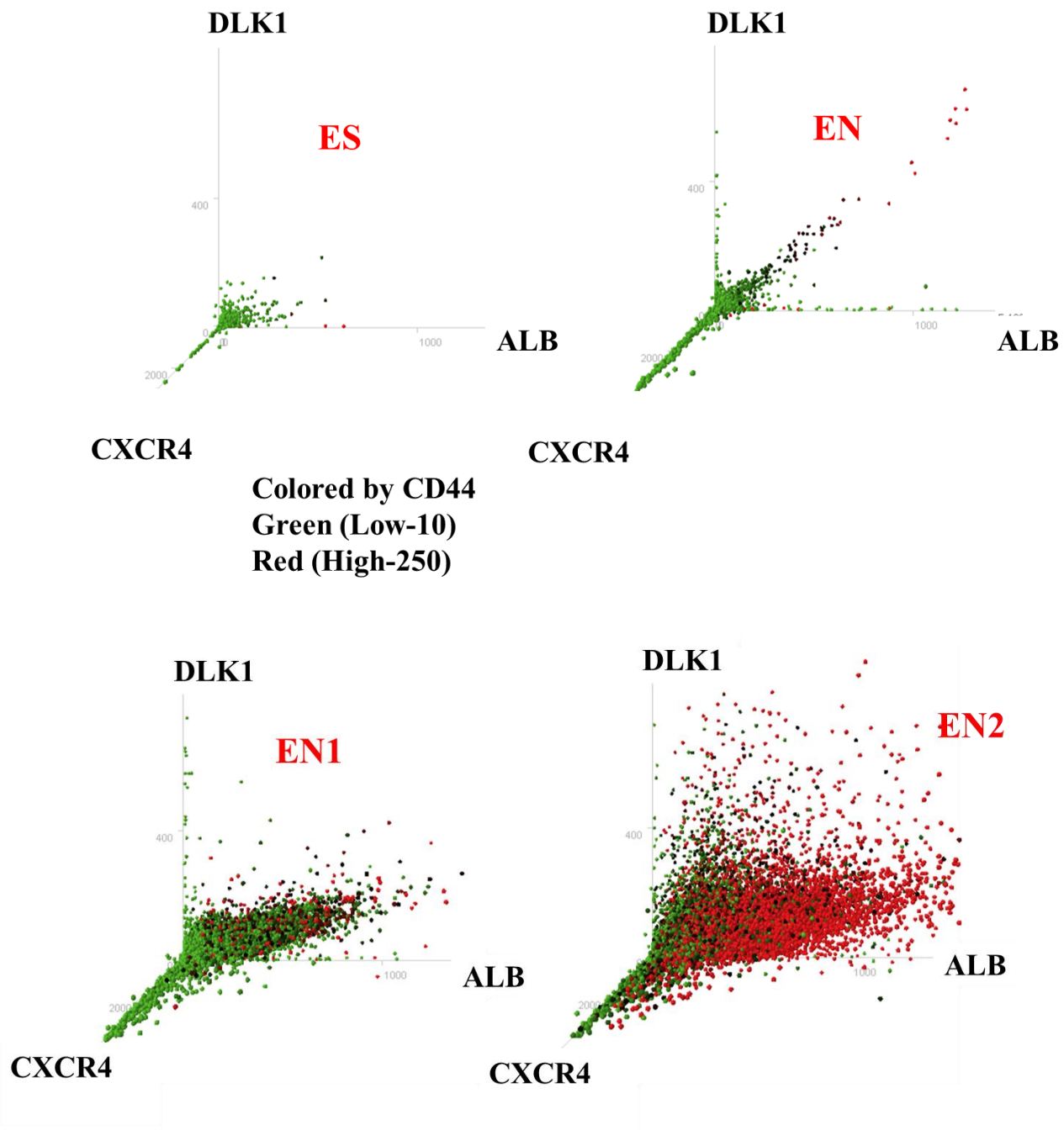


Figure 4-9: Mass Cytometry data for hESCs,EN,EN1,EN2 cells is shown where the x, y and z axes represent ALB, DLK1 and CXCR4 and the color represents CD44 expression. Green indicates low and red indicates high expression. hESCs are shown as the negative control. The endodermal genes decrease in most cells with increase in hepatic gene expression in EN1 and EN2 cells when compared to the EN cells.

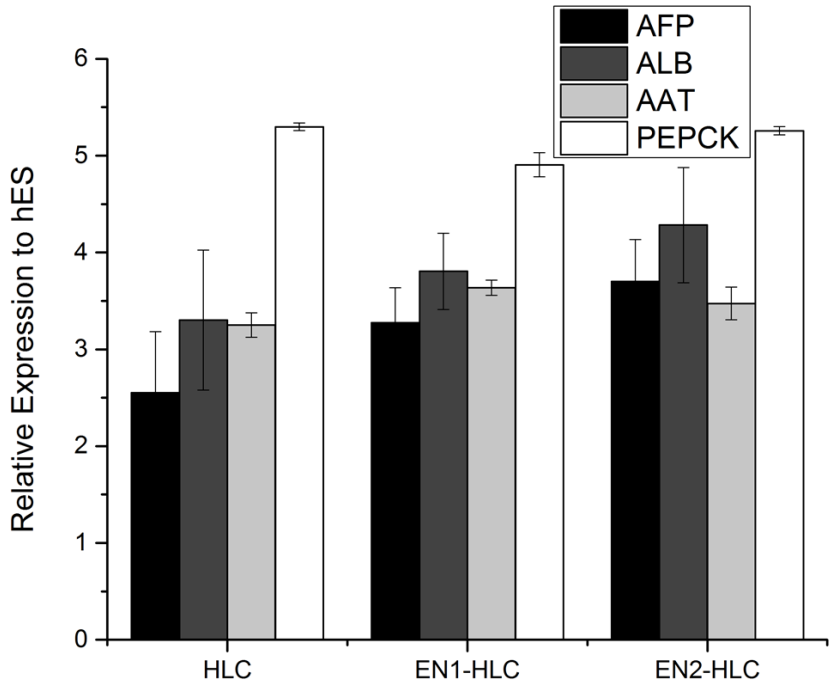


Figure 4-10: Differentiation properties comparing hepatocyte-like cells obtained from the traditional method of differentiation and the method of differentiation using a intermediate stage of expansion at the endoderm stage. A) HLCs, EN1-HLC, and EN2-HLC show similar levels of hepatic markers at the transcript level where we saw an even high expression in EN2-HLC compared to the other two samples

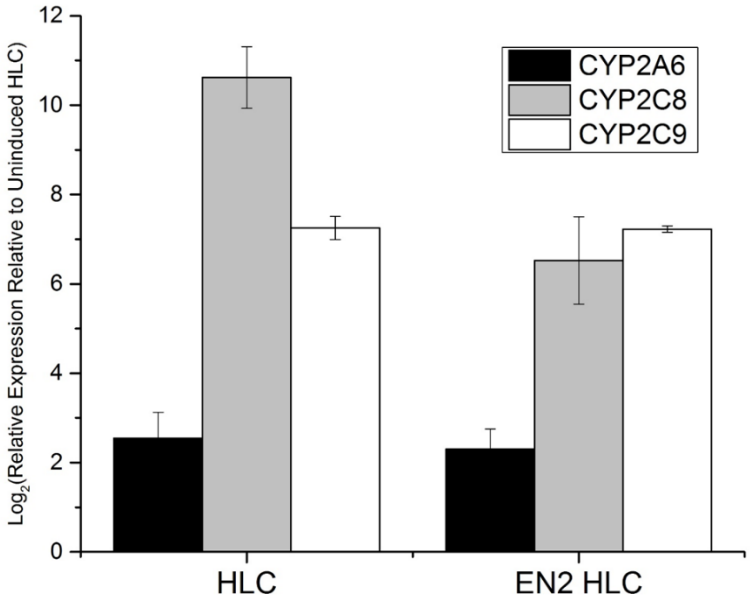


Figure 4-11: HLCs and EN2-HLCs were subjected to rifampicin and revealed similar drug metabolism abilities

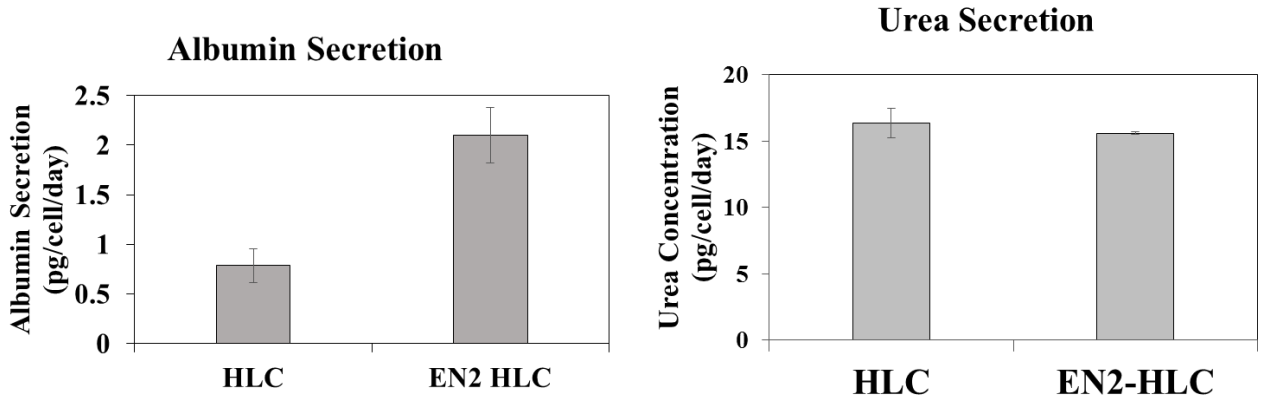
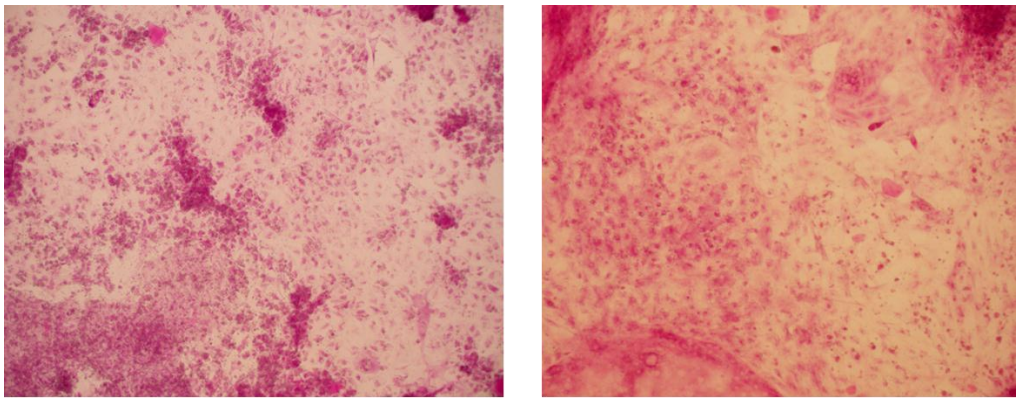


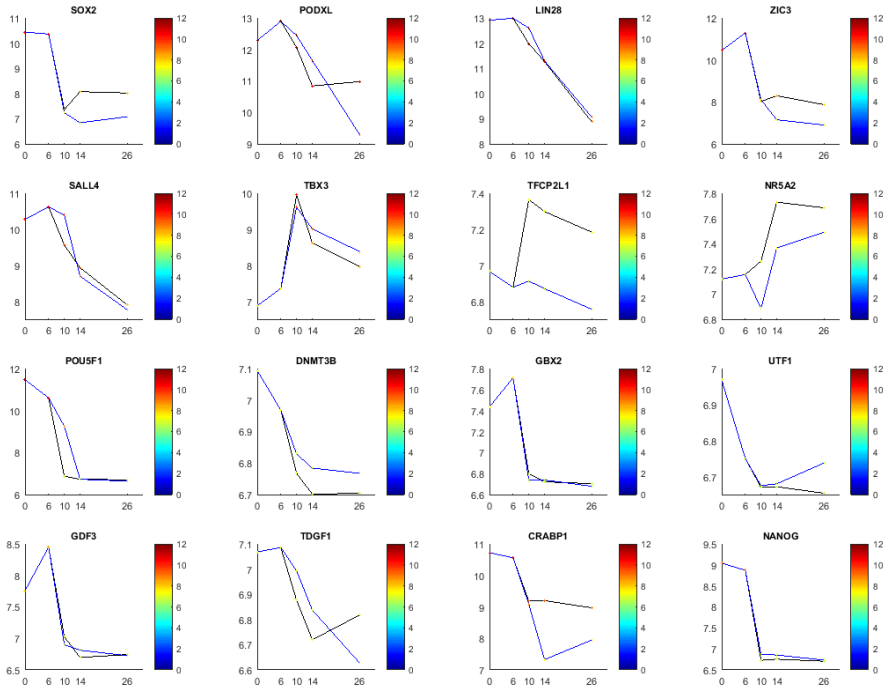
Figure 4-12: Functional behavior of albumin secretion and urea secretion also showed that HLCs and EN2-HLCs have similar function



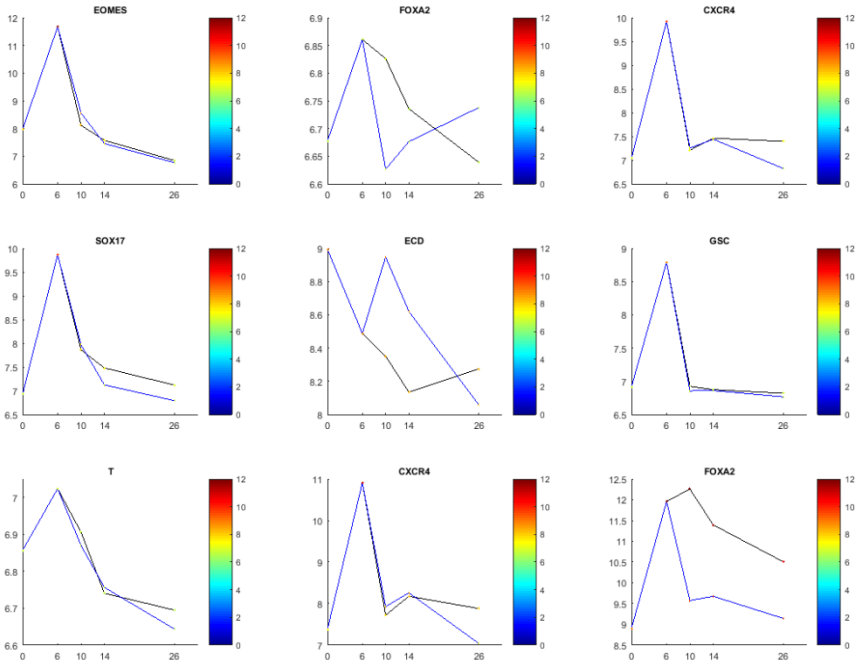
HLC EN2-HLC

Figure 4-13: Functional analysis was carried out to compare HLCs and EN2-HLCs illustrating both differentiated cells were able to perform glycogen storage

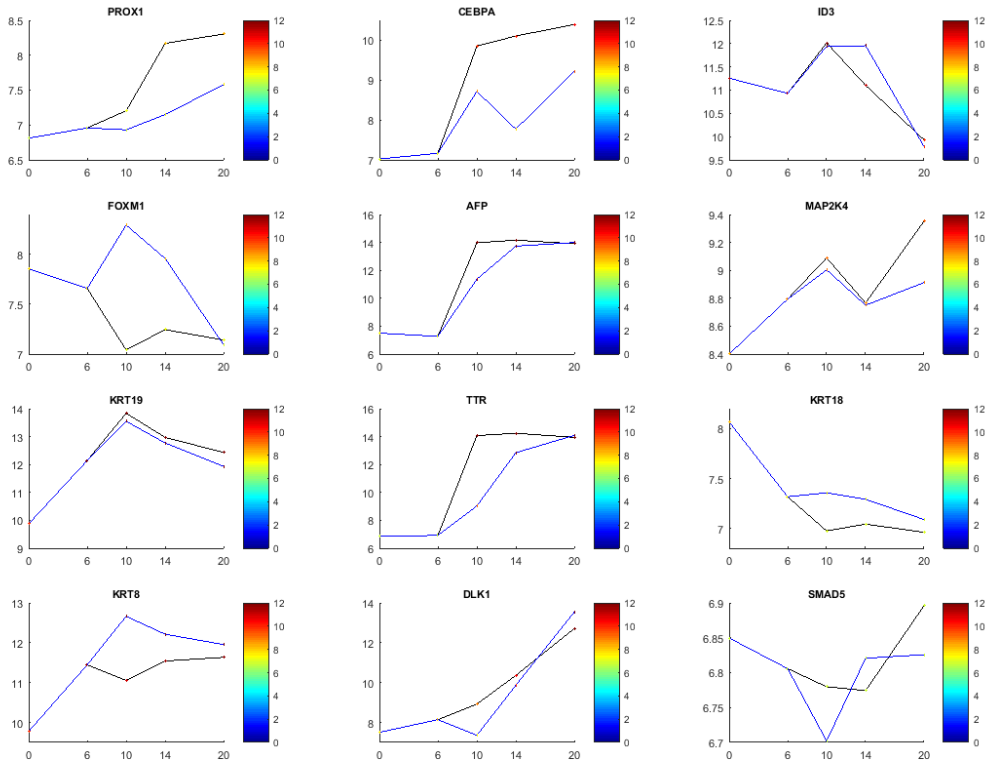
A


SOX2
PODXL
LIN28
ZIC3
SALL4
TBX3
TFCP2L1
NR5A2
POU5F1
DNMT3B
GBX2
UTF1
GDF3
TDGF1
CRABP1
NANOG

B


EOMES
FOXA2
CXCR4
SOX17
ECD
GSC
T
CXCR4
FOXA2

C


PROX1
CEBPA
ID3
FOXM1
AFP
MAP2K4
KRT19
TTR
KRT18
KRT8
DLK1
SMAD5

D

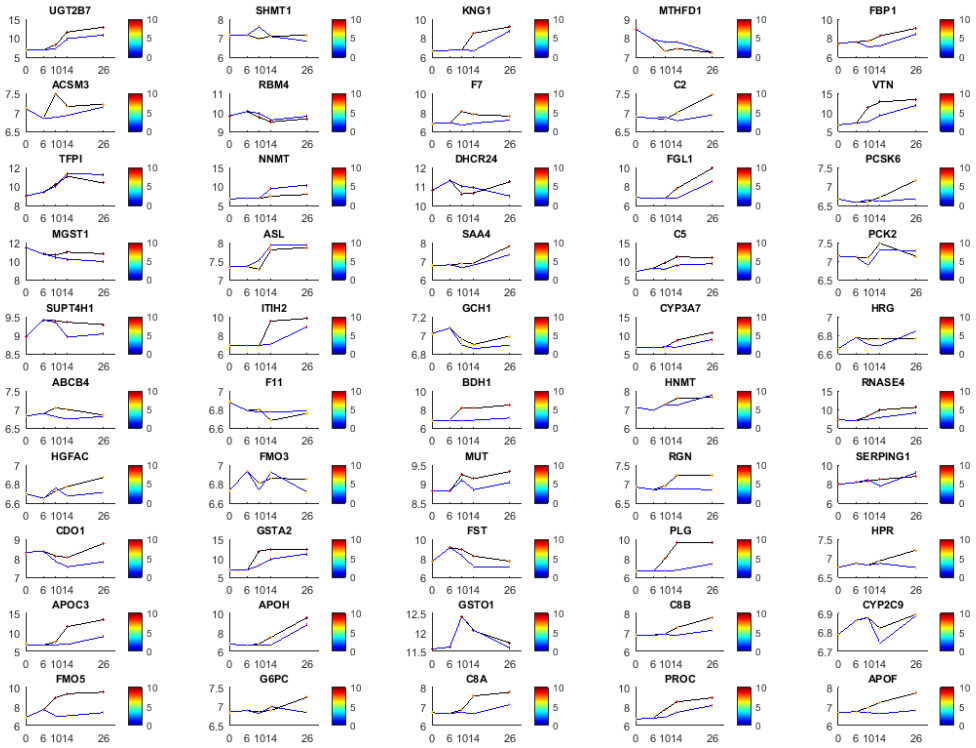

UGT2B7
SHMT1
KNG1
MTHFD1
FBP1
ACSM3
RBM4
F7
C2
VTN
TFPI
NNMT
DHCR24
FGL1
PCSK6
MGST1
ASL
SAA4
C5
PCK2
SUPT4H1
ITIH2
GCH1
CYP3A7
HRG
ABCB4
F11
BDH1
HNMT
RNASE4
HGFAC
FMO3
MUT
RGN
SERPING1
CDO1
GSTA2
FST
PLG
HPR
APOC3
APOH
GSTO1
C8B
CYP2C9
FMO5
G6PC
C8A
PROC
APOF

Figure 4-14: Transcriptome analysis comparing differentiated cells from traditional protocol and the newly established expansion protocol. A) Pluripotent genes B) Endoderm genes C) Hepatoblast genes D) Hepatic genes. The color bar represent the transcript intensities. The black and blue curves represents the modified and original protocols.

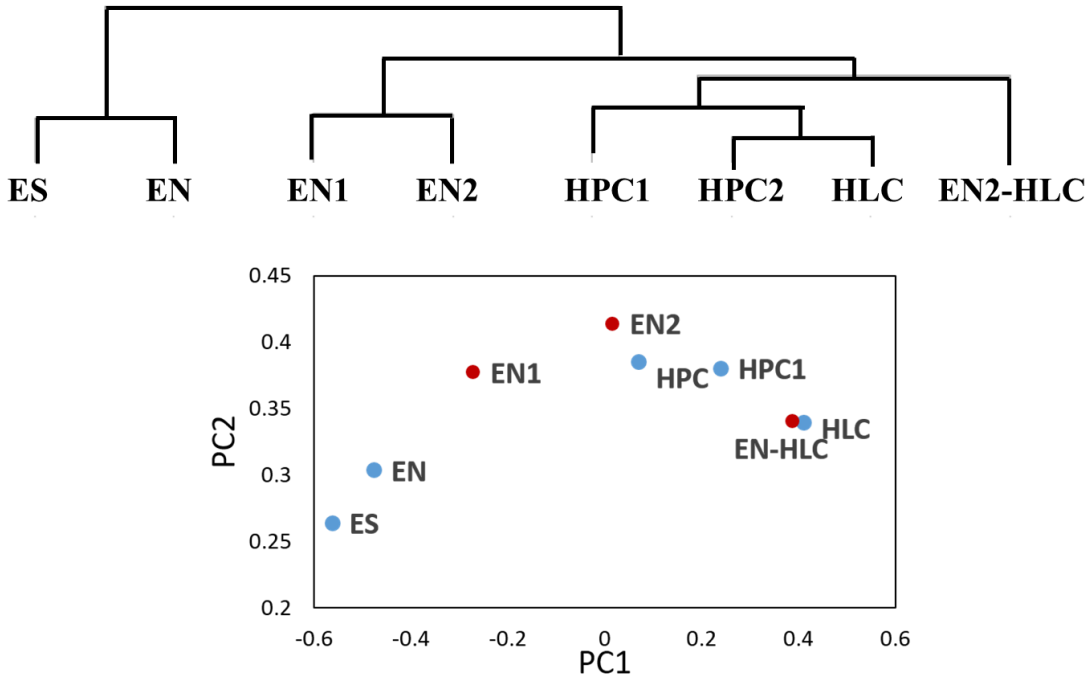
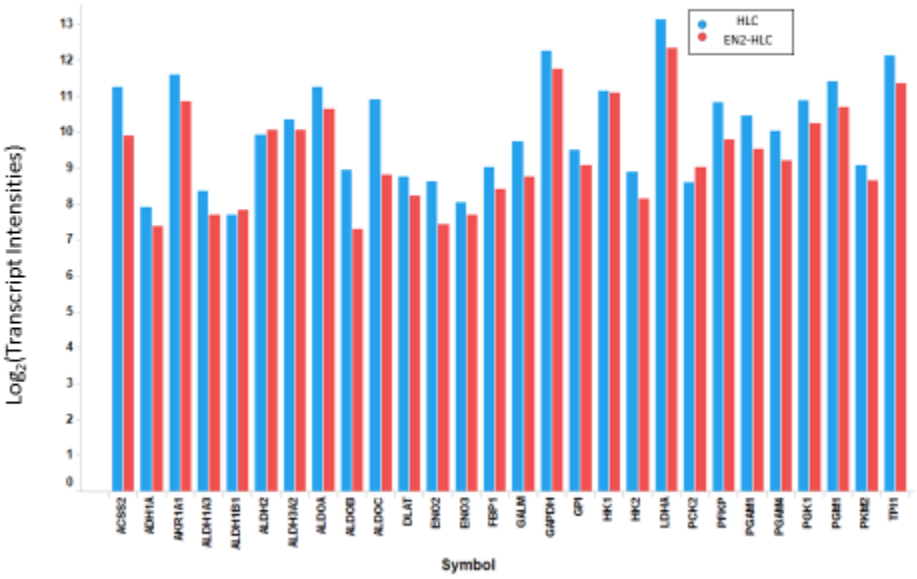
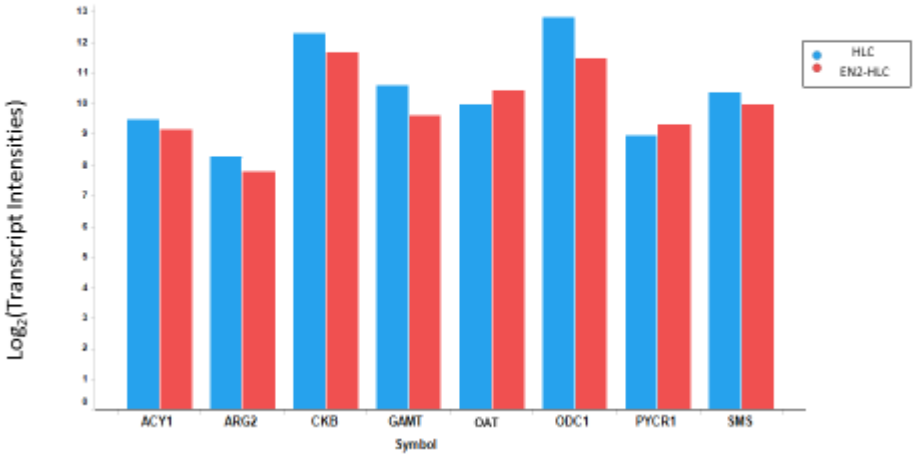


Figure 4-15: Transcriptome analysis of different samples reveal similarities of hepatocyte-like cells between the existing protocol and the expansion protocol through A) hierarchical clustering B) Principal Component Analysis. The first two PC's of all the samples are plotted in the B. The samples from the original protocol and modified protocol are marked in red and blue

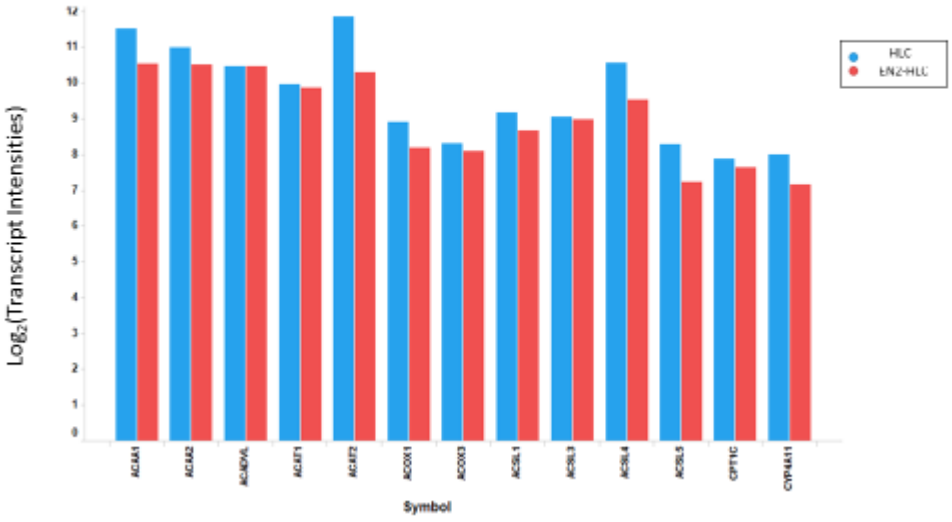
GLYCOLYSIS/GLUCOGENESIS



UREA CYCLE



FATTY ACID METABOLISM



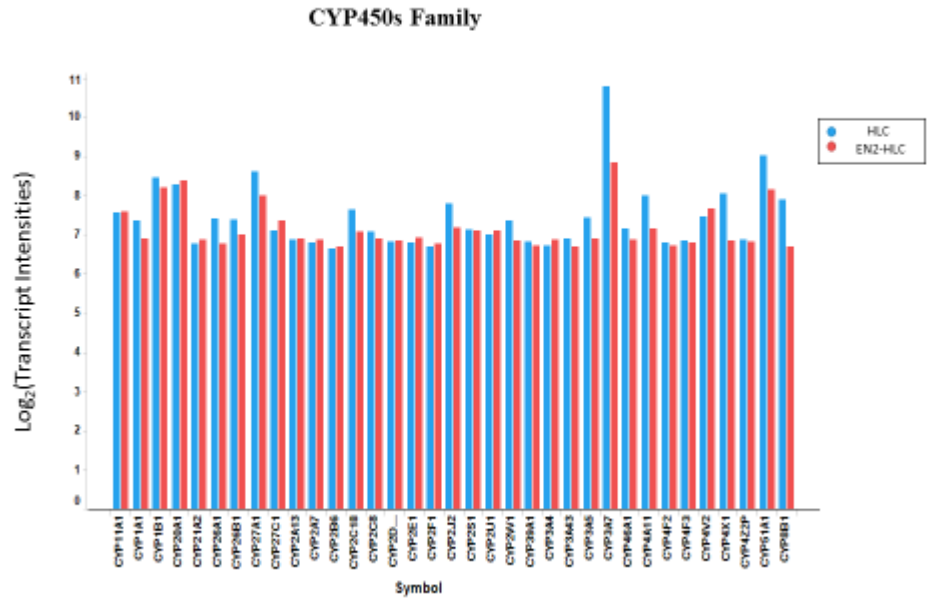


Figure 4-16: Expression levels of different functions were compared between hepatocyte-like cells and hepatocyte-like cells from endodermal expansion.. Expression levels of A) Glycolysis B) Urea cycle C) Fatty acid metabolism D) CYP450 enzyme genes showed similar level between HLCs and EN2-HLCs. The HLCs and EN-HLCs are shown in blue and red respectively.

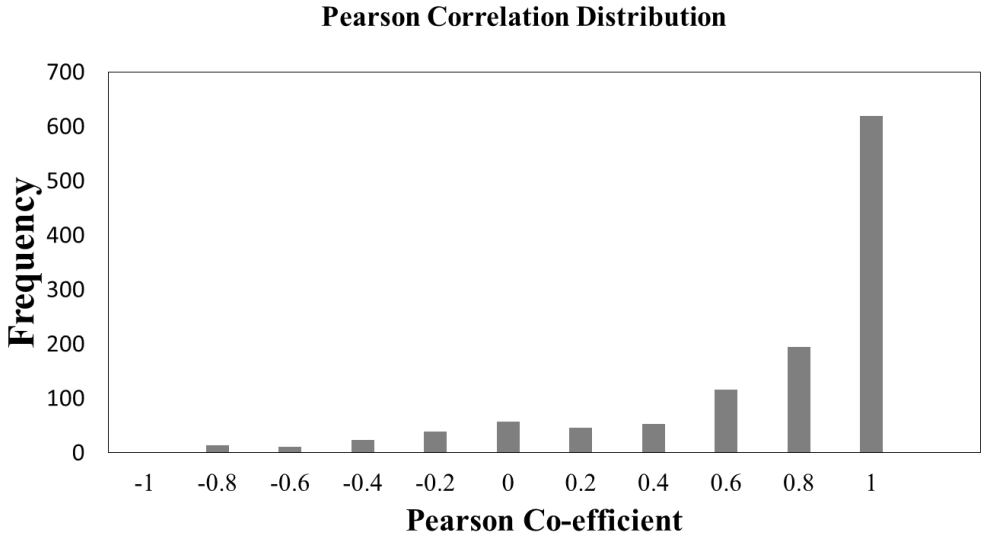


Figure 4-17: Pearson Co-efficient was computed among different time points in both protocols to identify genes with similar or contrasting dynamics. A distribution

of coefficients is shown. 1 indicates a strong positive correlation and -1 indicates a strong negative correlation.

Table 4-1: Genes that are negatively correlated from the pearson correlation analysis reveal that these genes do not participate in important biological functions

Symbol	Definition
AK3L1	Adenylate kinase 3-like 1
TRIL	TLR4 interactor with leucine rich repeats
AK3L1	Adenylate kinase 3-like 1
ALDOC	Aldolase C, fructose-bisphosphate
MXRA5	Matrix-remodelling associated 5
CLDN3	Claudin 3 (CLDN3).
MAP1B	Microtubule-associated protein 1B, transcript variant 2
SHRM	Shroom (SHRM)
RARRES2	Retinoic acid receptor responder (tazarotene induced) 2
RAB3IL1	RAB3A interacting protein (rabin3)-like 1
S100A16	S100 calcium binding protein A16
SNX26	Sorting nexin 26

Table 4-2: Geneset Enrichment Analysis results show the biologically relevant pathways upregulated in the different cell types

NAME	FDR q-val (%)	Enriched
KEGG_ECM_RECEPTOR_INTERACTION	0	EN2
KEGG_HEMATOPOIETIC_CELL_LINEAGE	0	EN2
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	0.0029	EN2
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	0.0026	EN2
KEGG_APOPTOSIS	0.0143	EN2
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	0.0221	EN2
KEGG_P53_SIGNALING_PATHWAY	0.0458	EN2
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	0.0707	EN2
KEGG_DRUG_METABOLISM_CYTOCHROME_P450	0.0833	EN2
KEGG_PPAR_SIGNALING_PATHWAY	0.1316	EN2
KEGG_CELL_ADHESION_MOLECULES_CAMS	0.1545	EN2
KEGG_WNT_SIGNALING_PATHWAY	0.0004	EN
KEGG_TGF_BETA_SIGNALING_PATHWAY	0.1354	EN

Table 4-3: Genes differentially expressed by five fold or higher in HLCs and EN-HLCs

Symbol	H9_D20	D6P2_Diff_d20	Log2(Fold Change)
ACTA1	11.1067	8.38899	2.71773
ACTG2	9.60846	12.7988	3.19034
ADAMTS1	8.65104	11.0802	2.4292
AHSG	13.5762	10.1642	3.41201
ANKRD38	10.2214	7.77595	2.44548
ANXA13	7.42085	9.81904	2.39819
APOA4	13.72	8.8395	4.88047
APOC2	10.0518	7.48163	2.57013
APOC3	13.3303	8.90385	4.42642
BHMT	9.20836	6.79231	2.41605
COL14A1	7.05913	9.41768	2.35855
CRYAA	9.95287	7.1319	2.82097
CST1	11.6516	7.88067	3.77094

EGFL6	8.48672	11.2259	2.73919
FMO1	10.3366	8.00172	2.33493
FOS	8.01669	10.4092	2.39256
GABRP	11.1903	6.9078	4.28249
GCGR	9.59124	6.77351	2.81773
HOPX	8.4149	11.4235	3.00865
HYAL1	10.3388	7.34903	2.98974
IGSF1	11.0816	8.33392	2.74765
IL1RL1	7.38202	10.4958	3.11374
MGP	8.50984	11.9262	3.41632
MST1	11.9363	9.51991	2.41634
NNMT	7.89405	10.3769	2.48287
SERPINA7	12.1384	8.9683	3.17013
SERPINC1	10.4065	7.61234	2.79411
SERPIND1	10.0648	7.3601	2.70465
SLC13A5	11.4297	8.9906	2.43914
SLC7A10	9.53267	7.17641	2.35626
TF	13.9989	11.0814	2.91749
TFF3	7.6737	10.4582	2.78447
TFPI2	7.31491	9.7234	2.4085
TSPAN8	7.0043	10.067	3.06273
TXNIP	7.24371	9.57456	2.33085
UGT2B28	10.8571	8.52376	2.3333
VTCN1	9.71793	7.04247	2.67546

5. ENRICHING HEPATIC LINEAGE FATED CELLS FROM STEM CELL DIFFERENTIATIONS

5.1 Introduction

Liver failure is a major cause of mortality for which the only cure is self-regeneration or organ transplantation. The transplant option is severely limited by the shortage of organ donors. Many liver disorders especially those that are related to metabolic deficiencies are caused by cellular malfunction in hepatocytes. Cellular transplantation in such scenarios could be a beneficial and less invasive alternative to whole organ transplantation [202]. Hepatocytes have been shown to successfully repopulate the liver in animal models with liver failure. Hepatocytes, derived from donors or cryopreserved cell banks, are typically infused through the portal vein and have been used to restore liver function or bridge patients to liver transplantation [11]. Hepatocyte transplantation however is also limited by the availability of donors and decline in cell viability and function with prolonged culture. Since, hepatocytes are terminally differentiated cells; their repopulation potential is limited when compared to a fetal hepatic progenitor or stem cell. Therefore, isolation of such cells from the fetal liver with high proliferative potential has been pursued as an alternative to hepatocytes for liver cell based therapy [203].

During embryonic liver development, hepatoblasts, a common progenitor cell for hepatocytes and cholangiocytes, undergo extensive proliferation during the formation of the liver bud. These cells are less likely to undergo de-differentiation in culture as primary hepatocytes and therefore are more suitable for transplantation [85]. Cells have mostly been isolated from embryonic day (E) 13-14 rodent fetal livers based on surface marker expression. Delta-like protein 1 (DLK1), typically expressed during early liver development, has been a popular surface marker to enrich rodent hepatoblasts [88-90] along with E-cadherin [91] , epithelial cell adhesion molecule (EPCAM) [92] and intercellular cell adhesion molecule (ICAM-1)[204]. These cells showed proliferative capabilities when cultured in appropriate culture conditions. A serum free medium developed by Kubota et al, was specifically tailored for the ex vivo propagation of

hepatic progenitor cells isolated from fetal liver [94,204,146]. However, the reliability on fetal liver sources has restricted these studies mostly to animal models.

Stem cells are an alternative source of generating hepatocyte like cells through differentiation strategies based on embryonic liver development. The differentiation protocol and the growth factors used have been described in the previous chapter. The differentiated hepatocyte like cells (HLCs) express key hepatic transcripts and functions. However, the key challenges with this protocol are the presence of heterogeneous cell populations after differentiation as well as limited functional maturity when compared to primary hepatocytes. We hypothesized that a similar strategy of enriching hepatic progenitor cells (HPCs), when applied to the *in vitro* differentiation process, could yield a homogenous population of cells, which can be used as a platform for future cell based therapies. These HPCs, if comparable to their *in vivo* counterparts could possess proliferative potential. Furthermore, the elimination of "unwanted" cells in culture would enable the pursuit of efficient differentiation strategies of these HPCs to high quality hepatocytes.

5.2 Results

5.2.1 HLCs are a heterogeneous population of cells

The human embryonic stem cell line (H9) was differentiated using a protocol previously published [182], and described in detail in the Material and Methods Section 3.2. Briefly, hESCs were passaged onto Matrigel coated plates. Differentiation was initiated by switching medium to a differentiation basal medium supplemented with stage specific growth factors or cytokines for specification of definitive endoderm (EN), hepatic progenitor cells (HPC) and hepatocyte like cells (HLCs). The hepatocyte like cells generated (HLCs) were found to express hepatic transcripts as shown by qRT-PCR in Figure 5-1A**.** The results are reported as log2 intensity values with respect to expression in hESCs. Alpha-fetoprotein, Delta-like homolog 1, Albumin and Alpha-1 antitrypsin are common markers of fetal hepatocytes or hepatic progenitor cells. Cholesterol 7-alpha hydroxylase (CYP7A1), Glucose-6-phosphatase (G6PC) and arginase (ARG) are mature hepatocyte markers that participate in cholesterol metabolism, gluconeogenesis and urea cycle respectively. Expression of fetal transcripts AFP, ALB, DLK1 and A1AT were

significantly higher by 1000 fold than mature transcripts CYP7A1, G6PC and ARG in the HLCs. The pluripotent stem cell marker OCT4 decreases with differentiation as expected. Asialoglycoprotein receptor (ASGR1) is a mature hepatocyte marker that plays a role in processing glycoproteins with terminal galactose or N-acetylgalactoseamine residues. About 23% of the HLCs were mature hepatocytes as quantified by the number of ASGR1 positive cells through flow cytometry as seen in Figure 5-1B. The secretory capabilities of HLCs are shown for albumin and urea in Figure 5-1C, D. However, similar functions in primary hepatocytes were at least an order of magnitude higher indicating that there is scope for further maturation. Transmission electron microscope imaging of HLCs showed hepatic features such as tight junctions and bile canaliculi as well as a high nucleus to cytoplasm ratio, which is indicative of a progenitor cell status.

The presence of an endothelial subpopulation was observed through co-staining for endothelial marker VE-cadherin along with fetal hepatic marker AFP, as early as D14 of differentiation, through flow cytometry (Figure 5-1E) This suggests that the differentiated progeny is a mixture of hepatic cells with varying maturity as well cells from other lineages. The presence of contaminating cell types during earlier stages of differentiation may have detrimental effects on the quality HLCs generated. Therefore, it is of interest to develop methods to enrich hepatic progenitor cells and then differentiate them to homogenous population of functional hepatocytes.

5.2.2 DLK1 as a surface marker for hepatic progenitor cells

Hepatic progenitor cells express a number of characteristic genes, including alpha-fetoprotein (AFP), Albumin, HNF-4α, CEBPα and CK-19 [1]. In addition to these intracellular markers, several surface markers characteristic of HPCs have been reported, such as CD29, CD34, CD133, c-kit, c-met, thy-1, N-cadherin, E-cadherin, EPCAM, NCAM, [127], CD44 [206], DLK1[88,89] and Liv2 [207]. Cell surface markers are desirable as they allow ease of cell enrichment without the genetic manipulation required for the development of hepatic promoter driven reporter cell lines. The gene expression dynamics of potential candidates in various stages of liver differentiation is shown in Figure 5-2. The expression values were obtained from time course microarray data collected using the protocol described in **Section 3.10** and are reported as fold change with respect to expression in hESCs. DLK1 is transmembrane protein with six epidermal growth factor (EGF) repeats, N-cadherin is a calcium dependent cell adhesion protein. CD44, ICAM and NCAM are cell surface glycoproteins that play a role on cell adhesion. DLK1I was the only marker that increased significantly during hepatic differentiation. Therefore, DLK1 was a promising marker for enriching hepatic progentitor cells based on its gene expression dynamics. The increase in DLK1 expression during differentiation was also confirmed by qRT-PCR Figure 5-3A). The intensity increased by 100 fold and 1000 fold in Day 14 HPCs and HLCs relative to hESC expression. The percentage cells expressing the DLK1 protein was found to be about 30% on day 14 or HPC1 cells by flow cytometry (Figure 5-3B). Finally, co-expression of DLK1 and AFP proteins was visualized through immunohistochemistry (Figure 5-3C). An isotype control was used the negative control for flow cytometry and immunohistochemistry experiments. Additionally, the protocol was optimized using hESCS and HEPG2 as a negative and positive control. Therefore, expression of DLK1, a hepatic progenitor marker, was indeed being upregulated during hepatic differentiation and enrichment of these DLK1+ cells would allow further investigation of their properties.

5.2.3 Cell surface marker based enrichment of Hepatic progenitor cells

Hepatic progenitor cells on day 14 of differentiation were harvested using a protocol described in **Section 3.7.** Magnetic activated cell sorting **(Section 3.8)** was used to enrich

the DLK1+ cell fraction from the HPC1 cells. Quantification of hepatic gene expression in DLK1+ cell fraction and the unlabeled (DLK-) cell mixture samples by qRT-PCR was used to confirm enrichment. Intensity values relative to expression in the unsorted day 14 bulk population are reported. DLK1 and AFP expression were expressed eight fold higher in the DLK1+ cell fraction than the unlabeled population as expected (Figure 5-4) However, the efficiency of cell sorting was low as DLK1+ cell yields recovered was only about 10% of the total population. Furthermore, upon culturing the enriched cells in Stage 4 medium conditions, cells began to loose their epithelial and gain a fibroblastic phenotype rapidly (Figure 5-5). This suggested that this method of cell enrichment was sub-optimal for enriching the hepatic progenitor cells.

5.2.4 Enrichment of HPCS using selective medium

Hepatocytes mediate cell-cell communication via different types of cell junctions [208]. Since these cell-cell junctions are important for hepatocyte integrity and function, dissociation into single cells for cell sorting may not be optimal. An enrichment medium is one that is selective for cells of a desired phenotype. A selective medium-based strategy would be preferential in the enrichment of hepatic progenitor cells, as this would eliminate cellular stresses caused by single cell dissociation and culture. Kubota et al optimized a serum free medium for ex vivo expansion of hepatic progenitor cells isolated from fetal rodent and human livers [94,204]. These cells formed epithelial colonies when cultured on plastic, possessed proliferative capabilities and maintained expression of fetal hepatic markers AFP and EPCAM [94,95,205,209]. It was therefore hypothesized that Kubota's medium could be selective for hepatic progenitors derived from stem cell differentiations. Hence, cell were harvested on day 14 and passaged as small clusters into different culture conditions.

5.2.4.1 Optimal extracellular matrix for hepatic progenitor cells

Extracellular matrix (ECM) is a critical component of the hepatic cellular niche and is essential for cellular attachment, growth and differentiation [101]. Cells, whether *in vivo* or *in vitro*, typically bind to an extracellular matrix (ECM), which anchors them, allows them to multiply or migrate and, most importantly, provides signals that can affect their fate (Chen et al, 2007). Progenitor cells maintained viability and liver specific gene

expression when cultured on appropriate biomatrices [210]. The native fetal liver matrix is composed mostly of collagens (Type I, III, and IV) along with laminin and other proteins. Stem cell hepatic differentiations are routinely performed on Matrigel, a complex mixture rich in ECM protein derived from mouse sarcomas, that has been found to promote cellular adhesion and viability[211].

To evaluate optimal culture conditions, cells harvested from day14 of stem cell differentiations were cultured on collagen I, collagen III, collagen IV, laminin and matrigel, in the presence of either Kubota's or Stage 4 medium. The cells were seeded at a density of 50,000 cells/cm^2. Hepatic gene and protein expression in cells cultured under these various conditions was evaluated by qRT-PCR and immunohistochemistry after six days in culture (Figure 5-6). The gene expression values are reported relative to expression in the traditional Stage 4 microenvironment to observe any improvement over this condition. The gene expression of cells with the combination of Kubota's with collagen I was the highest for hepatic markers (AFP, ALB, EPCAM and DLK1) compared to the other conditions. From immunohistochemistry, it is evident that majority of cells cultured in Kubota's medium on collagen I were expressing AFP, which is indicative of increased cell homogeneity. Cells cultured on other ECMs, with stage 4 conditions clearly showed the presence of an AFP- cell type (Figure 5-7). Therefore, Kubota's medium and collagen I were deemed suitable for culture of a homogenous population of hepatic progenitor cells. To evaluate the HPCs cultured in Kubota's further, gene expression of additional hepatic markers CYP3A4 and CYP3A7 was measured. CYP3A7 and CYP3A4 are CYP450 enzymes that are predominantly expressed in the fetal and adult livers. In differentiating HLCs, the CYP3A7 transcript is expressed at approximately 30 fold higher levels than CYP3A4 consistent with their immature status. Both these transcripts increased by tenfold in cells cultured in Kubota's medium compared to the Day 14 HPCs (n=4) (Figure 5-8). Furthermore, cells displayed an epithelial morphology upon culture in Kubota's medium, with the absence of other fibroblastic subpopulations. This was confirmed again using immunostaining for AFP (hepatic marker) and DAPI (nuclei stain) for cells in Kubota's medium and from Figure 5-7, where only AFP positive colonies are seen in Kubota's medium. The cells plated in

regular Stage 4 medium are also shown for comparison and consisted of several AFP negative cells (DAPI positive cells) alongside the AFP+ cells in the culture. The presence of a large non-hepatic sub-population suggests that these cells may have a growth advantage over HPCs in the Stage 4 conditions. To confirm this, D14 cells were harvested and stained with dye, Vybrant Dil labeling solution (Life Technologies) which is incorporated into the cell membrane. The labeled cells were replated into either Kubota's or Stage 4 medium conditions and visualized after two days. From Figure 5-10, the epithelial colonies in Kubota's largely retain the dye while the presence of unlabeled cells in Stage 4 conditions suggests that some cells may have undergone faster proliferation thereby diluting the dye in cell membranes. This along with the previous results indicate that some components of the Stage 4 medium may confer a proliferative advantage of undesired cells, which is absent in Kubota's medium conditions, thereby conferring the selectivity for hepatic cells.

Finally, the number of AFP and Albumin positive cells was found to be higher in Kubota's at 30 and 15% respectively relative to 15 and 5% respectively, for D14 cells and cells cultured in Stage 4 medium as seen in Figure 5-11. These results indicate that the Kubota's medium is useful as a selective medium for enriching hepatic cells.

5.2.4.2 Effect of growth factors on enhancement of hepatic phenotype

The combination of Kubota's medium and collagen I was found to be selective for hepatic cells. To optimize the culture conditions, the effects of growth factors was investigated. Hepatocyte growth factor (HGF) is a growth factor that allows hepatoblast survival through interactions with its receptor c-met, which is present on the surface of these cells. C-met activation enhances hepatoblast survival by inhibiting signals responsible for hepatoblast apoptosis [212]. Epidermal growth factor (EGF) is another factor implicated in cell proliferation and survival. Chen et al and Sasaki et al have demonstrated the use of a serum free media based on DMEM/F-12 including growth factors HGF and EGF in the culture of hepatoblasts isolated from rat and human liver [206,213]. Hepatoblasts isolated from E12.5 mouse embryonic livers were shown to proliferate in media supplemented with HGF and EGF [91]. The day 14 HLCs were found to express the HGF receptor c-MET and the EGF receptor EGFR transcript levels

about twofold higher when compared to hESCs based on microarray data. Therefore, the effect of supplementing growth factors, EGF and HGF to Kubota's medium on hepatic cells was investigated. The conditions tested were Day 14 HPCs on Collagen I treated plates with a) Kubota's medium without any growth factors b) Kubota's with EGF (20 ng/ml) and c) Kubota's with HGF(50ng/ml) and EGF(20ng/ml). Hepatic gene expression after six days in culture was compared with respect to the bulk D14 HPC population (n=2) as shown in Figure 5-12. Hepatic gene expression was lower for all the markers when cells were cultured only in the presence of EGF, and was approximately twofold higher when they were cultured in the presence of HGF and EGF when compared with cells cultured in Kubota's medium only. Therefore, a significant increase in hepatic phenotype was not observed upon growth factor supplementation to Kubota's medium.

While the combination of Kubota's medium and collagen I is selective for hepatic cells, it fails to confer the proliferative ability observed in the counterparts obtained from fetal liver, even in the presence of hepatic growth factors. This is in contrast to results for hepatic progenitor cells isolated from the fetal liver. This may be a consequence of differences in cell surface topology for hepatic progenitor cells obtained from *in vitro* and *in vivo* conditions.

5.2.5 Differentiation of hepatic progenitor cells to hepatocytes

The hypothesis for selectively enriching the hepatic population on Day 14 was that they could differentiate to a homogenous population of functional hepatocytes subsequently. The Kubota's medium enriched cells fated towards the hepatic lineage. To differentiate the HPCs to HLCs, the medium was changed from Kubota's to Stage 4 differentiation medium consisting of maturation growth factors HGF and Follistatin. Hepatic gene expression was evaluated after four days. The transition in expression of hepatic transcripts from D14 HPCs, to HPCs cultured in Kubota's medium for six days and then being differentiated to HLCs in Stage 4 medium is shown in Figure 5-13. The gene expression of D20 HLCs derived from the original protocol is also shown for reference. The highest change in expression of hepatic transcripts was observed between Day 14 HPCs and HPCs enriched after culture in Kubota's medium. Significant changes were not seen after the switch to the differentiation medium. Albumin, CYP3A4 and CYP3A7

showed a moderate two-fourfold increase in expression levels after treatment with Stage 4 medium when compared with the HPCs enriched in Kubota's medium (n=4) (Figure 5-13). The error bars reflect the inherent variability in stem cell differentiations. Effects on differentiation potential from addition of growth factors during HPC culture in Kubota's was also investigated and significant impact on hepatic transcripts was not observed as seen in Figure 5-14, n=2). Furthermore, transcript levels of HLCs derived from cells cultured in Kubota's medium were comparable to those of HLCs derived from the original protocol, in spite of starting with a more homogenous HPC population. While the initial results of enrichment using Kubota's is promising, further optimization of culture conditions for the proliferation and differentiation of HPCs will be necessary.

5.3 **Discussion**

Stem cells are a promising source of hepatocytes for clinical applications. Cellular therapies will entail transplantation of at least 10^9 cells per patient. Progenitor cells arising during differentiations are sought after, with the hypothesis that they can be proliferated like their *in vivo* counterparts and differentiated to hepatocytes, to sustain clinical demands.

Conventional bioprocessing knowledge can be applied to develop a robust process to generate stem cell derived products for clinical applications, as discussed in the literature survey. A key difference in a conventional bioprocess for recombinant protein production and stem cell culture lies in the product isolation stage. While the aim of cell separation operations in recombinant protein production, is to recover the desired products from the cell supernatant, the aim of a stem cell bioprocess, is to isolate desirable cell types while removing other undesired cell types and medium components. As previously emphasized, stem cell differentiation often results in a heterogeneous population, and the target cell population needs to be separated from the other cells. Notably, even a small level of contamination with undifferentiated or partially differentiated cells can be highly detrimental to therapeutic applications as these cells can potentially form teratomas in the host as reviewed in [214]. Thus, it is important that product isolation be as selective as

necessary in addition to being efficient. The isolation and enrichment of the product cell type is largely based upon cellular characteristics, such as surface marker expression.

One strategy is based on selective isolating or killing unwanted cells through the use of monoclonal antibodies [215]. Using an antibody against a surface glycoprotein specifically expressed on hESCs, the undifferentiated cells at the end of differentiation can be targeted for killing thereby reducing the risk of teratoma formation. On the other hand, the surface markers Stro1 and Stro3 are used to isolated specific MSC populations [216,217].

Another technique for isolating the desired cell types is by using selectable markers through genetic manipulation. Using a transgene that provides selective advantage, driven by a tissue specific promoter, a highly enriched population of target cells can be obtained. This approach is particularly useful for non-clinical applications. Mouse cardiomyocytes car thus be enriched from a differentiated population of mouse embryonic stem cells through the use of aminoglycoside phosphotransferase driven by α-cardiac myosin heavy chain promoter [218]. This technique has also been used for achieving the enrichment of other cell types such as neurons [219] or hepatocytes [220]. However, for this method to be suitable for clinical applications, a transient vector would need to be employed that would be reliably lost from the cell after a certain number of cell doublings is reached.

Cell surface markers are frequently used in the isolation of desirable cells from heterogeneous population, especially at the laboratory scale. Fluorescence-activated cell sorting (FACS) is a popular and effective tool for separating cells on the basis of fluorescent-antibody labeling of surface markers [106,221]. Magnetic cell sorting (MACS) is complementary to FACS in utilizing cell surface markers for selection and enrichment [222]. Cells are subjected to less stress during MACS isolation when compared to FACS procedures [223]. This makes it better suited for clinical and industrial scale up. It is noteworthy that this approach has been tested in a multicenter clinical trial conducted by the Bone and Marrow Transplant Clinical Trials Networks in which the CliniMACS® system was used to achieve sufficient levels of T cell depletion and isolating large numbers of $CD34^+$ cells [224].

As previously emphasized, the fact that these techniques rely on specific cell surface markers makes their wide spread use restricted by the limited existing knowledge about the relationship between cell surface markers and cellular phenotype and function. For embryonic and induced pluripotent stem cells, cellular markers have been identified through extensive research that enables assessment of cell quality [225]. These include surface markers such as SSEA-3/4 and TRA-1-60/81, as well as intracellular transcription factors such as Oct4 and Nanog, for which even the relationship between phenotype and function is well understood [225]. For liver, and other adult stem cell types, even though multiple markers have been identified for cell characterization, the relationship between their expression and cellular functions is not always understood [61]. This lack of knowledge about the functional significance of those markers poses some uncertainty on the nature of cells isolated based on those markers.

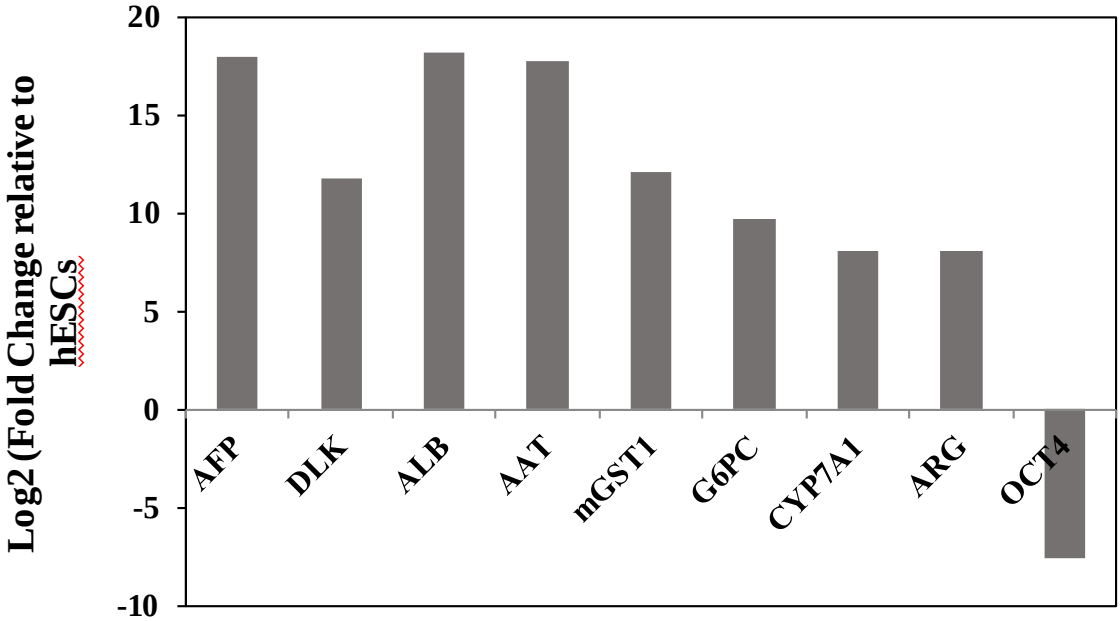
In this study, an extensive literature survey and analysis of transcriptome data were used to shortlist potential markers to enrich hepatic cells. Magnetic activated cell sorting was employed using DLK1 as a surface marker was shown to have increased hepatic transcripts compared to the unlabeled (DLK1-) cell fraction. However low cell yields and the inability to maintain a sustained hepatic phenotype in standard differentiation conditions was observed. Since cellular tight junctions play an integral role in hepatic integrity and function, dissociation of the hepatic cells to single cells and subjecting them to cell sorting could have effected cell yields and viability. Additionally, the Stage 4 microenvironment that was developed for differentiation may not have been optimal for prolonged cell culture of the enriched hepatic cells. Multiple fibroblastic cells were seen upon culture in the Stage 4 medium, suggesting that either an undesired cell subpopulation was proliferating faster than the hepatic cells or the hepatic cells themselves were undergoing epithelial mesenchymal transition. To preserve cellular integrity and eliminate cell sorting related stresses, a gentler method was sought to enrich HPCs. This would also enable easier downstream processing for product purification, when translating this approach into a therapy.

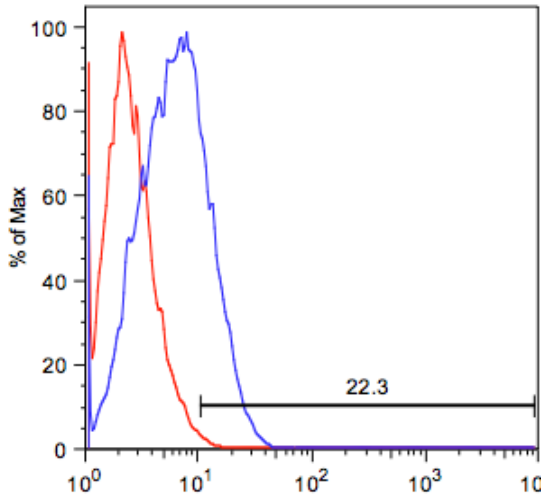
Kubota's medium combined with a collagen I surface were found to selectively enrich hepatic cells from stem cell hepatic differentiations. Culture of day 14 HPCs in Kubota's

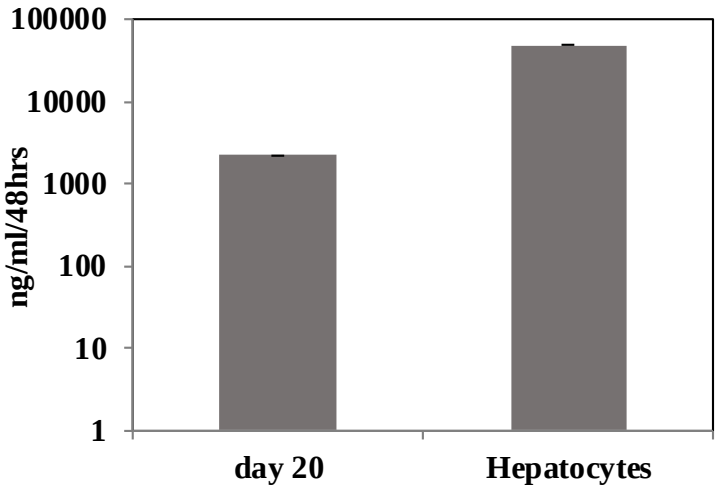
medium was accompanied by initial cell death. This reduction in cell yields is compensated by a limited 1.5 to 2-fold proliferation of the hepatic cells over six days in culture. This was accompanied by an increase in hepatic transcripts and proteins thereby indicating a selective pressure favoring cells fated towards the hepatic lineage. Interestingly when the same cells were cultured in Stage 4 conditions, a subpopulation of non hepatic cells were found to proliferate similar to the DLK1+ cells cultured in the same conditions. The additional surface area as well as the Stage 4 microenvironment could confer a growth advantage for these proliferating cells. In contrast, such cells are not observed in the Kubota's medium and Collagen 1. Apart from the difference in the extracellular matrix, differences in the medium components could be responsible for this. While the basal medium composition is largely the same for both media, Kubota's medium is additionally supplemented with nicotinamide, hydrocortisone and free fatty acids (linoleic, linolenic, palmitic, stearic and oleic acid), which could play a role in selectivity. Furthermore, the origin of these non hepatic cells is unclear as they could also arise as a result of dedifferentiation or epithelial mesenchymal transition of hepatic cells. In some conditions, the presence of HGF was found to stimulate EMT of epithelial cells[226].

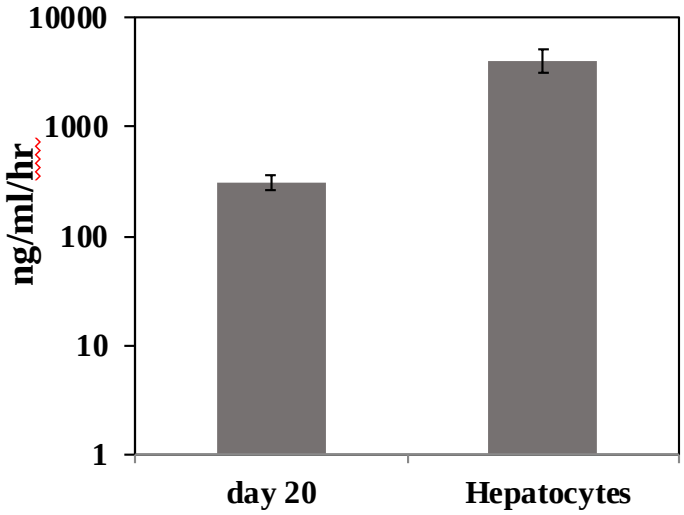
The hepatic cells that were enriched did not possess the highly proliferative properties of their *in vivo* counterparts cultured in the same culture conditions. In addition, when these cells were differentiated further, the HLCs generated were very similar to HLCs generated using the original protocol. The homogeneity of the starting cell population was not reflected in an increase in differentiated status of HLCs after differentiation. While the Kubota's medium and Collagen I microenvironment provide an encouraging start, further optimization of culture conditions for maintaining proliferative capability and successful differentiation will be needed.

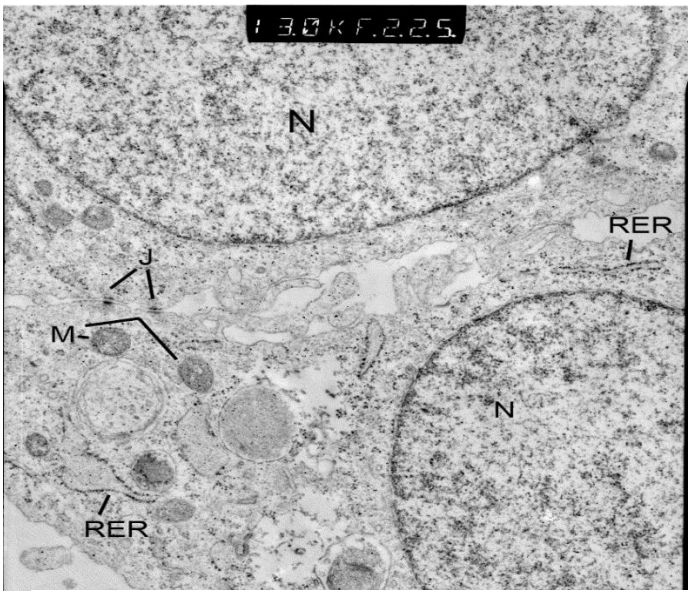
5.4 **Figures**


Log2 (Fold Change relative to hESCs
20
15
10
5
0
-5
-10
AFP
DLK
ALB
AAT
mGST1
G6PC
CYP7A1
ARG
OCT4


% of Max
100
80
60
40
20
0
22.3
10^0
10^1
10^2
10^3


100000
10000
1000
100
10
1
ng/ml/48hrs
day 20
Hepatocytes


10000
1000
100
10
1
ng/ml/hr
day 20
Hepatocytes



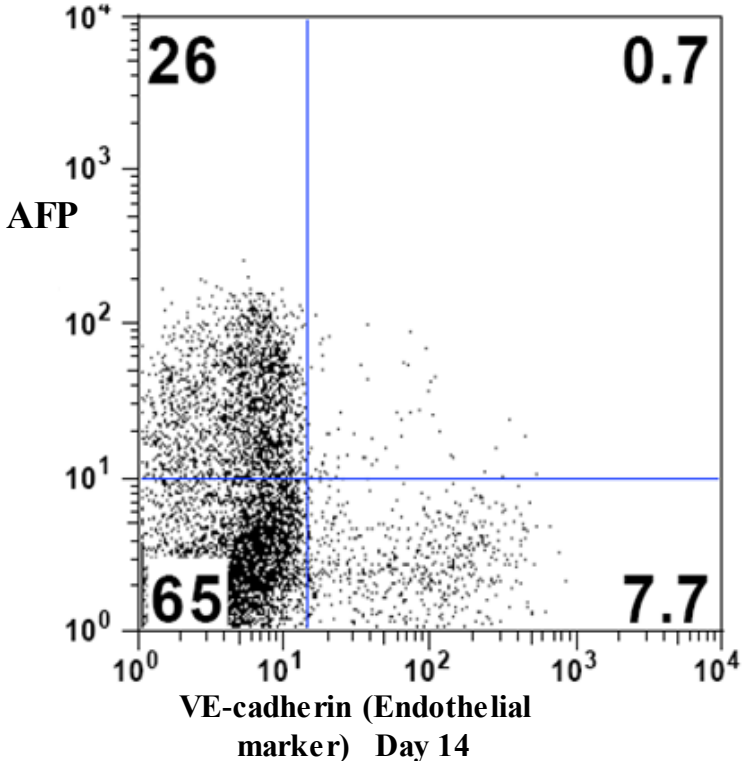


Figure 5-1: Characterization of stem cell differentiated HLCs. (A) Gene expression of pluripotent OCT4, fetal hepatic (AFP, DLK, A1AT, ALB) and mature hepatic (ARG,CYP7A1,G6PC, mGST1) are shown relative to their expression in hESCs (B) The number of cells expressing mature hepatic marker ASGPR by flow cytometry is shown by the red histogram and the isotype control is shown in blue (C) Comparison of ALB secretion in HLCs and primary hepatocytes (D) Comparison of urea secretion in HLCs and primary hepatocytes (E) A transmission electron

microscopy (TEM) micrograph of cells showing hepatic features along with high nucleus to cytoplasm ratio (F) Co-staining for hepatic marker (AFP) and endothelial marker (VE-cadherin) on Day 14 of differentiation by flow cytometry is shown.

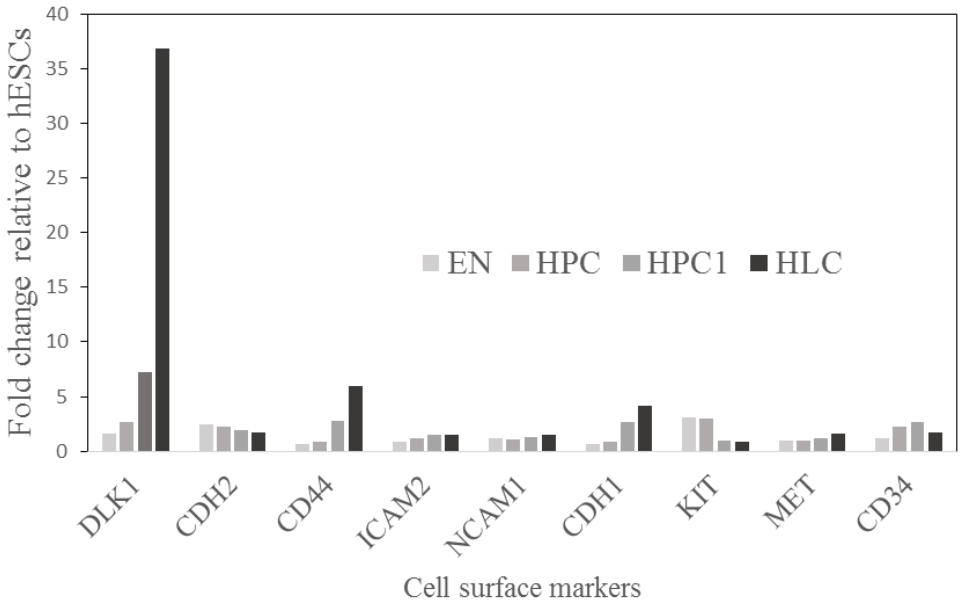
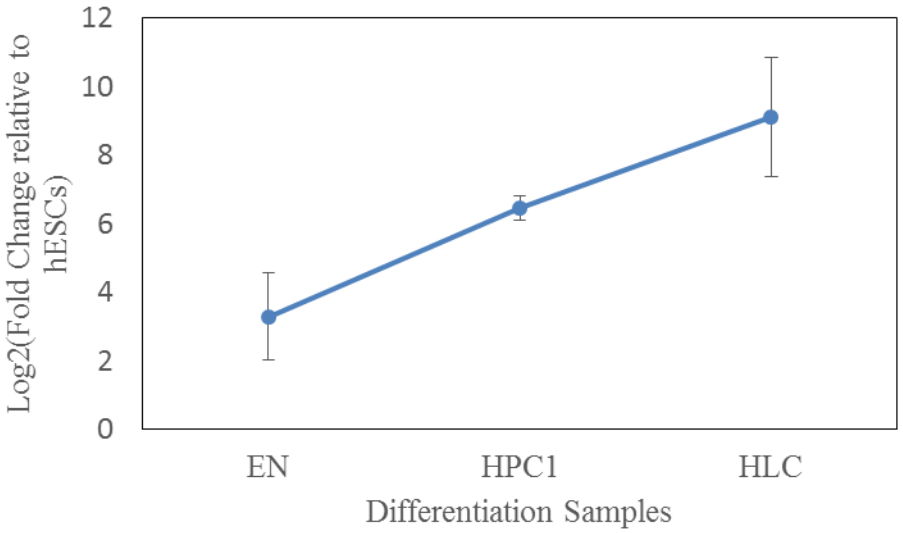
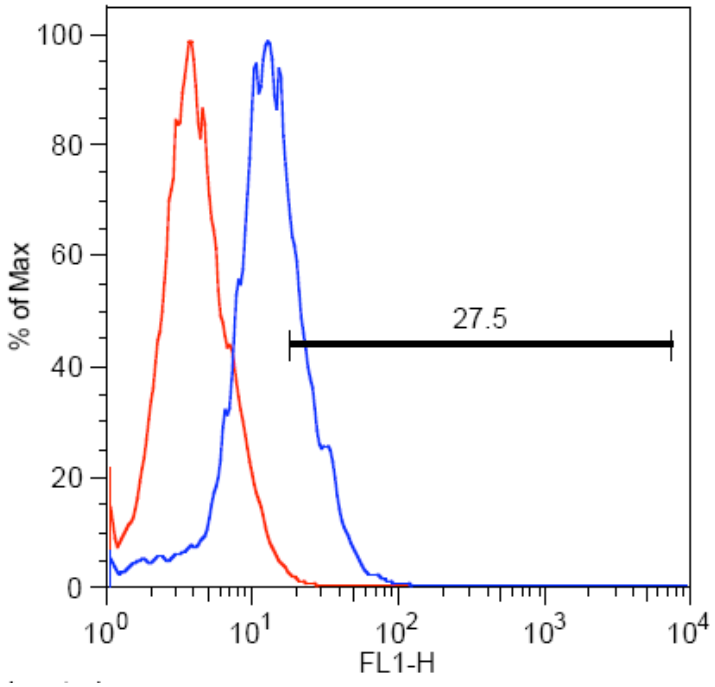


Figure 5-2: Dynamic gene expression of cell surface marker candidates for D6 endoderm (EN), D10,D14 hepatic progenitors (HPC,HPC1) and D20 hepatocyte like cells (HLCs) are shown. These values were obtained from transcriptome array data and are represented as fold change with respect to expression in hESCS.





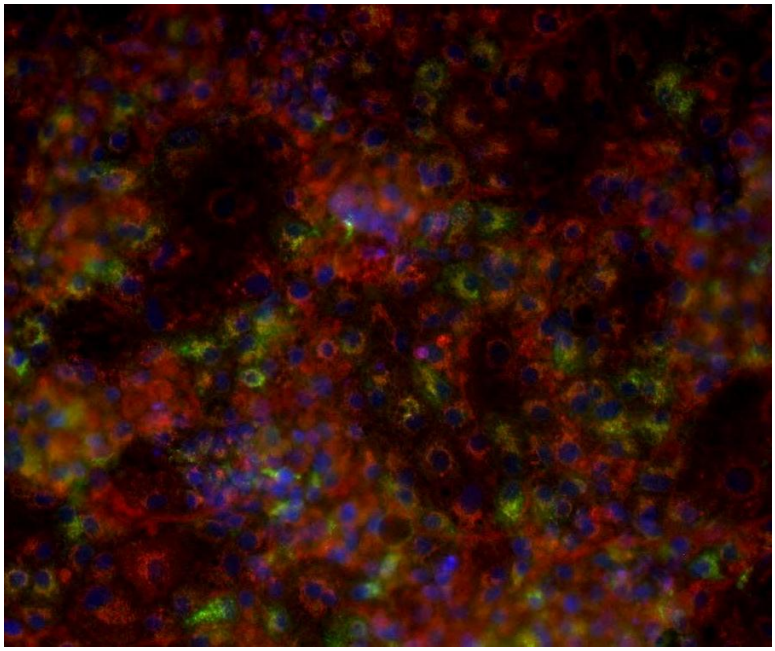


Figure 5-3: DLK1 expression during differentiation. A) Expression of DLK1 through various time points of differentiation was quantified by qRT-PCR. The values are reported as log2 intensity values relative to hESCs. B) About 30% of cells were found to express DLK1 on day 14 of differentiation. C) Immunostaining for AFP(red) with DLK1(green) on Day 14 of differentiation showed co-expression of these markers in many cells. The nuclei were stained with DAPI (blue)

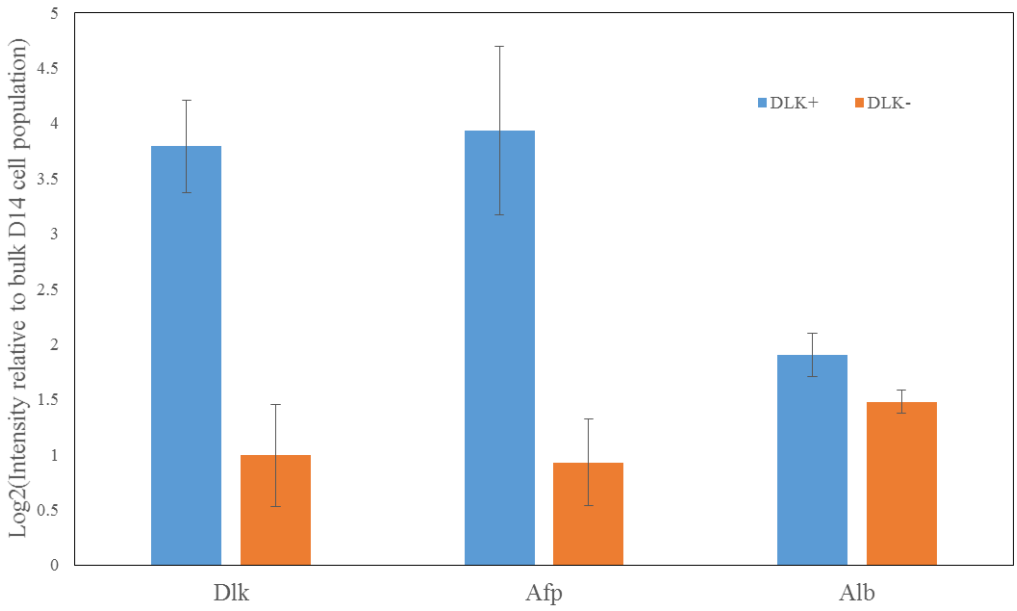


Figure 5-4: Hepatic gene expression in DLK1+ and unlabeled fractions after cell sorting on Day 14 of differentiation using MACS is shown (N=2). The values are reported as log2 values relative to the unsorted bulk Day 14 cell population. DLK+ cells express higher transcripts for the markers probed.

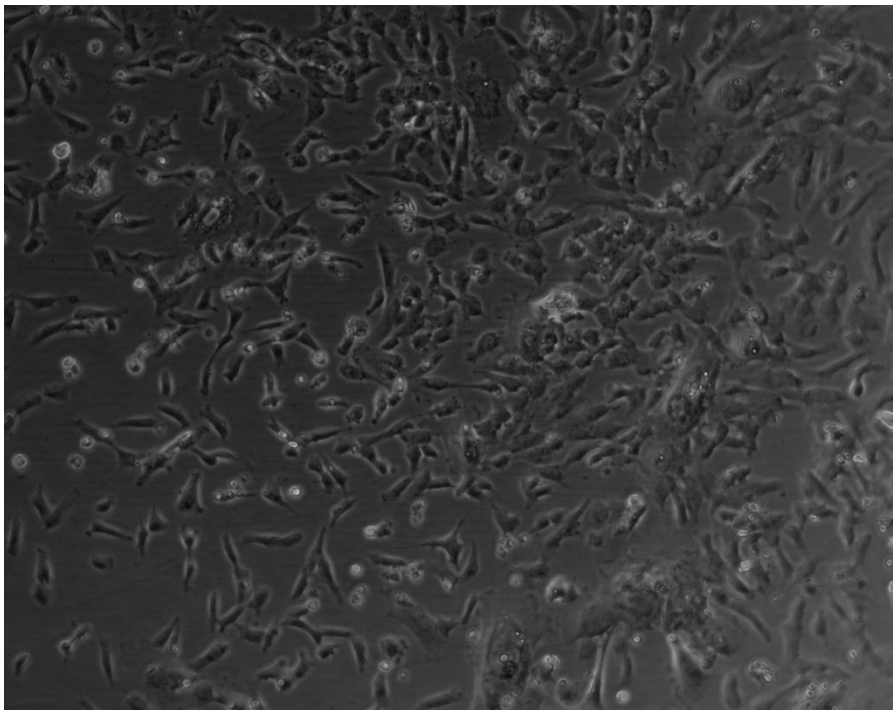


Figure 5-5: DLK1+ cells gain a fibroblastic phenotype within three days of culturing in Stage 4 medium.

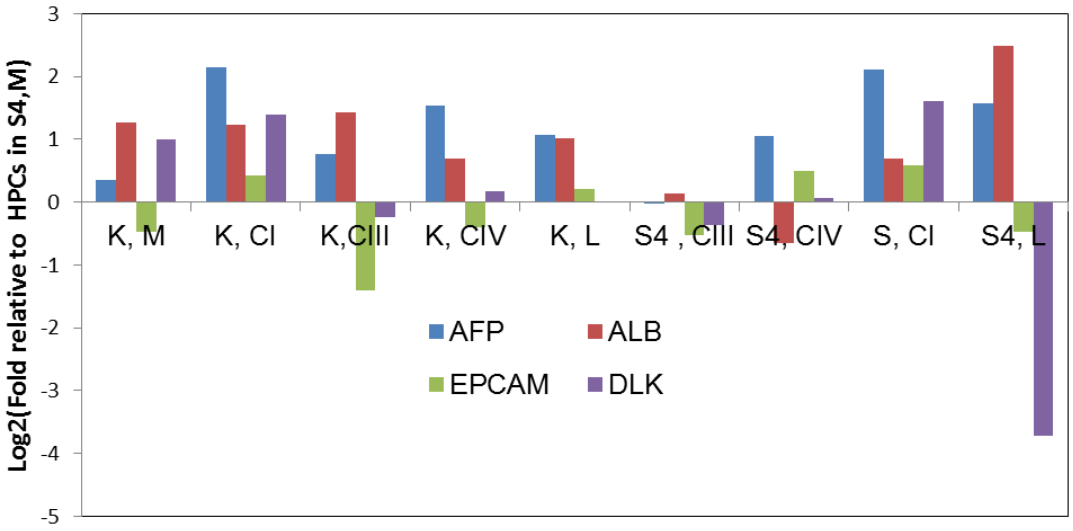


Figure 5-6: Hepatic gene expression in Day 14 cells that were harvested and cultured on Matrigel (M), Collagen I (CI), Collagen III (CIII), Collagen IV(CIV), Laminin (L) in either Kubota(K) or Stage 4(S4) medium for six days

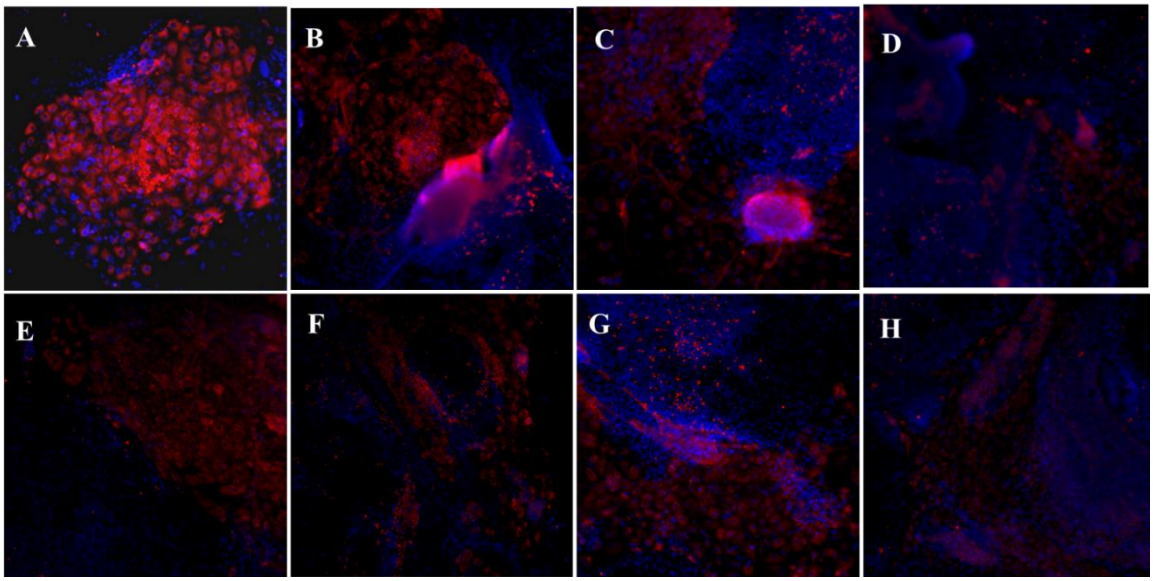


Figure 5-7: Immunostaining for AFP (red), DAPI (nuclear stain, Blue) in different culture conditions after six days is shown. A) Kubota, Collagen I B)Kubota, Collagen III C)Kubota, Collagen IV D) Kubota, Matrigel E) Stage 4 medium, Collagen I F) Stage 4 medium, Collagen III G)Stage 4 medium, Collagen IV H)Stage 4 medium, Matrigel

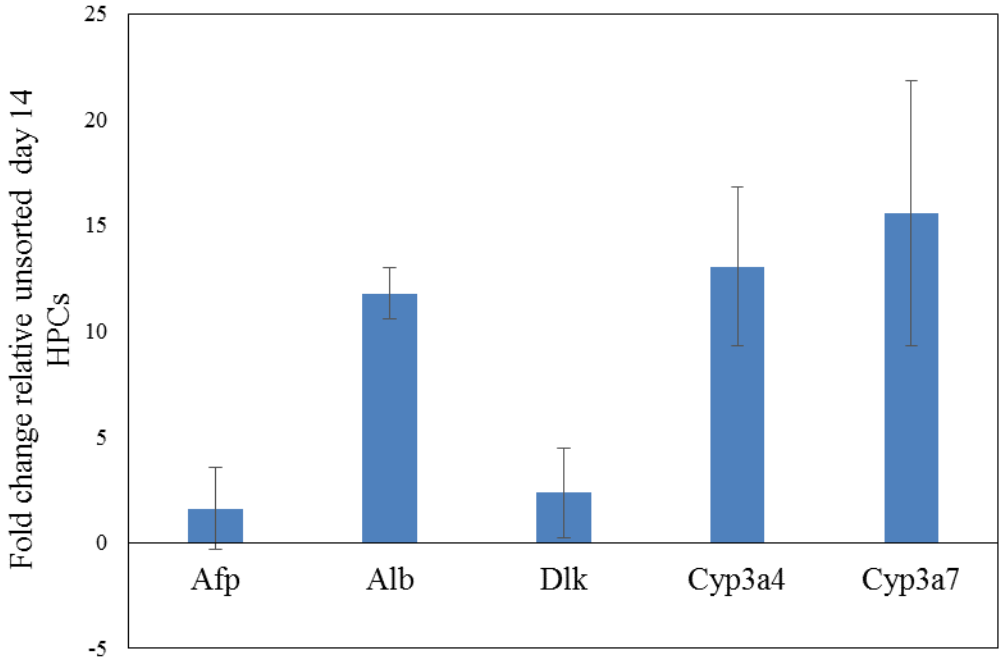


Figure 5-8: Hepatic gene expression of Day14 HPC cells cultured in Kubota's medium on Collagen I for six days (N=2). Increase gene expression relative to unsorted Day 14 cells was observed for cells in Kubota's medium

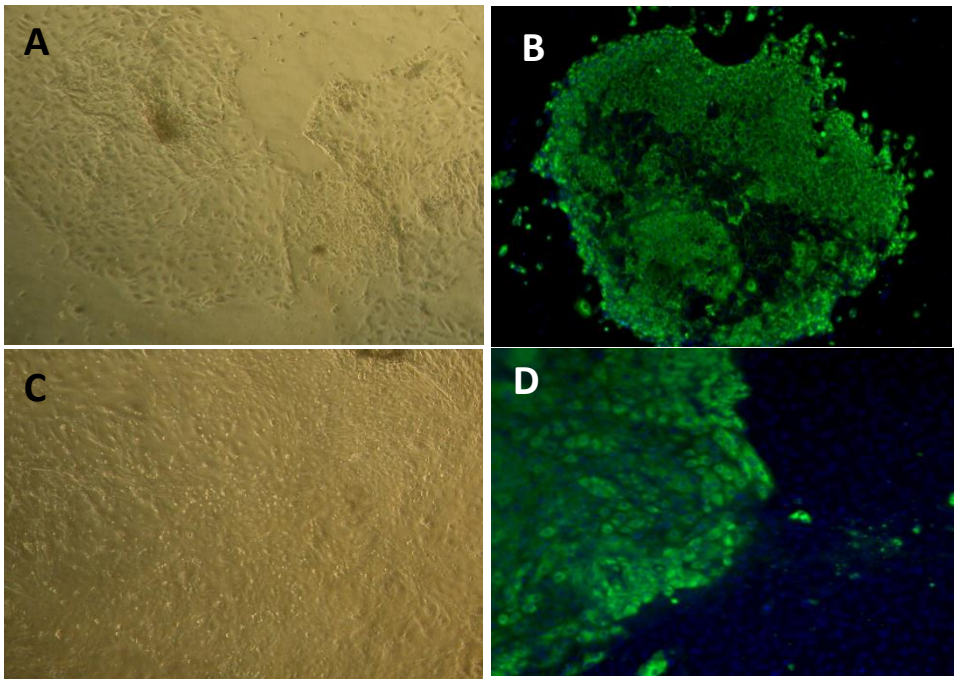


Figure 5-9: Morphology and immunostaining for AFP (Green) and nuclear DAPI (blue) is shown for cells cultured in Kubota's medium for six days. For comparison, cell cultured in Stage 4 medium conditions are displayed. The Kubota's medium is more selective for cells of hepatic morphology

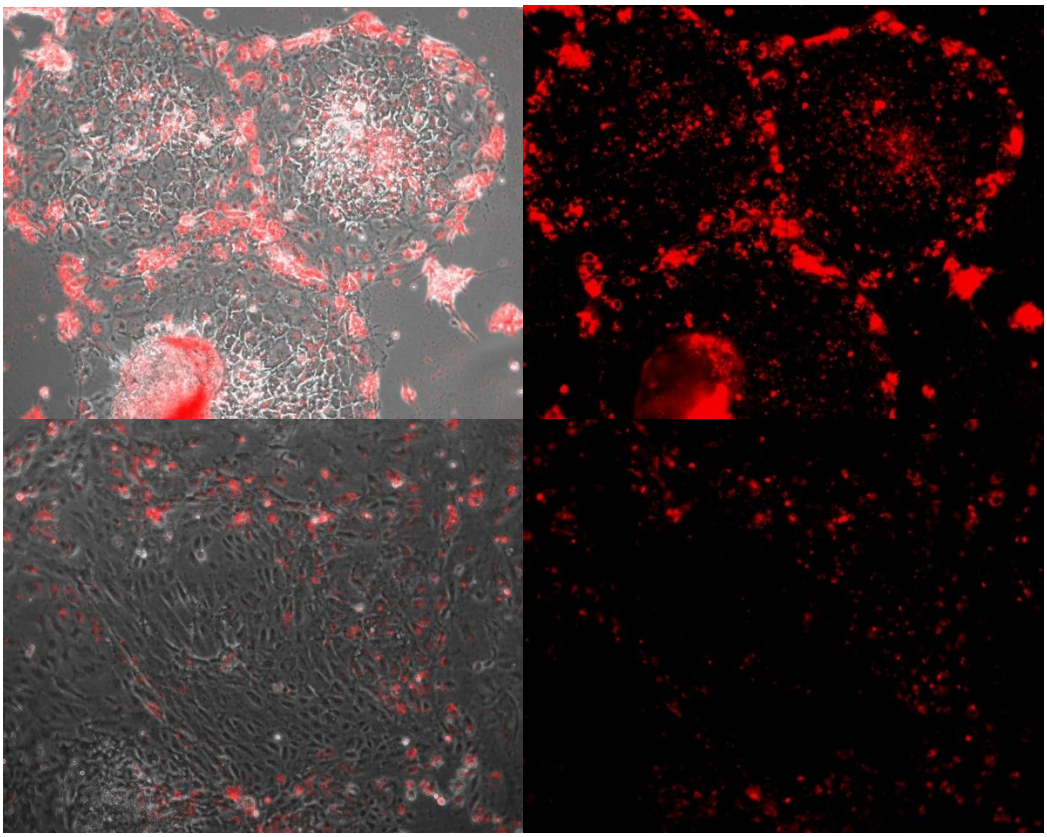


Figure 5-10: D14 HPCs were labeled with cell membrane Vybrant dye (Red) and cultured in different conditions for two days (A, B) Most cells cultured in Kubota's retain the dye (C,D) A non-hepatic subpopulation seems to be proliferating thereby diluting the dye in the Stage 4 conditions

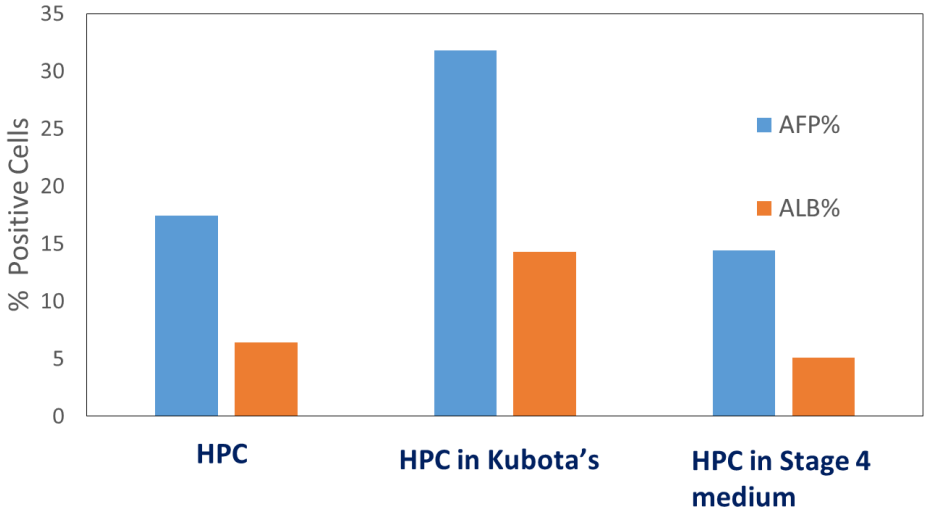


Figure 5-11: A higher number of cells expressing AFP and ALB were seen by flow cytometry in Kubota conditions vs the Stage 4 conditions after six days in culture

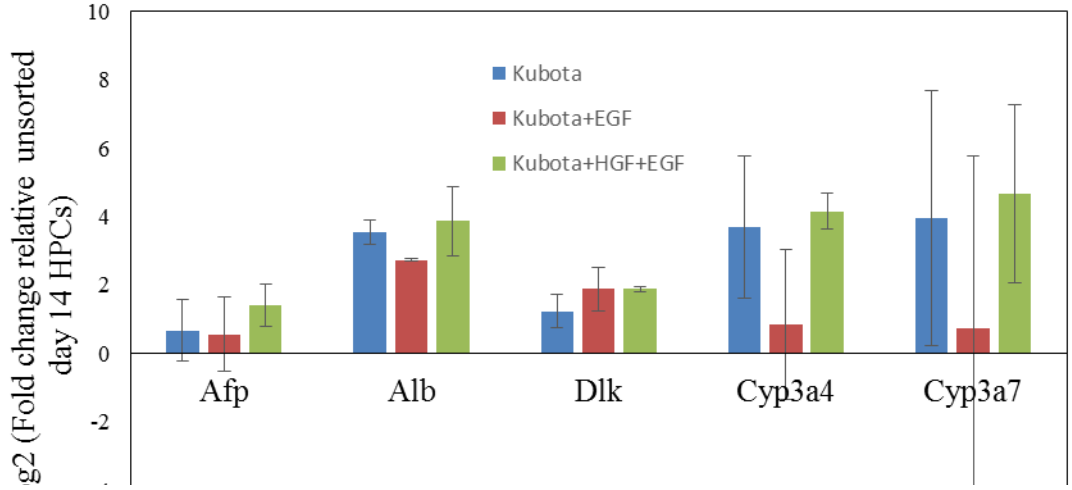


Figure 5-12: Effect of HGF and EGF growth factor supplementation in Kubota's medium on hepatic phenotype of Day 14 HPCs after six days in culture N=2. The error bars reflect inherent variability in stem cell differentiations.

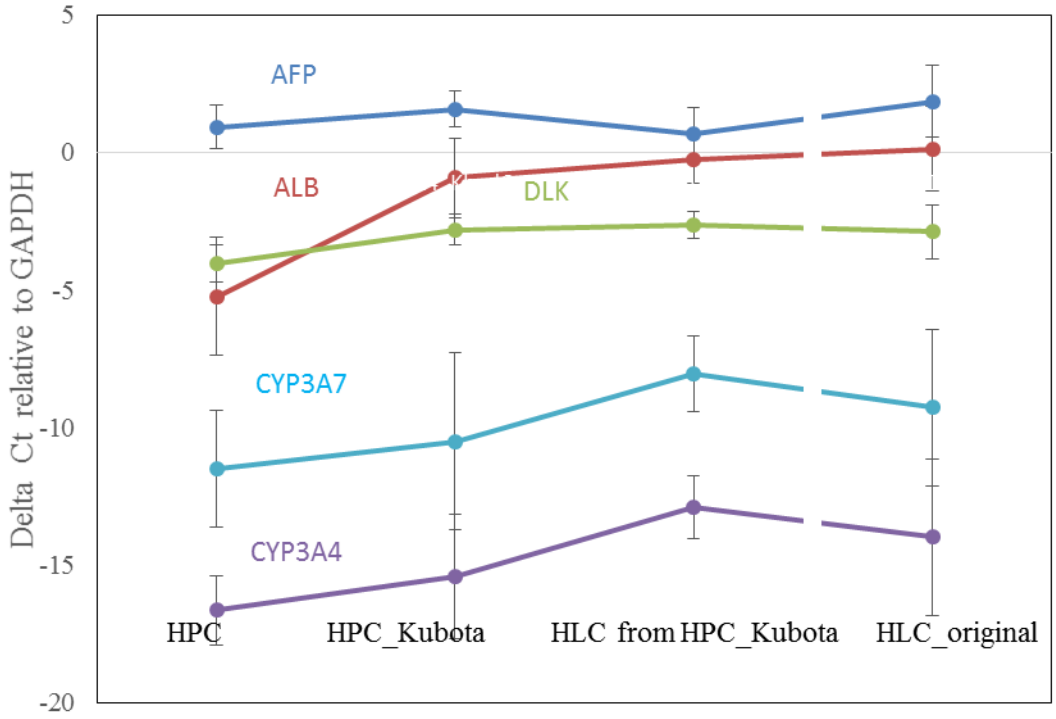


Figure 5-13: The change in hepatic transcripts of Day 14 HPCs through the process of enrichment in Kubota's medium and subsequent differentiation in Stage 4 medium is shown (N=4). The intensities are reported as Delta Ct values relative to a housekeeping gene GAPDH. Higher values indicate higher gene expression. The values for the original HLCs are also displayed and can be differentiated by the break in the curve.

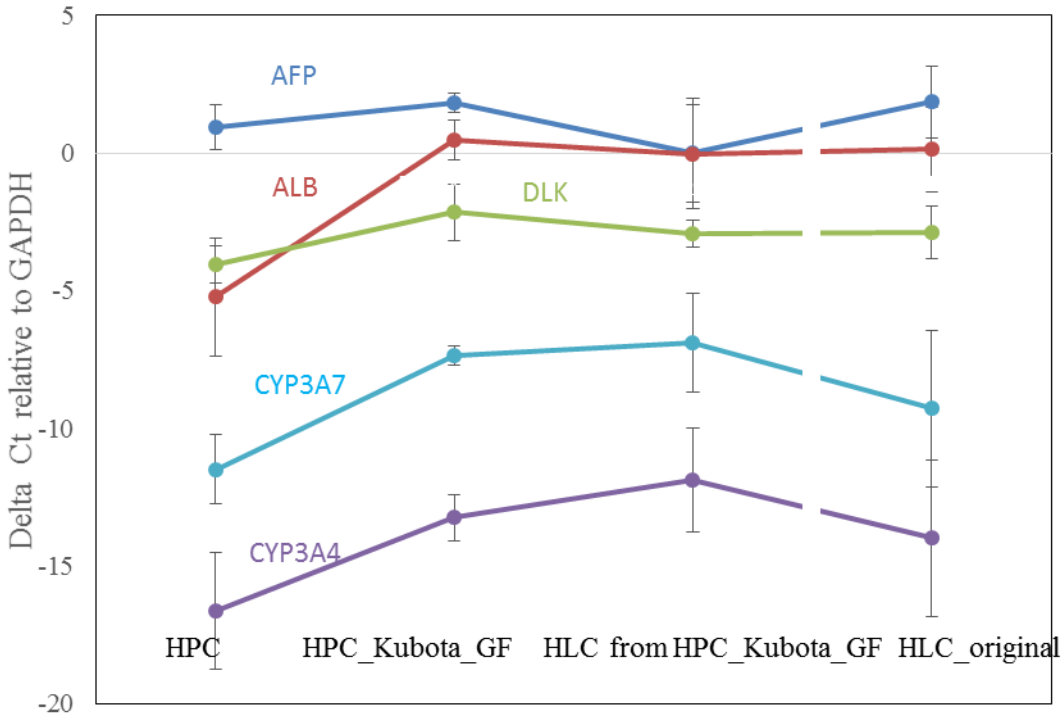


Figure 5-14: The change in hepatic transcripts of Day 14 HPCs through the process of enrichment in Kubota's medium supplemented with HGF and EGF and subsequent differentiation in Stage 4 medium is shown. The intensities are reported as Delta Ct values relative to a housekeeping gene GAPDH. Higher values indicate higher gene expression. The values for the original HLCs are also displayed and can be differentiated by the break in the curve.

6. CROSS-SPECIES TRANSCRIPTOME META-ANALYSIS WORKFLOW

6.1 Microarray Sample Processing

Total RNA was extracted from human and mouse liver differentiation samples at a various time points of differentiation using the RNeasy Mini kit (Qiagen). Human (H9, HSF6) samples were hybridized to the Illumina HT12 bead array v3 (Illumina Inc.), HES3 samples were hybridized to Human Genome U133 v2 array (Affymetrix) and the mouse samples were hybridized to the WG-6 v2 (Illumina) array.

6.2 Sample Collection from Public Depositories

The datasets and the samples, from which the data were acquired, are listed in Table 1. These data, encompassing transcriptome microarray data of different platforms Affymetrix, Agilent and Illumina, were downloaded from the Gene Expression Ontology (GEO) database (http://www.ncbi.nlm.nih.gov/geo). The human cell data included transcriptome of different human embryonic cell lines differentiating to hepatic lineage using different protocols [123,227,228]. Human differentiation data encompassed time series data from two sources, data from HLCs generated using various protocols, including regular differentiation, endodermal cell expansion followed by differentiation and formation of three dimensional spheroids upon HLC differentiation. Additionally, transcriptome data for liver bud generated through co-culturing iPS-derived HLCs with mesenchymal and endothelial cells was obtained [229]. Recently, HLCs generated through direct reprogramming of fibroblasts to HLCs by hepatic transcription factor transduction has been reported as an alternate source to stem cell derived HLCs. This data was also included in the analysis [201].

The mouse liver development data covered developing liver tissues from prenatal (E9.5 to E19.5) and postnatal embryonic development [78,229]. RNA-Seq data for mouse E8.5, E14.5 and adult liver (6 months female) was obtained from (http://www.alexaplatform.org/alexa_seq/Morgen/Summary.htm) [230]. *In vitro* differentiation data for mouse HLCs derived from iPSC differentiation and direct reprogramming of fibroblasts was also incorporated into the analysis [124,125]. Both the

human and mouse datasets consisted of primary hepatocyte and adult liver data that allowed for direct comparison.

6.3 Data Processing

6.3.1 Microarray Data Processing

The raw data of Affymetrix, Illumina and Agilent array platforms were processed using the *affy, lumi and limm*a packages in R respectively to convert to expression intensity values. Logarithmic transformed (log2) intensity from different platforms were corrected for depth of digital precision (16 bit) [231]. The expression level of each gene is condensed by the median expression value of multiple specific probes. Common ENSEMBL identifiers established the same genes from arrays of different platforms. The cross platform data was then combined into a master expression dataset based on matching these ENSEMBL identifiers. Each of the human and mouse datasets were linear normalized. The initial data processing until normalization was performed using the *virtualArray* package in R [231]. The average expression values from replicates were used in further analysis. The mouse and human datasets consisted of 16415 genes and 17683 genes respectively, of which 14,333 were common based on orthologous ENSEMBL identifiers. Transcript intensity values for mouse RNA-Seq embryonic development data were directly obtained from [230] (http://www.alexaplatform.org/alexa_seq/Morgen/Summary.htm)

6.4 Removal of batch dependent effects in combining data of different platforms

To eliminate platform-dependent bias in combining all hESC differentiation data, mouse differentiation and embryonic data respectively, empirical Bayes methods were used as means of location and scale adjustment [232]. The ComBat algorithm in the *sva* package in R was used due to its robustness in handling batches with fewer samples [232]. All the human *in vitro* differentiation was combined into a master set based on matching ENSEMBL identifiers prior to platform correction with ComBat. Datasets from different sources were treated as separate batches, therefore human data from seven sources were

treated as 7 different batches for correction. A similar platform correction was performed on the mouse embryonic as well the *in vitro* differentiation dataset respectively.

The human *in vitro* and mouse *in vivo* datasets had to be combined prior to further analysis. Towards this goal, the platform corrected human and mouse datasets were first combined, based on homologous genes using the *biomaRt* package in R [233]. Batch correction using empirical Bayes methods was performed again on the collective human and mouse dataset by treating data from each species as one batch [234]. The same procedure was followed for combining the mouse *in vitro* differentiation and *in vivo* datasets, except the *in vivo* and *in vitro* data were treated as separate batches. The combined human and mouse dataset consisted of 14,312 genes while the combined mouse dataset consisted of 16,627 genes.

6.5 Data Analysis

Global analysis to elucidate trends in the data was performed using hierarchical clustering, principal component analysis and non-negative matrix factorization. Prior to subjecting the data to these techniques, genes with static expression profiles in that dataset were removed. Only those genes changing greater than fourfold among any sample pair were retained in the global analysis.

6.5.1 Hierarchical Clustering

Hierarchical clustering analysis was performed on the batch corrected datasets of human and mouse separately using the statistical software R and its *hclust* function. Euclidean distance between the expression value of all gene (i) was used as the metric for the distance between different pairs of samples (a and b).

$$\|\text{a-b}\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$$

In a second clustering Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was used as the distance metric.

$$\frac{1}{|A||B|}\sum_{a\in A}\sum_{b\in B} d(a,b)$$ where d is the metric chosen (i.e. Euclidean distance)

6.5.2 Principal Component Analysis

Principal Component Analysis (PCA), a multivariate statistical technique commonly used to reduce the dimensionality of a dataset, was performed using the *prcomp()* function in R. Principal component calculation is based on a singular value decomposition (SVD) of the mean centered gene expression data [235]. The transcriptome data of all sample was organized in n gene x m sample matrix. Let x_{ij} be the expression value of gene i in sample j and A be the n x m mean centered expression matrix, where n is the number of genes and m is the number of samples. The elements of A are thus $a_{ij} = x_{ij} - \bar{x}_i$

By single value decomposition,

$$A = UEV^T$$

Where U is the (n x m) eigenvector matrix of A^TA, E is an (m x m) diagonal matrix and V is the (m x m) eigenvector matrix of AA^T. Elements of U are referred to as eigen vectors and are ordered by their corresponding eigenvalues, which capture the variance of that element. Each eigenvector corresponds to a principal component (PC). The projection of the samples in the principal component space is given by

$$U^TA = EV^T$$

EV^T contains the coordinates or scores for the samples in the PC space, which were calculated using the *prcomp()* function in R. If the first few PCs can capture most of the variance, one can visualize the coordinates of the samples defined by the first few PCs without losing much information. The n x m dimensions of A is therefore reduced to an m x v matrix, where v represents the number of PCs reflecting the largest variance in the dataset. In our analysis, the first two PCs captured about 90% of the variance in all datasets. Therefore, the multidimensional transcriptome data was reduced to m x 2 matrix, where m represents the number of samples in the dataset. The x and y co-ordinates of this matrix are represented by PC1 and PC2, the components which retain the most variance in the data. These co-ordinates were then used to visualize the ordering of the samples in the two dimensional PC space.

6.5.3 Non-negative matrix factorization

Non-negative matrix factorization, another dimension reduction analysis method [236], was used to analyze the batch corrected human or mouse datasets using a NMF R/Bioconductor package [237].

Briefly, A , an n x m transcriptome expression data matrix of n genes of m samples was resolved as a product of two matrices containing only non-negative values W and H of size n x k and k x m respectively. k is the optimal factorization rank or the number of clusters that A can be represented by. This is a critical parameter, its value is usually determined by plotting the cophenetic coefficient corresponding to different values of ranks, and selecting the rank at which the cophenetic coefficient begins to decrease [236]. The samples are then assigned to these clusters by performing several iterations. In the human and mouse NMF analysis, 100 iterations were used to confirm sample assignments.

The elements of matrix W are referred to as metagenes while the matrix H corresponds to the metagene expression profiles. [237]. To evaluate metagenes that contribute most significantly to the biological variance in the phenotypes of interest, they were assigned scores based on a scoring schema outlined in [238] between values 0 and 1. Higher values correspond to larger contributions of the gene to that cluster. A threshold criteria was set and genes above this threshold were designated as significant metagenes [238].

6.6 Developmental time

A two dimensional plot was used to examine how samples at different time points are segregated, by plotting the value of principal components (PCs) 1 and 2 of each sample on a PC1 vs. PC2 plane. The same procedure was used for individual datasets of *in vitro* stem cell differentiation, mouse embryo development and the combination of the two. For each scenario, a second order polynomial curve was fit to the data points of all samples. The arc length with respect to the earliest time point was computed as the developmental time (DT) of each sample using the equation shown below, where x corresponds to PC2 values of samples for that dataset. Developmental time of the human and mouse *in vitro*

data (DT^h, DT^m) was computed with respect to the earliest differentiation time point EN. The development times for the mouse embryonic data (DT^{dev}) as well as the combined *in vitro* and *in vivo* data (DT^{hm}) for human and (DT^{mm}) for mouse differentiations was computed with respect to E9.5 as reference. Each sample irrespective of its origin is assigned an arc length to allow for direct comparison of gene dynamics especially between *in vitro* stem cell differentiation and mouse *in vivo* development.

$$P(x) \simeq a + bx + cx^2$$

and now compute the arc length

$$\int \sqrt{1 + (P'(x))^2}\, dx = \int \sqrt{1 + (b + 2cx)^2}\, dx$$

$$\int \sqrt{1 + (P'(x))^2}\, dx = \frac{(b + 2cx)\sqrt{(b + 2cx)^2 + 1} + \sinh^{-1}(b + 2cx)}{4c}$$

where x corresponds to PC2 , P(x) corresponds to the polynomial fit of the curve in the PC1/PC2 space and a ,b and c correspond to the coefficients of the polynomial function.

6.7 Differential expression of transcriptome of HLCs with E19.5 and primary hepatocytes

Differential expression analysis of human (13 samples) or mouse differentiated HLCs (4 samples) with mature cells was performed using Significance Analysis of Microarrays using the 'samr' package in R. All HLCs were treated as one group and mature cells as another group for this analysis. The mature group consisted of E19.5 from mouse development and species-specific primary hepatocytes. A criteria of false discovery rate (q-value <0.05) and fold change of four or higher in gene expression between the two groups was used to identify differentially expressed genes.

6.8 Dynamic differential gene expression analysis based on developmental time

Limited time series data was available for human and mouse *in vitro* differentiations for dynamic trend analysis. These time series human and mouse differentiation data were aligned with mouse development based on the similarity in developmental times (DT^{hm}, DT^{mm}) respectively. For both species, the culmination of differentiation coincided with E15.5 of mouse development. Hence, gene dynamics of differentiation data from D5 to

D20 was compared with E9.5-E15.5 of embryonic development. Using these DT^{hm} values as reference, the data was separated into a human *in vitro* differentiation dataset incorporating time series data from both sources and a mouse development dataset. To identify orthologous genes whose time dynamic profile is different between human stem cell differentiation and mouse embryo development, the expression profile (i.e. the intensity value of the transcript) of each gene was expressed as a second order polynomial function of the unified developmental time, for each of the differentiation and development datasets individually. The difference of the dynamic expression profile of a pair of orthologous genes was computed in three ways: Pearson's coefficient, Spearman's coefficient and Euclidean distance as shown below. Curves generated using the polynomial functions obtained for human differentiation and mouse development datasets were sampled at fifty equal intervals between the minimum and maximum value of DT^{hm} to compute the different correlation coefficients.

Pearson's correlation where Xij represent expression intensities for gene i and species j

$$r = \frac{\sum(x_{i_m} - \overline{x_{\iota_m}})(x_{i_h} - \overline{x_{\iota_h}})}{\sqrt{\sum(x_{i_m} - \overline{x_{\iota_m}})^2 \sum(x_{i_h} - \overline{x_{\iota_h}})^2}}$$

x_{i_h} is the gene in human and x_{i_m} is the corresponding gene in mouse

Spearman's correlation where Yij represent the ranks of gene expression intensities for gene i in species j

$$r = \frac{\sum(y_{i_m} - \overline{y_{\iota_m}})(y_{i_h} - \overline{y_{\iota_h}})}{\sqrt{\sum(y_{i_m} - \overline{y_{\iota_m}})^2 \sum(y_{i_h} - \overline{y_{\iota_h}})^2}}$$

Euclidean Distance where Xij represent expression intensities for gene i and species j

$$d = |(x_{i_m} - x_{i_h}| = \sqrt{\sum_{i=1}^{n} |x_{i_m} - x_{i_h}|^2}$$

Genes for which the Pearson's or Spearman's coefficient is < -0.9 were identified as candidates whose expression dynamics between human *in vitro* differentiation and

mouse embryo development follow opposite trends; while genes for which the Euclidean distance is > $\mu+2\sigma$ (where μ and σ are average value and standard deviation of Euclidean distance), were identified as candidates of dynamically differentially expressed genes. Genes with expression values changing fourfold in both human and mouse data were further shortlisted as these represent highly expressed dynamically different genes. A Matlab program, Time view, was then used to plot the expression profiles of each gene for visual inspection to confirm the differential expression profile [239].

6.9 **Functional Analysis**

Relevant KEGG pathways for were downloaded to provide functional context to differentially expressed genes. Additionally, gene lists from published data encoding for 1391 transcription factors [240], 400 cellular transporters [241], 3700 genes on the cell surfaceome [242] were incorporated to group the data into functional groups. Additionally, data visualization and functional analysis were performed using Spotfire (Tibco), Database for Annotation, Visualization and Integrated Discovery (DAVID) and Ingenuity Pathway Analysis (IPA) (Qiagen). Functionally enriched clusters from DAVID with enrichment scores greater than 2 were considered significant. DAVID was also used to predict representative pathways as well as transcription factors (TFs) regulating the differential expressed genes via TF binding site prediction from the UCSC Genome Browser (http://genome.ucsc.edu/ENCODE/).

6.10 **MiRNA Analysis**

Microarray assay was performed for D6, D14 and D20 (HLCs) using a service provider (LC Sciences). The miRNA intensities were normalized using a LOWESS filter (Locally-weighted Regression) by the provider [243]. The data consisted of 2615 probes targeting all the human miRNAs present in the latest miRNA depository miRBase 21. Since miRNAs are global regulators, differential expression analysis was performed by imposing a high cutoff for tenfold change during differentiation and requiring that miRNA intensity in at least one sample was greater than 1000.

6.10.1 mRNA-miRNA Target Analysis

mRNA-miRNA interaction information was downloaded from various databases mirTARBase[244], miRDB(v5) for the genes with differential dynamics between human and mouse. Of these, the interaction information of miRNAs from miRDB v5 was based on the most recent miRNA annotation release. miRDB is based on support vector machine (SVMs) approach using high-throughput experimental data to train the target prediction model and using that to predict genes downregulated by miRNAs[245]. Scores for each interaction was calculated from the target prediction algorithm and ranged from 50-100; values above 80 were recommended for high confidence interactions. Hence miRNA-mRNA interactions with miRDB target prediction scores greater than 80 were used in the analysis. mirTARBase and IPA databases consisted of a collection of experimentally verified miRNA-target interaction information. The target genes for highly differentially expressed miRNAs were therefore compiled from all the above sources.

The interaction of a miRNA with its target mRNA is valid if both are expressed with opposite dynamics, i.e one increases while the other decreases. miRNA microarray data was combined with the transcript intensities of differentially expressed genes from the previous analysis to investigate miRNA-mRNA interactions. A fourfold change in miRNA or mRNA intensities in the opposite direction was considered a valid interaction.

6.10.2 Cross species miRNA comparative Analysis

RNA-Seq data for mouse liver developmental stages E8.5 and E14.5 consisting of both transcript and miRNA information was obtained from Hoodless et al [230]. This study was based on older version of miRBase 18, hence only 200 mouse miRNAs were available for comparison. This was consolidated with the human differentiation transcriptome and miRNA data using the conserved miRNA nomenclature across species for comparative analysis. Each species was first individually surveyed for valid miRNA-target interactions using the procedure described above. Those miRNAs exhibiting contrasting miRNA-mRNA regulation in either species were shortlisted as potential targets for intervention.

7. UNVEILING THE ROAD BLOCKS OF STEM CELL DIFFERENTIATION THROUGH CROSS SPECIES TRANSCRIPTOME DATA META-ANALYSIS

7.1 Introduction

The prospect of deriving functional cells from pluripotent stem cells has raised hope for cell therapy to treat various ailments [181]. Using different combinations of growth factors in stages to mimic the chemical environment leading endoderm commitment, liver specification and maturation, stem cells can be directed to differentiate to the hepatic lineage [182,227]. Those differentiated cells resemble hepatocytes in many ways, they secrete albumin, have some cytochrome P450 activities, and are capable of synthesizing urea. However, those cells derived from directed differentiations of embryonic stem cells (ESCs) and other stem cells are still marked by functional immaturity, many mature liver markers are either not expressed or are expressed only at very low levels. They are often referred to as hepatocyte-like-cells (HLC) because they still lack many mature liver functions [246].

Many approaches have been explored to enhance maturation of HLC, including the formation of tissue-like 3D structure [177] and co-culture of HLC and endothelial cells [229], small molecule screening [247] and transfection with transcription factors [131,248,249]. When HLC were transplanted into animals under selective conditions, engraftment was observed [122,229]. Using a cell aggregate formed by co-culturing HLC and endothelial cells for transplant, some mature liver markers were seen in some transplanted cells [229]. However, these HLCs are still immature in most of their functional capabilities when compared to their primary counterparts and adult liver.

Comparison of transcriptome data of differentiation to embryonic development can elucidate the genetic roadblocks preventing stem cells from reaching the functional maturity of their tissue counterparts. While transcriptome data for *in vitro* stem cell differentiation is easy to obtain, human *in vivo* developmental transcriptome data is hard to come by. Such *in vivo* data may be obtained from other species. A cross-species comparison of human stem cell *in vitro* differentiation data with rodent *in vivo* liver development data may reveal the deviation of the *in vitro* process from that *in vivo*, with

the hypothesis that the embryonic development in rodents and in human bears a high degree of conservation in gene expression dynamics.

Human gestation occurs over 280 days compared to 20 days in mouse. A typical directed stem cell differentiation process for both human and mouse ESCs lasts for about 20 days [250,251]. Although these processes differ in time scales, they share similar progression through different stages (Figure 7-1). If the dynamics of gene expression of the *in vitro* and *in vivo* processes of the two species bear a high degree of similarity, it should be possible to identify cells of corresponding stages through cross-species meta-analysis. Such a comparison may provide us crucial insights on possible gene targets for intervention that can enhance the maturity in stem cell differentiations.

Systematic variations and random errors in the assay of samples of different sources often bear the characteristics of the course. Frequently samples of different sources, or of the same source but attained at different time are assayed with different measurement platforms and also bear the characteristics of the platform. These sample source and assay platform derived system variations from multiple batches of data need to be removed before those data can be combined and analyzed. The ComBat algorithm [232] is an empirical Bayes method that performs a location and scale adjustment by pooling information across each gene in every batch and using this information to center the data to an overall grand mean. This technique can be applied to multiple batches and is robust for small sample sizes [232]. ComBat consistently outperformed the common batch correction techniques, including distance-weighted discrimination (DWD) [252] which uses a support vector machine based approach, mean-centering (PAMR) which relies on gene based analysis of variance [253], and surrogate variable analysis (SVA) which uses a combination of single value decomposition and linear model analysis [254-257].

In this study, we employed the Combat algorithm to integrate transcriptome data from human stem cell differentiation and mouse embryonic liver development obtained using different assay platforms. The batch correction allowed data of mouse embryonic liver and HLC derived from human ESCs to be clustered according to their degree of maturation. We identified the genes whose expression profiles in stem cell hepatic differentiation deviate from of embryonic liver development and may contribute to

blocking further maturation of HLC in stem cell differentiation. Such meta-analysis of highly heterogeneous transcriptome data from mouse and human, *in vivo* and *in vitro*, provides clues of genetic intervention, either suppression or overexpression to advance directed differentiation of stem cell to hepatic and possible other lineages.

7.2 Results

7.2.1 Compilation of Human *in vitro* hepatic differentiation Data

The transcriptome data of the differentiation of human stem cells to the hepatic lineage were assembled from many published studies and consisted of 101 samples. Included in the datasets are embryonic and induced pluripotent stem cell differentiation using similar but somewhat different protocols [123,166,227,228]. Two of the studies included data from various time points during the stage-wise differentiation; others presented the transcriptome data of final differentiated cells (HLCs). One study performed a comparative study between HLCs and HLCs generated from endodermal cell expansion and subsequent differentiation (Hu et al, in preparation). Another study utilized formation of three-dimensional spheroids to enhance maturity [177]. Transcriptome data for all HLCs and HLC derived spheroids was also obtained. All have shown that the resulting cells exhibited key hepatic functions. Another study employed co-culture of endothelial cells, mesenchymal cells and iPSC derived HLCs, which were termed as an organoid or liver bud [229]. These are shown in Table 7-1. However, all studies generated hepatocyte-like cells (HLC), which, although expressing hepatic lineage markers, still lack mature liver functions such as key energy metabolic genes (glucokinase, aldolase B) and cytochrome P450 genes. Also included was the transcriptome of differentiated cells derived from hepatic lineage reprogramming of fibroblasts, primary human hepatocytes (pooled mRNA after 1 day of culture), fetal liver (18 weeks of gestation) and adult liver. The data from different sources were derived using various platforms with different probe design and gene coverage. A total of 17,683 genes were commonly present in all arrays were combined into human hepatic differentiation dataset used in this study.

7.2.2 Batch correction of human *in vitro* data

The transcriptome data from various sources cannot be directly compared. To remove the systematic bias of different sources and platforms batch correction was performed as

described in the Materials and Method section. The effect of batch correction is vivid by examining the clustering of samples in the entire dataset. Without batch correction, data obtained from the same studies clustered together as seen from Figure 7-2. After batch correction, data of samples obtained from similar stages of differentiation from different sources, clustered together (Figure 7-3). All the differentiation protocols employed growth factors to guide stem cells first to differentiate to endoderm stage, then to hepatic progenitors, and finally to hepatocytes. Most protocols treat cells for about 20 days using similar but not identical differentiation protocols. It can be seen that the time course data from two different sources clustered based on their differentiation day indicating a similar progression through differentiation. Data from Duncan et al for day 5 (D5_D) and day 10 (D10_D) clustered with our data for day 6 (D6_H) and day 10 (D10_H) in spite of being from different sources. Furthermore, most data of differentiated HLCs from all studies cluster in a separate group, with the exception of a few HLC samples clustering near primary hepatocytes. However, the HLC data from all sources consistently clustered separately from tissue samples of fetal and adult liver.

Additionally, NMF analysis was performed on the data to further validate the results. The optimal number of clusters for the data was chosen based on its cophenetic coefficient as described in the Materials and Methods (Figure 7-4). NMF analysis categorized human data into three groups as seen in Figure 7-5, which depicts the average of results from 100 iterations. Additionally, hierarchical clustering on these three categories is shown, further grouping samples based on similarity. The smallest group consists of data from liver and primary hepatocytes, the other two groups consists of transcriptome from early (D6, D10) and late differentiation (D14, D20) stages respectively. Only two samples, the HLCs from [123] and HLC spheroids based on [177] were separately grouped into tissue group instead being grouped together with the other HLC samples. However, from the clustering, it is evident that though these HLCs are assigned to the tissue group, they are separated during hierarchical clustering, indicating a slight but not complete advanced differentiation status. Notably, the grouping of early and late differentiation samples and the separations from tissue samples are similar to the results of hierarchical clustering shown in Figure 7-3.

To further examine the NMF classification scheme we surveyed the 969 metagenes used in the classification as listed in Table 10-1 and used DAVID to probe their functional significance. The number of human metagenes is high due to the number of samples and their inherent variability. The metagenes used to classify the early differentiating cell states were involved mainly in developmental processes, those dictating the late differentiation states were mainly involved in cell differentiation, adhesion, extracellular matrix reorganization, epithelial specification and drug response. The metagenes classifying the tissue group were involved in mature liver functions such as CYP450 drug detoxification, electron carrier activity, carbohydrate metabolism among many others.

Therefore, two independent unsupervised classification methods classified the batch corrected transcriptome data into similar groups. Data from different sources were categorized based on their functional relevance. This gives credence to the data processing method we adopted to conduct the meta-analysis.

7.2.3 Alignment of human *in vitro* data along a differentiation scale

The batch corrected transcriptome data of human hepatic differentiation was subjected to PCA. Two principal components, PC1 and PC2, captured 90% of data variance, suggesting that the two components are sufficient to display the variability of transcriptome data of those samples. PC1 and PC2 of each sample were then plotted on a PC1 v. PC2 graph. Interestingly all samples, starting from endodermal cells to HLCs were aligned along an arc while primary hepatocytes and adult liver laid further out to the high PC2 region (Figure 7-6). The two data sets that included differentiation data of different stages both align from left to right in order their time duration of differentiation. The samples of iHEP and the liver bud aligned with earlier stages of D10 and 14. Interestingly all HLC samples, regardless of the source, all align within a narrow region in the principal component space, suggesting that they all had similar degree of hepatic maturity. The results also suggest that the distance along the arc from the initiation of differentiation can possibly be used as a metric of human embryonic stem cell differentiation toward hepatic lineage.

7.2.4 Alignment of the transcriptome dynamics of rodent fetal liver development

Mouse embryonic development transcriptome data (E9.5 to post-natal) were compiled from two sources [78,229]. $CD45^{-}Ter119^{-}$ liver cells were isolated from C57/BL6/Tg mice embryos at different stages of development in the first study while whole embryonic livers were dissected from C57/B6 mice embryos in the second study. RNA samples for expression analysis were prepared by pooling several embryos at the same stage of development in both these studies. The first study consisted of samples E9.5, E10.5, E11.5, E13.5, E15.5, E17.5, E19.5 while the second study consisted of samples E11.5, E12.5, E13.5, E14.5, E15.5, E16.5, E17.5 and E18.5 along with some postnatal samples.

The data processing pipeline described for human dataset was used to process the dataset of genome-wide transcript profiles during the development of mouse embryonic liver. The batch corrected data of 47 samples from 2 studies using two strains of mice, and with 16,415 genes common in all samples, were then subjected to hierarchical clustering, NMF and PCA. Similar to the observation made with the human dataset, batch correction removed the effect of different platforms and sources; the transcriptome data, which were clustered largely by their sources and platforms without batch correction, were clustered according to their developmental stages (Figure 7-7). Furthermore, the samples now clustered based on the level of their maturity into early (E9.5-E14.5) and late (E15.5-E19.5) stage (Figure 7-8) upon hierarchical clustering. NMF also classified the batch corrected dataset into two groups, early and late development, which was identical to hierarchical clustering results (Figure 7-10). The optimal value of two clusters was determined based on the cophenetic coefficient shown in Figure 7-9. Only E14.5 and E19.5 were reversed in their order with E13.5 and E19.5, which could be a result of heterogeneous sample preparation protocols in both studies.

This technique also identified 129 metagenes that contribute significantly to the difference between the two states [21] and thus contain key liver development information. Indeed these metagenes belonged to important liver functional classes including development, carbohydrate metabolism, cholesterol homeostasis, response to drugs and urea cycle. This classification agrees with prior understanding of liver

development that the E14.5 stage represents a transition from primarily hematopoiesis to hepatocyte maturation [39, 40]. These metagenes are tabulated in Table 10-2. Of these 129 mouse metagenes, 96 were common with those from the human analysis, mainly overlapping with metagenes classifying mature cells or tissue and not HLCs. This hints that the HLCs correspond to an early state of liver development.

The first two principal components of each sample from PCA are plotted. In this case, samples, in order of their developmental stage line up with increasing PC2 (Figure 7-11). PC1 and PC2 essentially can constitute a developmental vector describing the stage of hepatic development. Identical results were seen with clustering, NMF and PCA analysis.

7.2.5 Integration with mouse RNA-Seq data

RNA-Seq data for mouse embryonic stages of endoderm (E8.5) and hepatic progenitor cells (E14.5) along with adult liver (6 months female) from [230], was integrated with the mouse development microarray data to add additional insights to the analysis. The data was integrated based on ENSEMBL identifiers and the platform correction using ComBat was applied. PCA was used to verify the overall data trends and the developmental stages and adult liver from both microarray and RNA-Seq datasets fall on the same developmental trajectory thereby validating the approach (Figure 7-12). While transcript intensities of individual samples may not be directly comparable for RNA-Seq and microarray studies, dynamics trends can easily investigated. Hence, this resulted in a comprehensive rodent embryonic liver development dataset to gain insights into mechanism of hepatic fate acquisition.

7.2.6 Compilation of mouse *in vitro* differentiation data

Transcriptome data at different time points of differentiation (D6, D14, D20 and D28) of mouse iPS cells differentiating to HLC based on [250] were obtained. Data from reprogramming of fibroblasts to HLCs by transfection with hepatic transcription factors was obtained from [124,125], and integrated into the mouse *in vitro* differentiation dataset using the same methods used from the previous analyses. The order of the samples in the PC space is shown in Figure 7-13. Similar to the human differentiation data, the samples were ordered by their differentiation state (DT^m) and away from the primary hepatocytes indicating a similar barrier to maturation across species.

7.2.7 Integration of mouse *in vitro* differentiation and *in vivo* development data

The mouse *in vitro* differentiation data was combined with mouse embryonic development data based on common ENSEMBL identifiers. The developmental and differentiation data clustered as separate group upon integration and hence were subjected to another batch correction to allow for meaningful analysis. These combined data were then subjected to PCA. On a space of PC1 and PC2 the *in vitro* differentiation data and *in vivo* development data can be seen to line up along an arc as seen in Figure 7-14. The mouse iPSC differentiation data spanned over the same range as the early stages of embryo development (E9.5 to E15.5), while the data points for further developed embryo continues to spread to the region with a higher value of PC2. The results suggest that stem cell derive HLC differentiated to an equivalent of ~E15 stage and still lack the maturity of fully developed prenatal E19 and postnatal liver.

7.2.8 Alignment of mouse *in vivo* and human *in vitro* developmental paths

Having shown that mouse iPSC *in vitro* differentiation data and *in vivo* embryo development data can be integrated and projected onto a PC1 vs PC2 plane, we next integrated the hESC *in vitro* differentiation data with mouse *in vivo* embryo development data. The batch corrected mouse embryo liver development (E9.5-19.5) transcriptome data was combined with the entire set of batch corrected human transcriptome data. A master dataset with 14,312 genes was generated by their homologous identifiers extracted using the BioMart database. Data from ESCs and fibroblasts were not included, as the mouse embryonic development data started at E9.5, which corresponds to the endodermal cell stage. The merged dataset of 35 samples (replicates averaged yielding nine mouse developmental and 26 human differentiation samples), were again subjected to batch correction using ComBat to eliminate bias of species.

The batch correction on the cross species gene expression profiles transformed the data to eliminate species-specific features. This can be verified by plotting the expression levels of hepatic marker alpha-fetoprotein (AFP) and albumin (ALB) before and after batch correction for mouse development and one of the time series of human differentiation (). Before batch correction the human and mouse data showed increasing trends in gene expression of these hepatic markers. After batch correction, while the expression profiles

of the two species were shifted slightly higher in human and lower in mouse, the increasing data trend for each species is strictly preserved (Figure 7-15, Figure 7-16).

PCA analysis was performed on the integrated dataset of human HLC differentiation and mouse embryo liver development. PC1 and PC2 of each sample of both species was projected into a two dimensional PC space. The data points of mouse and human aligned along the same arc, and lined up in order of their development or HLC differentiation stages respectively (Figure 7-17).

From the alignment of human HLC and mouse embryo liver development data, the plausible corresponding stages in human HLC differentiation and mouse liver development can be postulated. HLC samples of the endodermal stage of differentiation, after stage I and around D6, are aligned with E9.5-E10.5 samples. The majority of fully differentiated HLCs (at the end of final stage of differentiation ~ D18-20 of differentiation) are aligned to the E13.5-E15.5 of development. The results strongly indicate that HLCs derived from stem cells were more similar to the fetal state of ~E14-15 or hepatic progenitor state than more mature hepatocytes. This was the case for all HLCs surveyed regardless of the different differentiation protocols used.

Interestingly HLC differentiation of both hESC and mouse iPSC appear to stop at the same corresponding state E15. This indicated that irrespective of species of origin, or the protocols used, all the HLCs encounter universal roadblocks preventing maturation. The integration of embryonic development with differentiation data across species allows for further investigation of these barriers.

7.2.9 Comparison of HLCs with mature cells

The transcriptome of all HLCs for human (hHLCs) and mouse (mHLCs) were separately compared to E19.5 and primary hepatocytes using SAM, where criteria of $q<0.05$, fold change>four was imposed to identify differentially expressed (DE) genes. For the human HLCs vs. mature cell states, 129 differentially expressed genes were found. Among 129 genes human DE genes, many functional classes including CYP450 drug metabolism, carboxylic acid, amine and lipid metabolic processes and complement and coagulation cascades were identified as enriched by DAVID. The differential expression pattern, in term of hepatic genes and functional classes, is rather similar to those seen when

comparing mouse HLC and mature mouse hepatocytes. From the mouse comparison, 159 genes were found to be significantly differentially expressed, of which 42 genes were common in both analyses. These are listed in Table 10-3. Prominent hepatic genes differentially expressed in both species include metabolic genes G6PC, FBP1 and cytochrome P450 enzymes CYP3A4 and CYP2C9 indicating the lack of the maturity in the metabolic profile of HLCs from both species.

7.2.10 Expression profile comparison on a Unified Developmental Time scale

The alignment of human *in vitro* HLC and mouse *in vivo* embryonic development data on a common platform presents the opportunity to compare the similarities and differences in their gene expression profile. We treated the trajectory or the arc formed by those data points as a common developmental path. Using the first point (E9.5) as a reference point, the distance of a sample from the reference point along the developmental path can be taken as a "unified developmental time scale". Each sample, irrespective of its species or differentiation state, is assigned a developmental time (DT). The DT for samples in each analysis can be calculated to obtain a measure of their differentiated state. DT^{h}, DT^{dev}, DT^{m} correspond the values calculated from the human and mouse *in vitro* differentiations and mouse development respectively. DT^{hm} and DT^{mm} represent the distance along the arc for the combined human and mouse *in vitro* with mouse development respectively. In all the analyses, a higher value of DT indicates a more differentiated status for that sample.

The PC2 of a few human differentiation samples and mouse developmental samples are plotted against their developmental time (DT^{hm}) is shown in Figure 7-18. This allows the transcript level of orthologous genes in human HLC differentiation and in mouse embryonic liver development to be plotted on the same unified developmental time scale and compared systematically.

Among all the human HLC differentiation, transcriptome data only two sets provided coverage over different time points. Both dataset ended at a developmental time (DT) of approximately 0.4 (equivalent to E15.5). We combined the two human HLC datasets, from DT=0 to DT=0.4 into a string of data, and all mouse development data in the same DT range (E9.5-E15.5) into another string. While some hepatic genes (e.g. Albumin)

show similar trends between the human *in vitro* and mouse *in vivo* data, others showed inconsistent trends (e.g.TGFBI) as seen in Figure 7-19 and Figure 7-20). Dynamic differential expression analysis (Pearson correlation, Spearman's coefficient and Euclidian distance) were performed to compare the two strings of data for each of the 14332 orthologous genes. The Pearson measures the linear correlation while the Spearman's coefficient is a measure of the monotonic relation between two variables. Threshold criteria were set and the difference in the time dynamics of expression between human HLC and mouse embryo was confirmed by inspection after the time series profile of the differentially expressed genes were plotted (Figure 7-23).

7.2.11 Functional Analysis and comparative gene expression

Genes whose expression were found to be dynamically differentially expressed between hHLC and mouse *in vivo* development, were 197 (Table 10-4), including 17 Transcription factors, 7 transporters and 33 other cell surface markers. The major functional classes or pathways that these genes represent were analyzed using DAVID [43]. Functional clusters with enrichment score greater two included genes participating in the developmental process, organ development, cell adhesion and communication as shown in Figure 7-21. ECM-receptor interaction, focal adhesion and arginine and proline metabolism were among the pathways represented in the differentially expressed genes as seen in Figure 7-21. This suggests that the surrounding microenvironment may have to be further optimized for hepatocyte differentiation.

Several important developmental transcription factors, including ALX1, CDX2, HAND1, PITX1, MSX2, SOX9, FOXA1 and SNAI2, are prominent among those. ALX1 in involved in forebrain development [258], CDX2 plays a role in lineage segregation of the inner cell mass and trophectoderm [259,260], intestinal fate and is involved in epithelial mesenchymal signaling [261], HAND1 regulates embryonic cardiac development [262], PITX1 and MSX2 participates in limb development [263,264], SOX9 regulates pancreatic development [265] and SNAI2 which regulates epithelial mesenchymal transition [266,267]. FOXA1 plays critical role in pancreatic and renal functional specification [268] while TSHZ1 in involved in pancreatic cell development and maturation [269]. All of those TFs were expressed at high levels in both HLCs studies,

but became low at DT~0.4 corresponding to E15.5 in *in vivo* development, suggesting that a down regulation of those TFs may be an *in vivo* developmental control that has not occurred in *in vitro* differentiation.

Genes encoding cell surface proteins are of interest, as these represent the main gateway for cellular signaling and crosstalk to occur. Typically, these are specialized proteins that can relay information about the outside environment to the cells, thereby eliciting cellular adaptation or apoptosis. The expression of an EMT regulator SPARC, an extracellular matrix secreted factor, activates a downstream transcription factor SNAI2 and regulates CDH3 (P-Cadherin) expression which then induces epithelial mesenchymal transition [267]. The expression of all three genes is upregulated in human differentiation but is downregulated by E15.5 of development. Cells surface receptors Frizzled (FZD2) and TGFBR3, receptors for Wnt and TGFβ signaling respectively, are highly expressed in HLCs at DT~0.44, but low in the *in vivo* sample. This along with the increase in expression of other participants of Wnt and TGFβ signaling such as WNT5A, PITX1 could indicate an untimely activation that might need to be suppressed to achieve further maturity. SLC2A2 (GLUT2), a liver specific glucose transporter, SLCO1B3, a bile transporter and SLC27A5, a fatty acid transporter all increased their expression levels during mouse development but not in HLC at DT~0.44, corresponding to D20 of differentiation.

A quantitative observation of the contrasting dynamics for these genes for differentiation (Hu et al) and development was obtained through plots of gene intensities against the corresponding developmental distance using the TimeView software. For ease of visualization, again the genes were categorized into transcription factors, transporters, cell surface markers. The expression profiles of genes in these categories are shown in Figure 7-22. The efficacy of the correlation analysis is high as seen from the contrasting trends the gene expression profiles between human stem cell differentiation (black curve) and mouse liver development (blue curve). The genes exhibiting stark contrast in their profiles and participating in important developmental or hepatic functions are likely the reason the *in vitro* hepatocytes are lacking in functional maturity and are suitable targets for genetic intervention.

7.2.12 Regulators of differentially expressed genes

The genes with disparate profiles were identified and were found to participate in important biological process of organ development, metabolism and drug detoxification. It is impractical to genetically manipulate about 200 genes, therefore transcriptional regulators, both transcription factors (TFs) and miRNA regulators of these genes were sought after.

7.2.12.1 Transcription Factors

TFs interacting with the genes of interest were predicted based on the overrepresentation of corresponding TF binding sites in the promoter regions of the genes. These interactions were sourced from the UCSC genome browser and mapped to the gene list through DAVID. The top TFs that were found to have overrepresented binding sites in the genes of interest including GATA, FOXO1, CREBPH, GATA6 are shown in Table 7-3: Transcription factors with overrepresented binding sites in the dynamically differentially expressed genes. These transcription factors along with the suppression of the developmental transcription factors can be used to modulate the gene expression of the potential targets.

7.2.12.2 MicroRNA Analysis

MicroRNAs are small non-coding RNAs about 22 bp in length, which play an important role in post transcriptional regulation[270]. These miRNAs can bind partially to the 3' UTR region of the target gene causing its repression. This partial complementarity is responsible for the ability of a single miRNA to modulate the expression of several genes simultaneously. miRNAs are therefore global regulators and have been implicated in the regulation of hepatic fate during liver development [271].

Human miRNA microarray data was obtained at EN (D6), HPC1 (D14) and HLC (D20) of differentiation to add another dimension to the meta- analysis. From the miRNA microarray data, 24 differentially expressed were found. These changed greater than tenfold throughout differentiation and at least one sample had miRNA intensity greater than 1000 and are displayed in Table 10-5. The most abundant liver miRNA, miR-122 increased by 10,000 fold through the course of differentiation [272]. Similarly, homologous miRNA's miR-192/miR-215 increase by 175 and 93 fold by D20 compared

to D6 of differentiation and are known to negatively regulate cell cycle genes [273]. miR-194 which was reported to be marker of hepatic epithelial cells increases by 340 fold [274]. Conversely, embryonic cell specific miRNAs miR-302 and miR367 decrease by 20 fold and 37 fold respectively through the differentiation process [275]. miR-302b is also expressed highly in mouse endoderm (E8.5) and participates in liver development by modulating TGF-β signaling [230].

From the miRNA-mRNA interaction data obtained from different databases, 20 of the differentially expressed miRNAs were potential regulators of 85 dynamically differentially expressed genes from the previous analyses. The table of these miRNAs and their gene targets is shown in Table 10-6. However, when assayed for valid miRNA-mRNA interactions, only miR-367, miR-122 and miR-335 showed opposite expression trends with their gene targets COL1A2, NODAL, CHD10 and IFITM1. These 20 differentially expressed miRNAs could therefore be potential targets for genetic intervention to regulate their target genes in a desirable manner. Overall, these 20 TFs and 20 miRNAs and are good candidates for developing a high throughput screening method to evaluate their effect on hepatic maturity.

7.3 **Discussion**

Stem cells derived hepatocytes have been sought after as a renewable source of cells for regenerative medicine and for *in vitro* toxicity studies. To date HLCs derived from stem cells still lack the functional maturity of hepatocytes. Several approaches have been attempted, including cultivation as 3D structure and co-culture with endothelial cells, and increased expression of some hepatic functions have been reported. Those hepatic differentiation studies employed an array of protocols and assessed the differentiation using different genetic and functional markers, making a direct comparison of their results difficult. With the increasing affordability of global gene expression analysis, public depositories have become populated with transcriptome data of embryonic development as well as stem cell differentiations from different species. We hypothesized that genome wide gene expression profile carries more relevant information on the state of differentiation or development than a few selected genetic or functional markers and set out to conduct a meta-analysis on those transcriptome data. A comprehensive

comparison of the maturity of HLCs derived from those studies and their relative position to *in vivo* fetal liver development can tell us much about the potential road block on *in vitro* differentiation.

The transcriptome data available on fetal liver development were of mouse origin, while those for *in vitro* differentiation were mostly from human and a limited number from mouse. The data were heterogeneous in their sources, assay methods and platforms. It was imperative to remove platform and source dependent effects to allow for meaningful comparison. For removal of those batch effects we chose to employ the empirical Bayes based Combat algorithm that has been reported to be its robustness [254] and has been shown to be efficacious when applied to gene expresion data obtained from different patients of two types of arthritis [256].

Batch correction did remove systematic variations for both human and mouse data sets as illustrated by unsupervised clustering before and after batch correction (Figure xx and xx). PCA resulted in two interesting observatios: first, differentiating HLCs align along their progression in differentiation in a PC1 vs. PC2 plot; second, the final HLCs of different protocols and sources all line up in a region that is well differentiated but at a distance from the mature hepatocytes of adult liver. With the small amount of transcriptome data available for mouse stem cell derived HLCs, the data was combined with mouse fetal liver development. Again, similar two observations were made as in human HLC, except that, in this case the *in vitro* differentiation could be aligned to *in vivo* development. Stem cell derived HLCs appear to cease to differentiate further beyond the equivlanet of E15 in embryo development. The results also suggest that PCs from PCA can be used as a measurement of the developmental stage comparing *in vitro* and *in vivo* data. We examined the identity of the metagenes used in the alignment of both human and mouse samples. They are related to development and liver functions, giving further credence to the biological relevance of the alignment of samples.

To identify gene expression differences between human HLC differentiation and mouse fetal liver development, we need to take one step further to combine human *in vitro* and mouse *in vivo* data. To accomplish that, we treated human dataset and mouse dataset as if each were one batch dataset, i.e. assuming that their transcriptomes were sufficiently

similar but bear systematic variations that can be removed by batch correction. The classification of samples into groups identifiable by their biological characteristics using unsupervised hierarchical clustering, non-negative matrix factorization and principal component analysis gives credence to the data processing scheme we employed in this meta-analysis. The alignment of samples along their PC2/PC1 space placed the differentiated human HLCs and E15 mouse embryonic liver at an equivalent stage. The alignment thus give the *in vitro* and *in vivo* data from the two species a common "scale" of their development.

The development of mammals bears many common morphological, biochemical and genetic hallmarks. Based on those marks, similar stages of development can be identified in spite of the vastly different time scale in different mammals. These are referred to as Thieler stages. For example, in human embryo development 22 day and 52 days are considered to be equivalent to mouse E9.5 and E14.5 respectively[276]. In an early study neural development events were codified and used to generate a regression model for predicting a translation table of times of corresponding stages across nine species ([277]). In another study feature measurements of leaves of different tomato species were subjected to principal component analysis to establish a developmental trajectory ([278]) Recently transcriptome data of nematode species were compared to morphological markers to establish embryo developmental milestones in different species of Caenorhabditis ([279]). A algorithm that takes RNA-seq data of a species' developing embryo and subject them to principal component analysis for alignment in PC plane was reported to describe the developmental stage ([280]).

In this work we extend those prior studies to tackle transcriptome of *in vivo* and *in vitro* samples from different species, and in multiple assay multiple platforms. Important in our data processing scheme is the batch correction that removes systematic variations inherent in the heterogeneous data. By aligning the human *in vitro* and mouse *in vivo* data in the same "unified time scale", the expression of orthologous genes between cells of the same "time" can be systematically compared.

Our metaanlaysis revealed that all HLCs, regardless of their differentiation protocol or species, were closer to early liver development (~E15.5) than to primary adult

hepatocytes. Among the differentially expressed genes between HLCs and mature cells (E19.5, PHH) are key CYP450 enzymes (CYP2C8, CYP2C9, CYP2E1, CYP2D6 and CYP3A4) and metabolic izozymes (ALDOB, SLC2A2 (GLUT2), G6PC and FBP1) that are the hallmarks of liver metabolism. This analysis is in concordance with the study by Baxter et al, where they demonstrate that HLCs were more representative of fetal hepatocytes, especially in the expression of CYP3A4 and CYP2D6 through proteomic and functional analysis [281].

The dynamic profile of gene expression during HLC differentiation was compared to that of early embryo development along the unified developmental time. Gene whose time dynamics follow opposite trend were identified as a pivotal gene set of about 200 genes. These genes spanned across important functional clusters of development and metabolism. Many transcription factors regulating organ development, including HAND1, PITX1, PITX2, MSX1, CDX2, SNAI2 and TWIST1 were highly expressed in early stage but subsided in E15 but remained highly expressed in D20 of *in vitro* differentiations. Genes of cell surface receptors (TGFBR3, FRZB) of the TGFβ and WNT signaling pathways increased their expression during HLC differentiation as compared to *in vivo* liver development, suggesting that these signaling mechanisms, which play key roles in early liver development, may be discordantly regulated in HLC differentiations. Furthermore, TGFβ signaling in conjunction with upregulation of SNAI2 and TWIST1 hint the occurrence of epithelial to mesenchymal transition (EMT) [266].

An upregulation of CDX2, a regulator of intestinal cell fate, was reported during lineage programming of fibroblasts to hepatocytes [282]. A recent study by Godoy et al performed transcriptome comparison of HLCs generated from three labs to different cell lineages using the gene regulatory network analysis using the CellNet platform [283,284] which is currently applicable only to Affymetrix expression platforms. They concluded that all HLCs display a mixed gene signature corresponding to different lineages. Notably, some genes reported by Godoy et al coincided with differentially expressed genes identified in our study. Godoy et al saw a similar increase in expression of intestinal transcription factor CDX2 along with EMT regulator SNAI2. While they compared HLCs with hESCs and primary hepatocytes, our study focused on investigating

gene dynamics of differentiation with early embryonic development to elucidate genes with contrasting trends. Hence, we discerned several unexpected TF's regulating not only intestinal but heart, limb, pancreas development. Such upregulation of TFs responsible for other endodermal and mesoderm cell lineages suggests that the strict control of developmental mechanisms occurring through crosstalk among developing organs in the embryo, is lacking during stem cell differentiation [285,286] . This combined with the previous comparative analysis with mature cells, provides a comprehensive view of genes that warrant intervention to achieve enhanced maturation.

Transcription factors and miRNAs regulators for the dynamically differentially expressed genes were identified. These global regulators and can initiate the path forward for developing uniform strategies for genetic manipulation, to tailor the cell fate specifically towards the hepatic lineage.

In conclusion, an unbiased cross species comparative analysis of miRNA and transcriptome of *in vitro* stem cell differentiations from heterogeneous sources with *in vivo* mouse embryonic liver development was performed. It was demonstrated that HLCs not only represent an immature hepatic cell fate, but a mixed cell gene signature more representative of early development, of which an exact *in vivo* equivalent may not exist. Finally, pivotal gene targets were identified that merit combinatorial genetic intervention to enhance maturation of stem cell derived hepatocytes.

7.4 **Figures**

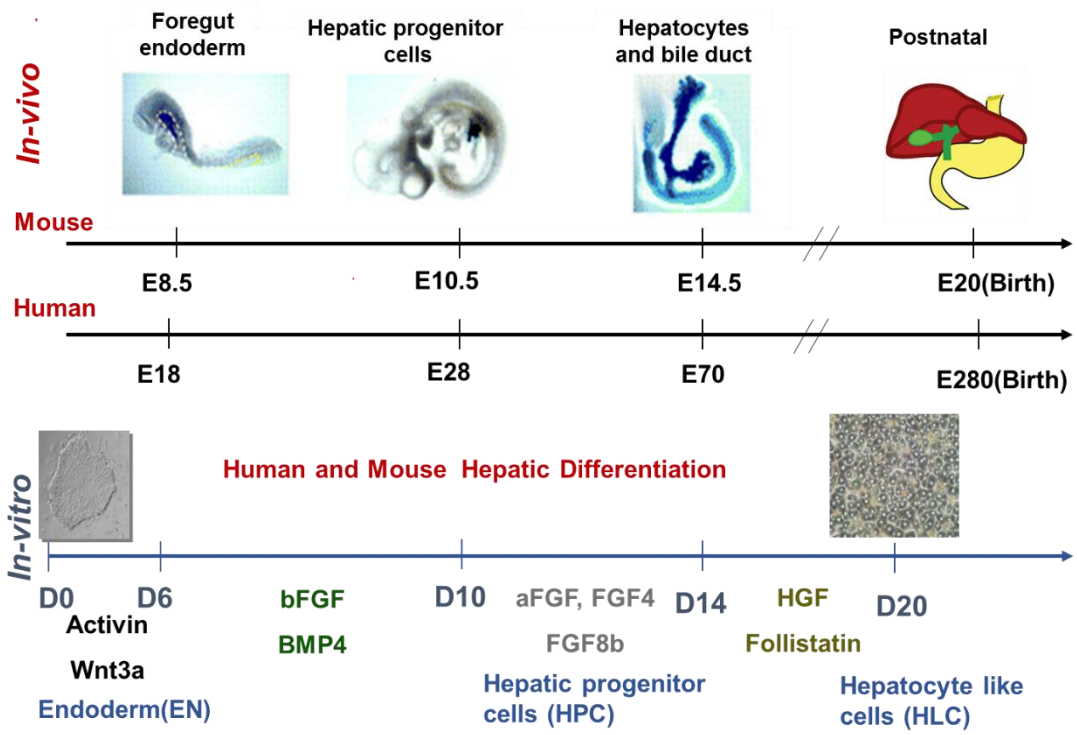


Figure 7-1: The timelines for mouse and human embryonic liver development and *in vitro* stem cell differentiations

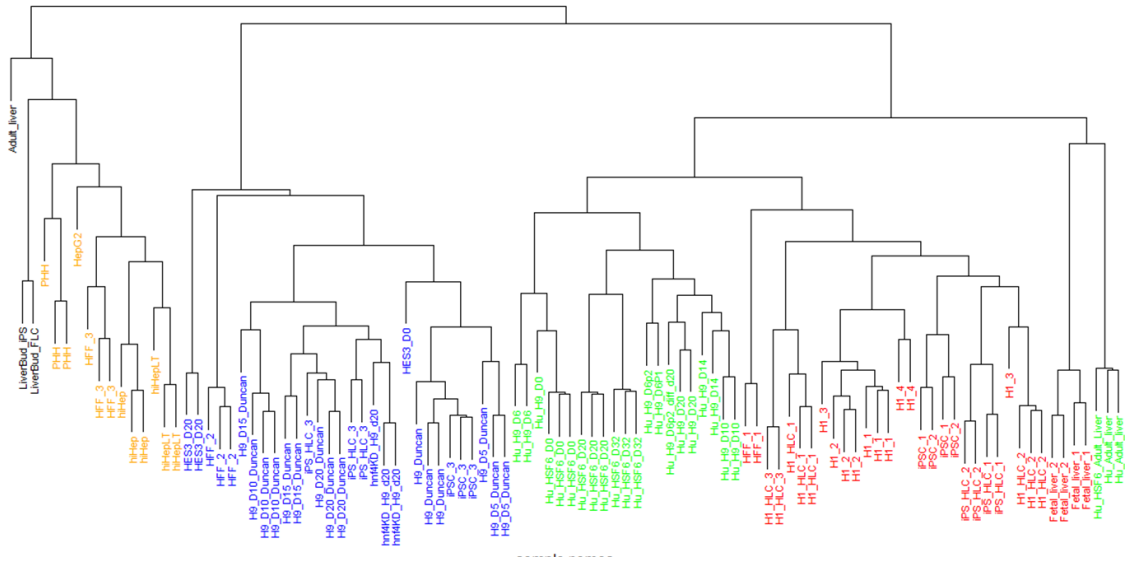


Figure 7-2: Platform dependent batch effect were evident for human *in vitro* differentiation data. The different colors indicate different platforms

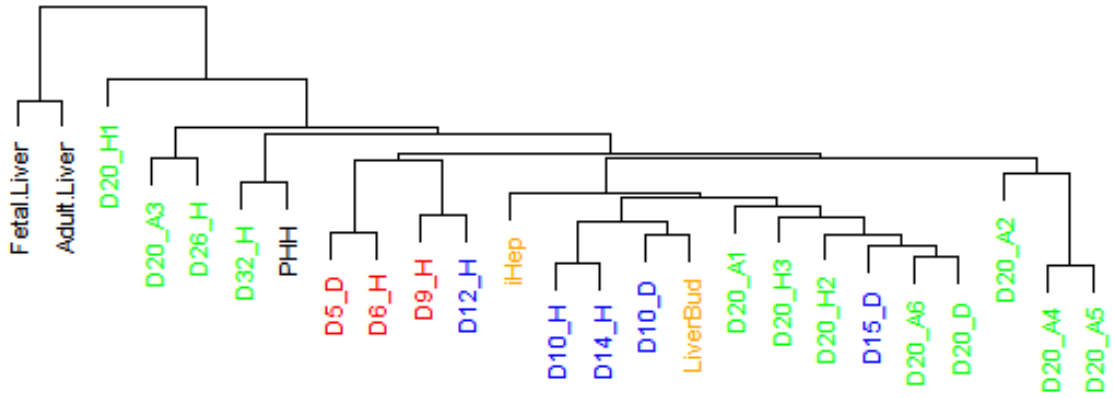


Figure 7-3: Human *in vitro* Data after batch correction. The labels indicate the group the differentiation samples were obtained from and color represents the differentiation state (Endoderm: Red, Hepatic Progenitor Cell: Blue, Hepatocyte like cells: Green and Mature Cells: Black)

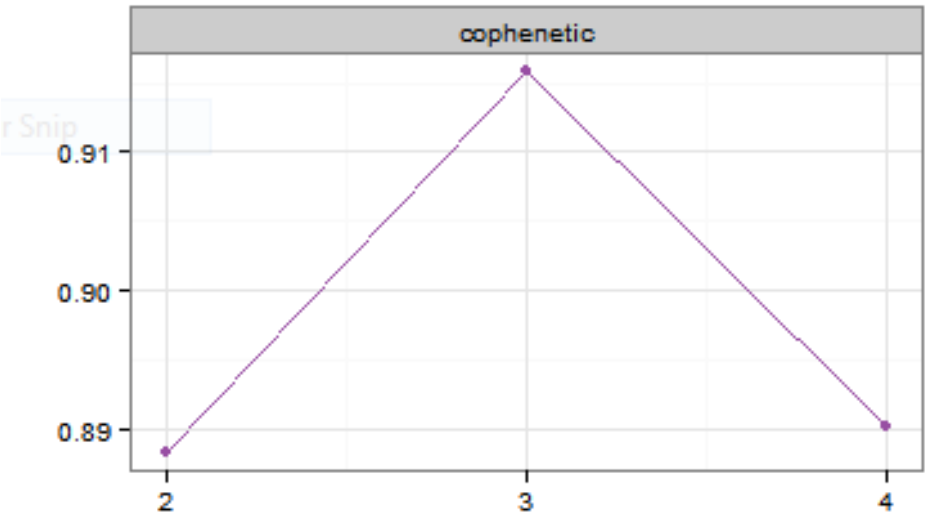


Figure 7-4: The cophenetic coefficient for the human differentiation dataset is plotted for different ranks. Since the coefficient begins to decrease at 3, the data was optimally clustered into three groups

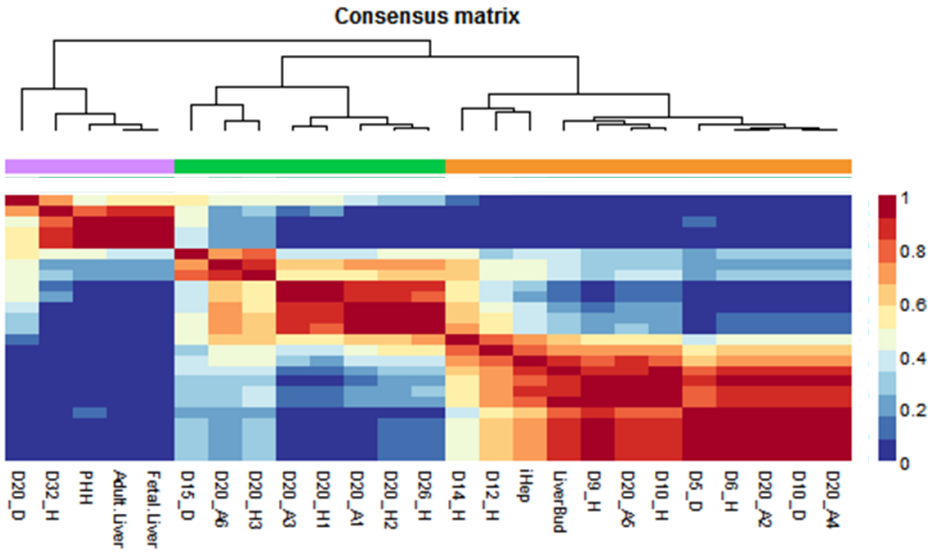


Figure 7-5: NMF on human *in vitro* data after batch correction separates the data into three groups of differentiating cells until D14, HLCs and mature cells and tissue. The consensus matrix shown was generated after 100 iterations. The value 1 (Red) and 0 (Blue) represent the confidence of sample assignment to that cluster. About 950 metagenes were found to classify the data into three groups.

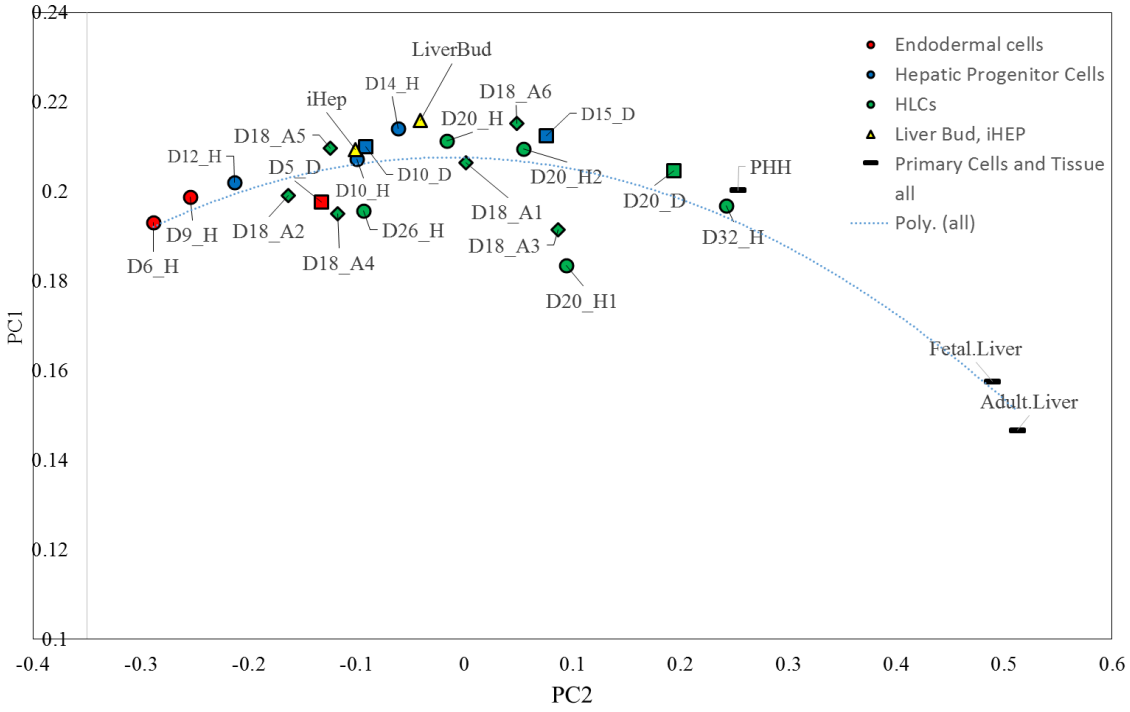


Figure 7-6: The first two principal components represent 90% of variance in data. The samples are arranged according to their differentiation status in the PC space, starting with the endodermal cell flowed by HLCs and mature cells to the far right. The colors represent different stages of differentiation and shapes represent different sources of data

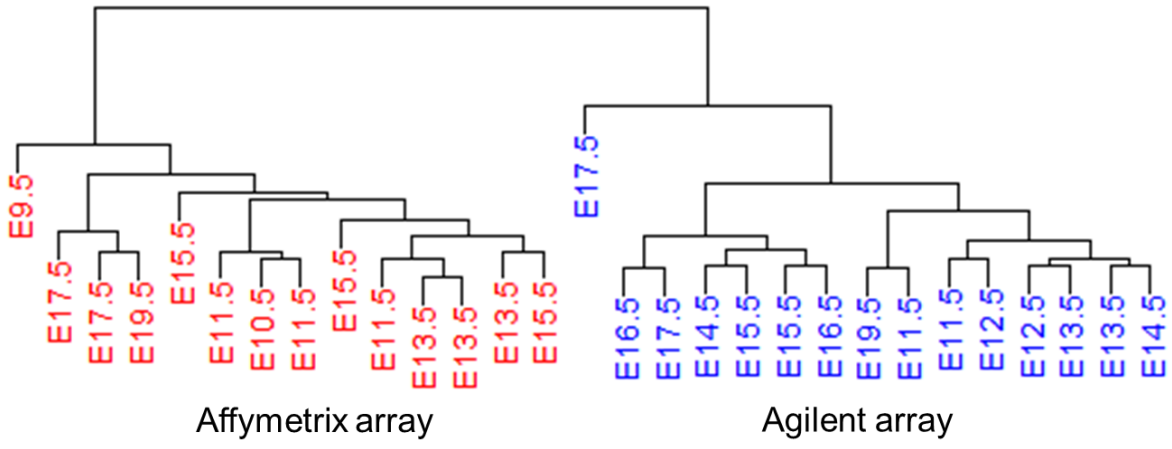


Figure 7-7: Platform dependent effects were seen in the mouse development data where both platforms are represented in red and blue

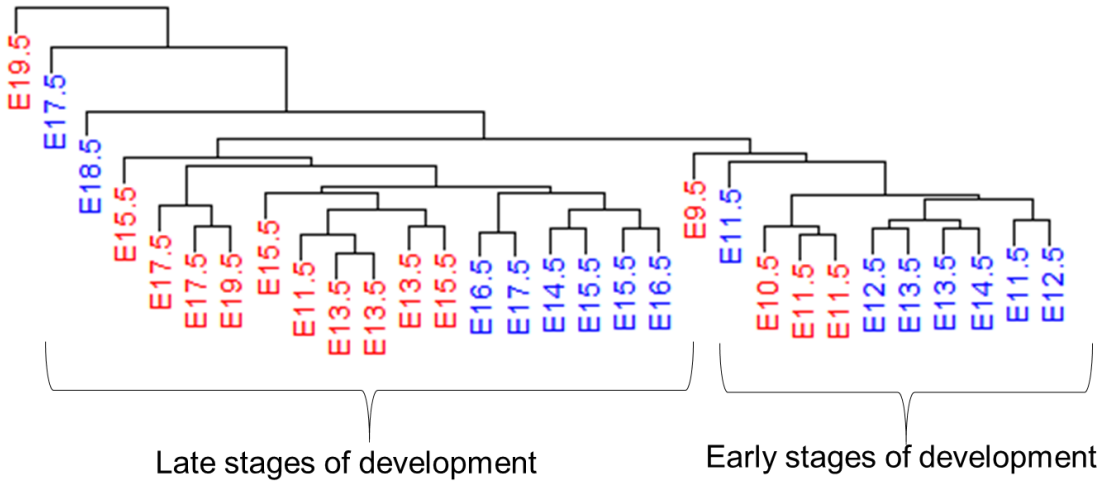


Figure 7-8: Platform correction for mouse embryonic development data separates early and late stages of development. The two platforms are represented in red and blue.

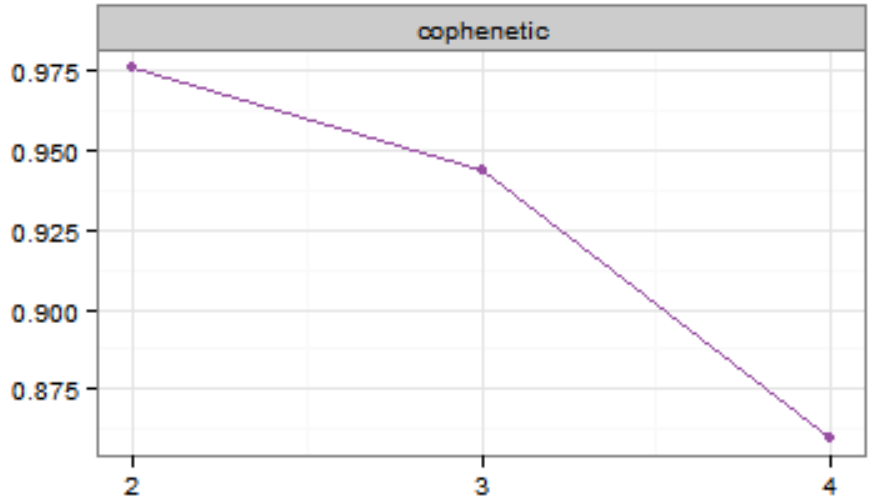


Figure 7-9: The cophenetic coefficient begins to decrease at 2 for mouse development, therefore data was optimally clustered into two groups

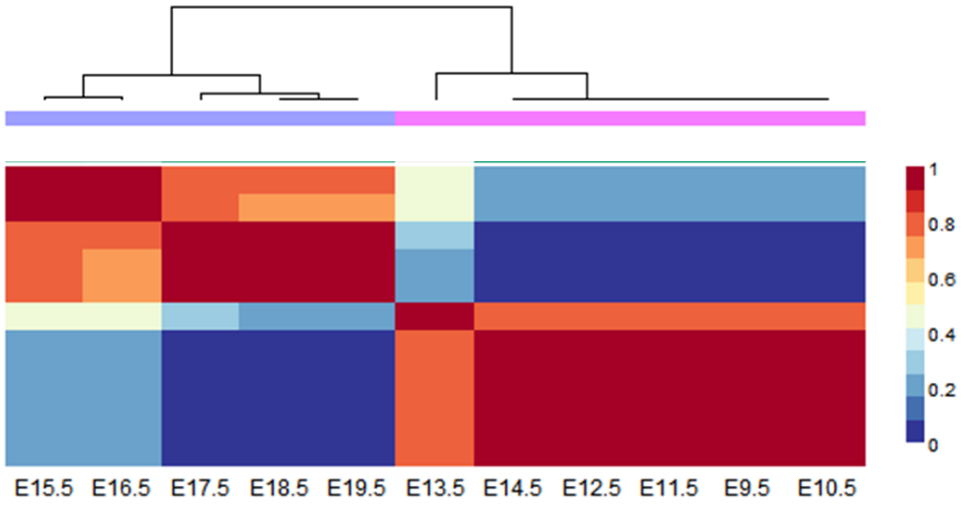


Figure 7-10: NMF optimally clusters mouse data into two groups of early and stages of development. The consensus matrix was generated after 100 iterations. About 120 metagenes were found to classify both the developmental states.

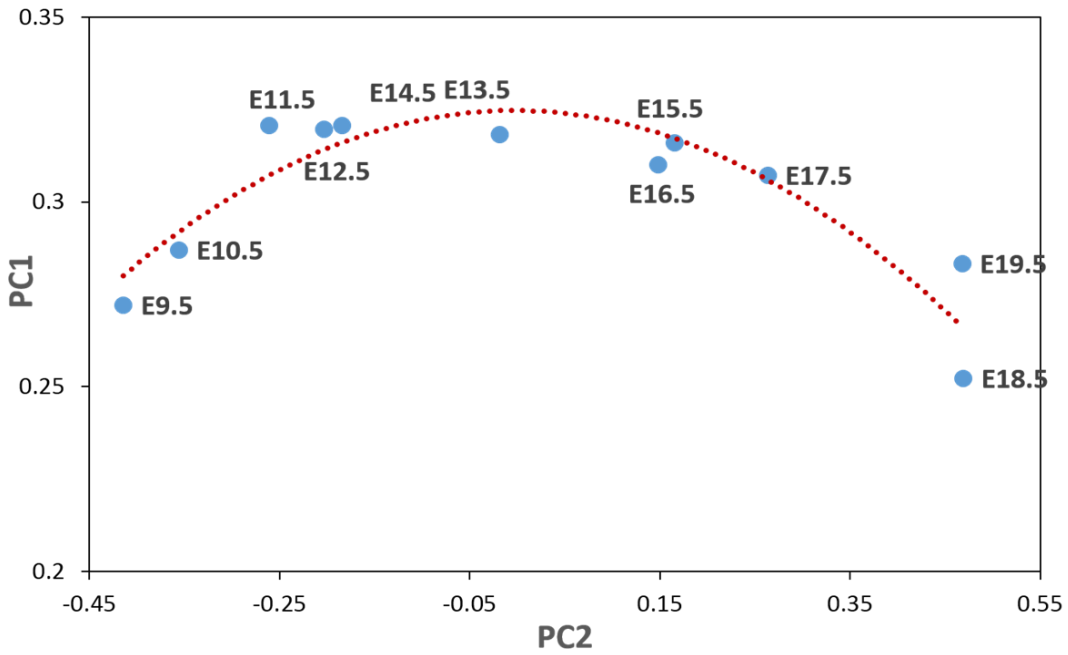


Figure 7-11: After PCA analysis on the batch corrected data, samples aligned according to their developmental stage in the PC space starting with the earliest (E9.5) to latest prenatal stage (E19.5)

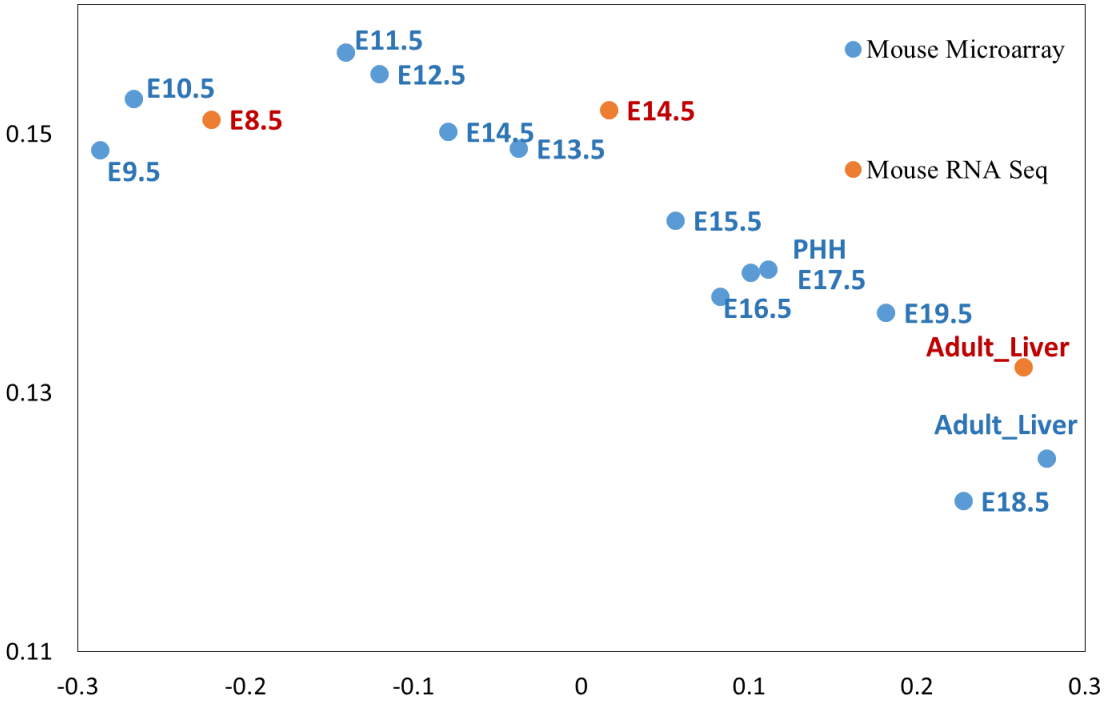


Figure 7-12: Transcriptome data for mouse embryonic development, both microarray and RNA-Seq were combined after correcting for platform effects. The samples align along the same developmental stages validating the technique

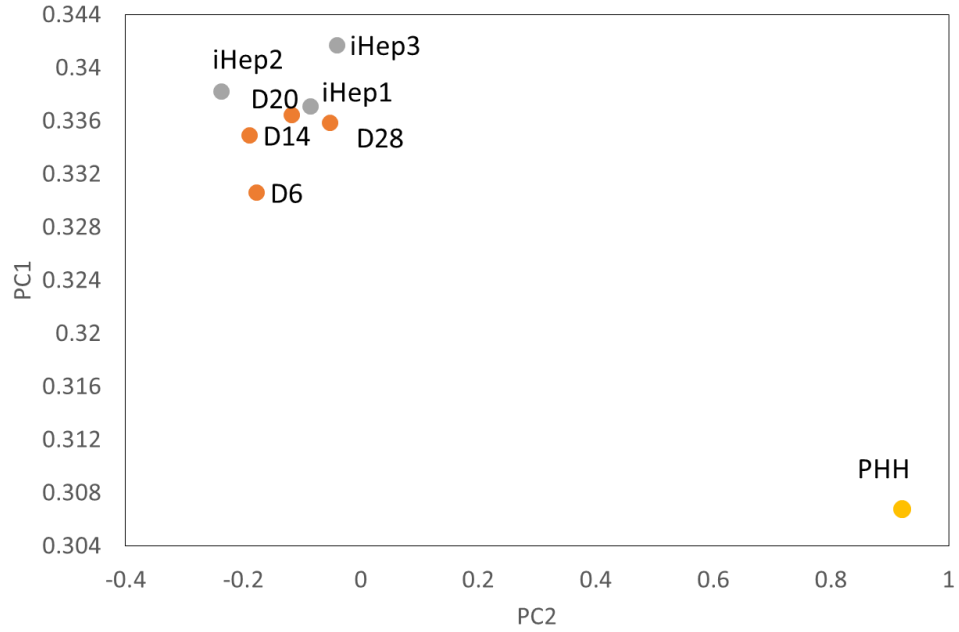


Figure 7-13: Mouse *in vitro* differentiation data was compiled and upon batch correction, the HLCs clustered together in the PC space and consistently away from primary hepatocytes

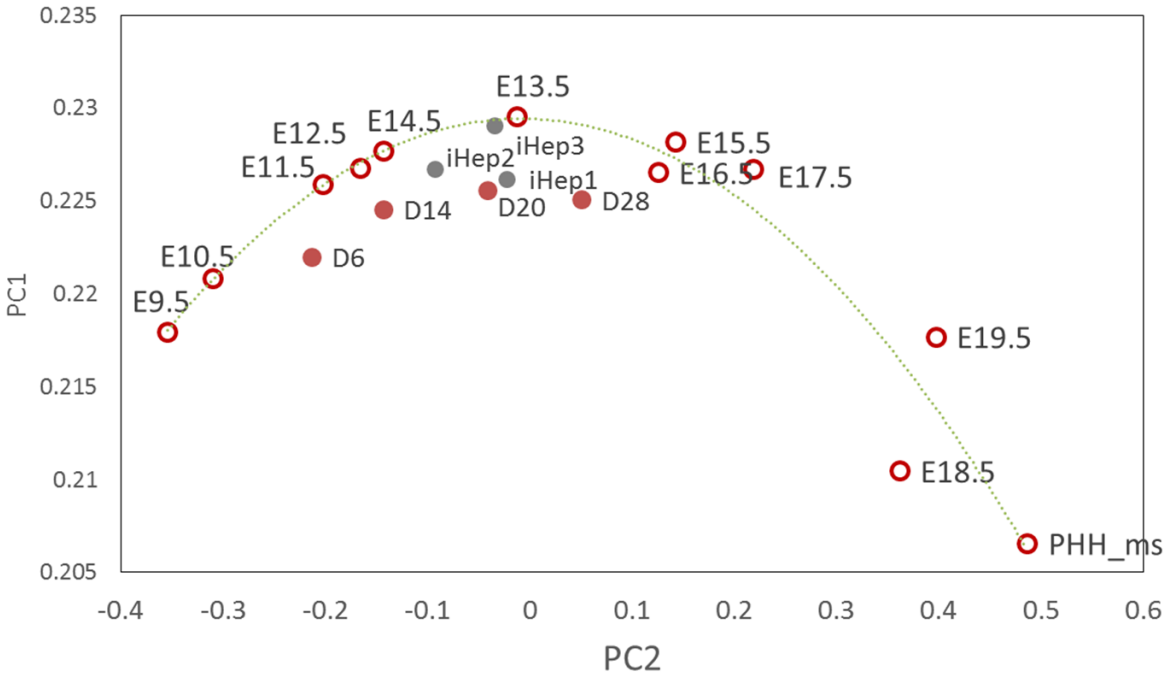


Figure 7-14: Mouse *in vitro* differentiation data was combined with embryonic liver development. After subjecting the entire dataset to PCA analysis, the samples aligned according to their differentiation status along the PC curve. The open circles represent embryonic development while the closed ones *represent in vitro* differentiation All the HLCs lay between E13.5 and E15.5 on this curve while E19.5, primary hepatocytes (PHH) lay in the region with higher PC2 values.

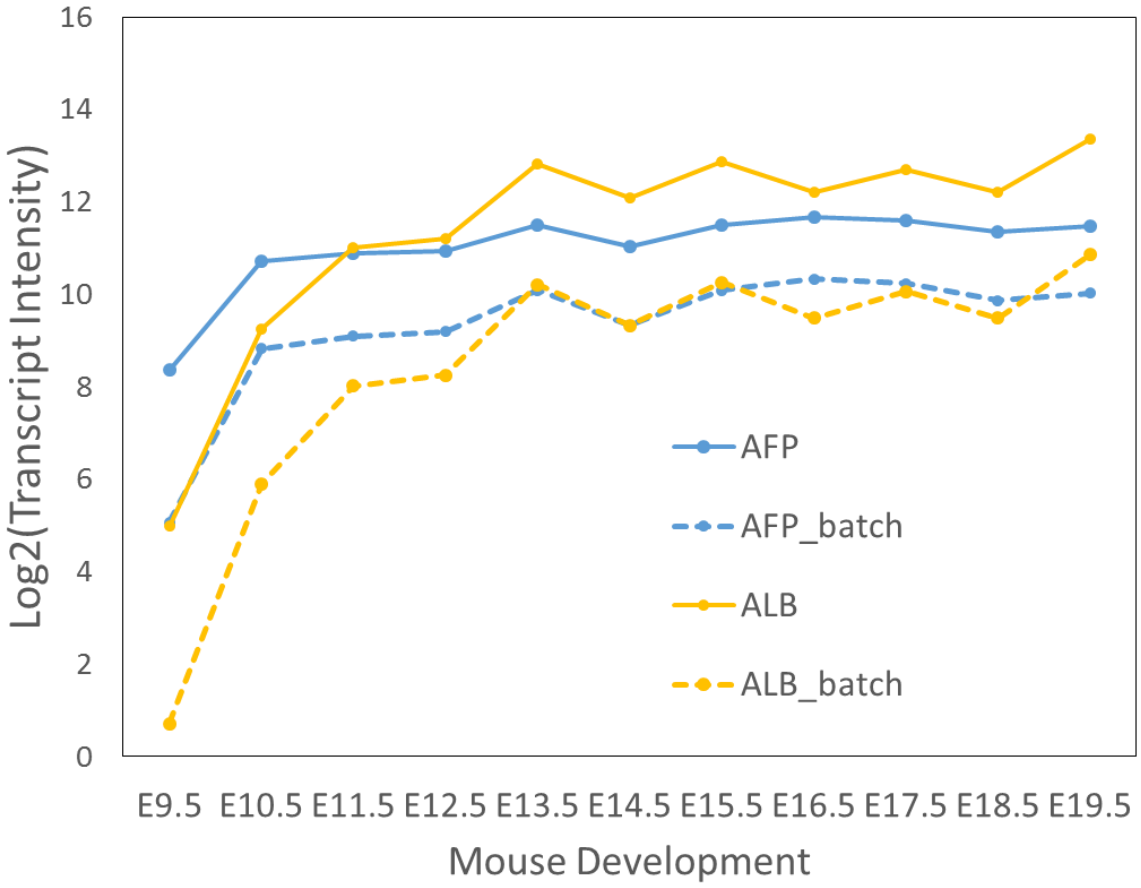


Figure 7-15: The gene expression profile in mouse developmental samples before and after combining the human *in vitro* and mouse *in vivo* data by performing batch correction for species effects. The gene profiles are shifted but are preserved in their dynamic trend.

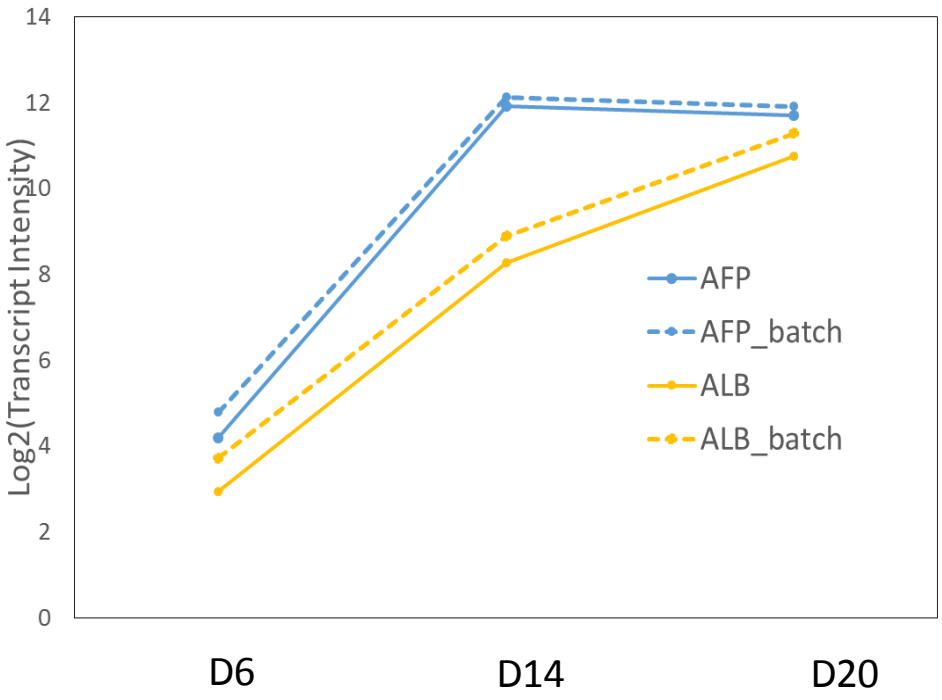


Figure 7-16: The gene expression profile in human differentiation samples before and after combining the human *in vitro* and mouse *in vivo* data by performing batch correction for species effects. The gene profiles are shifted but are preserved in their dynamic trend.

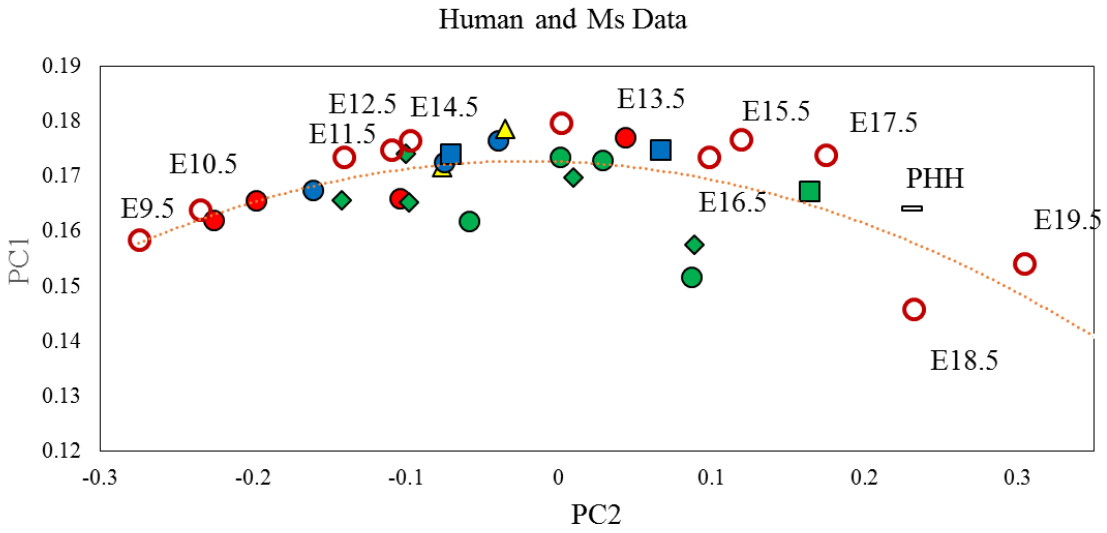


Figure 7-17: Human *in vitro* differentiation was combined with mouse development. Endodermal cells, hepatic progenitor cells and HLCs are represented in red, blue

and green. The circles, squares and diamonds correspond to samples from Hu et al, Duncan et al and Adjaye et al. Most HLCs lie before E15.5 in the PC space.

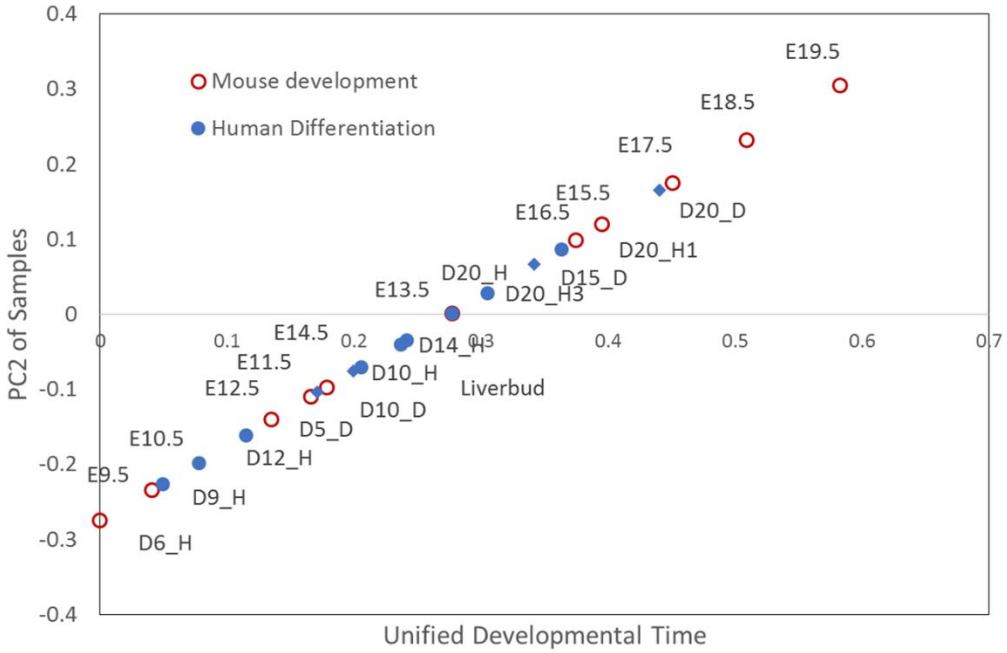


Figure 7-18: The PC2 values for mouse and human samples are plotted against their unified developmental time (DT). The DT increases with increase in maturity, and most HLCs fall within DT corresponding to E15.5

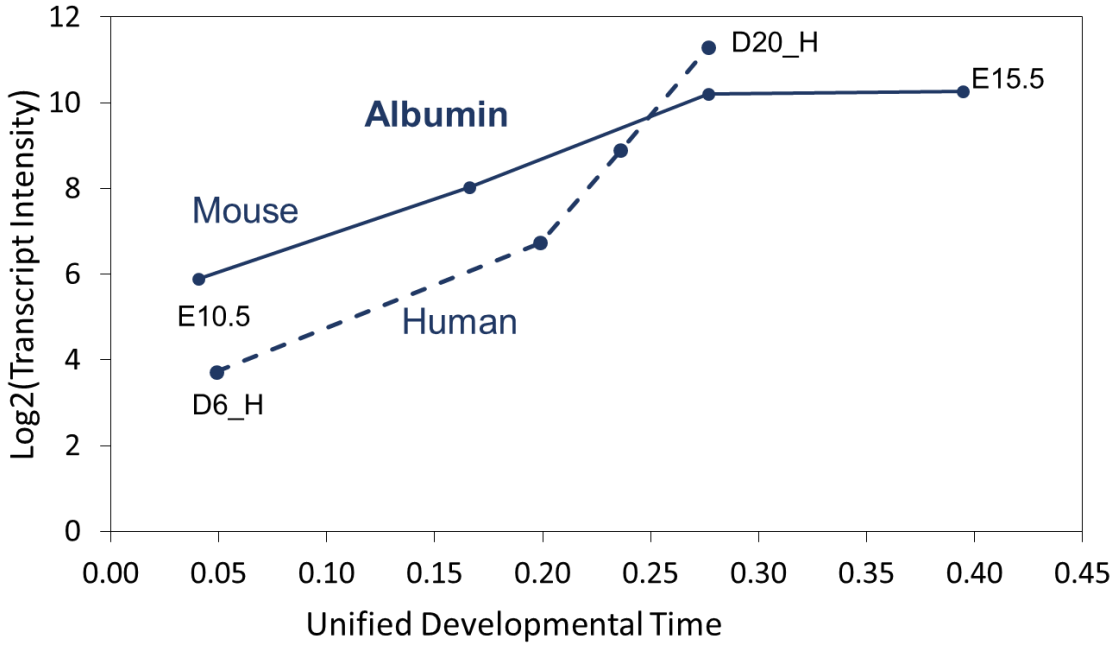


Figure 7-19: Human and mouse gene profiles can be evaluated on the same developmental scale. Albumin follows similar dynamics in development and differentiation

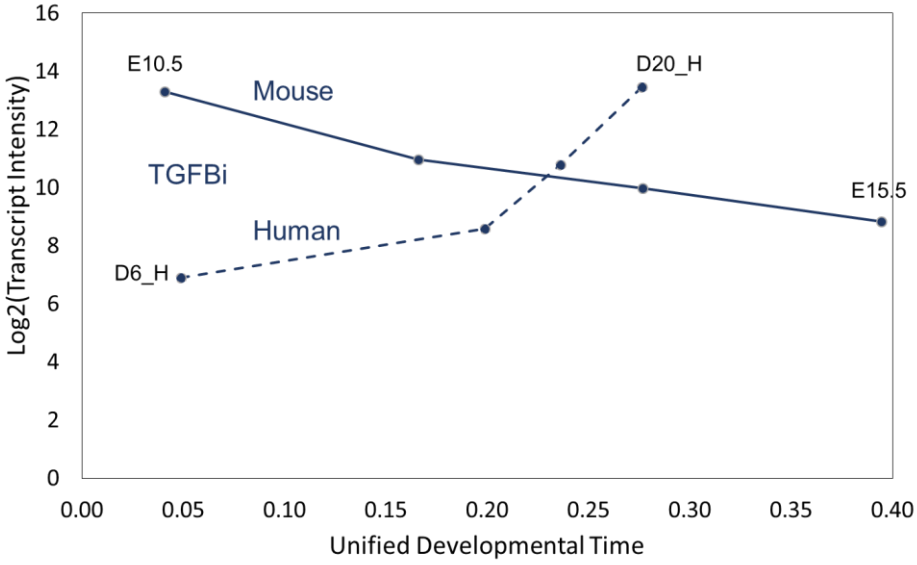
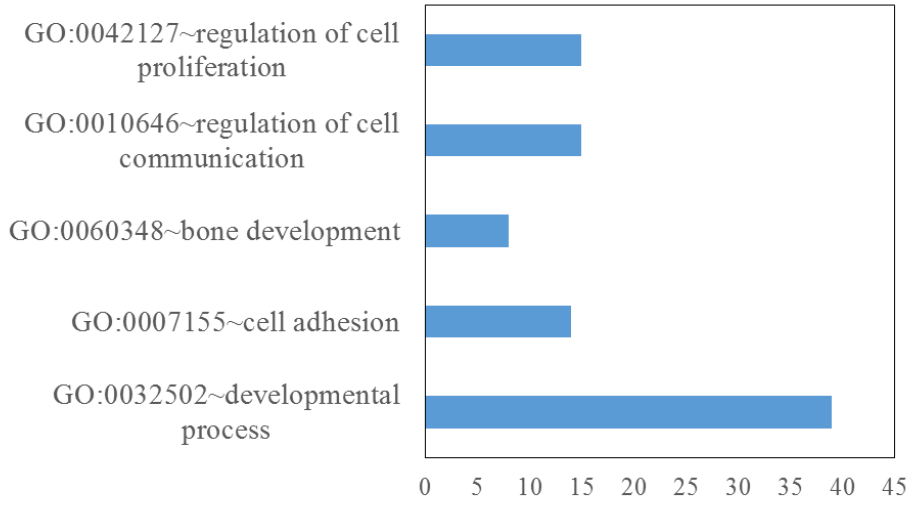


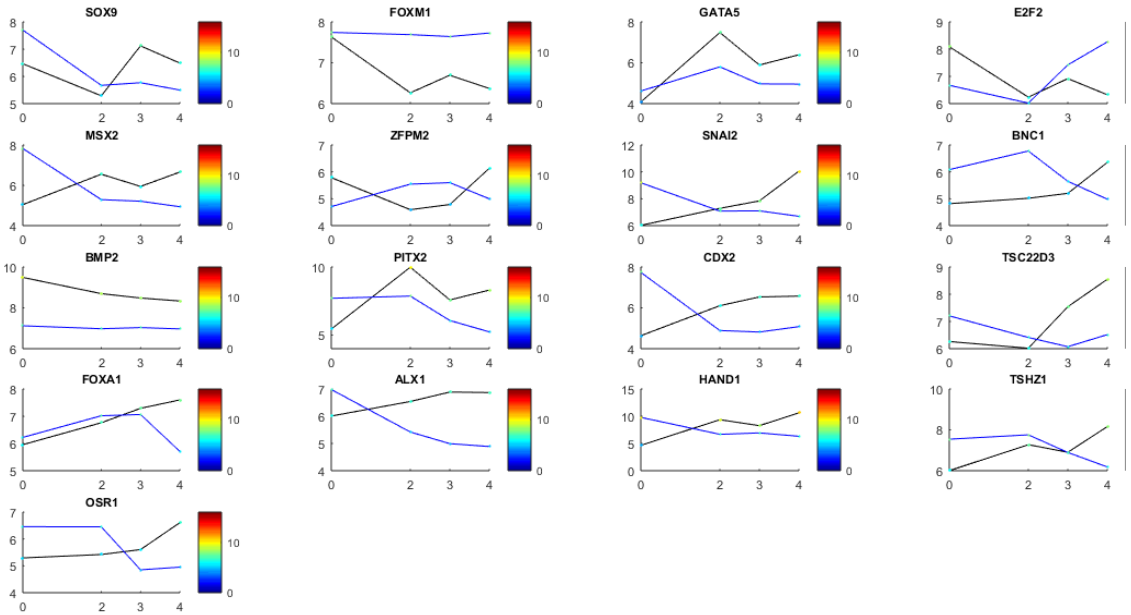
Figure 7-20: Human and mouse gene profiles can be evaluated on the same developmental scale. TGFBI follows opposite dynamics in development and differentiation



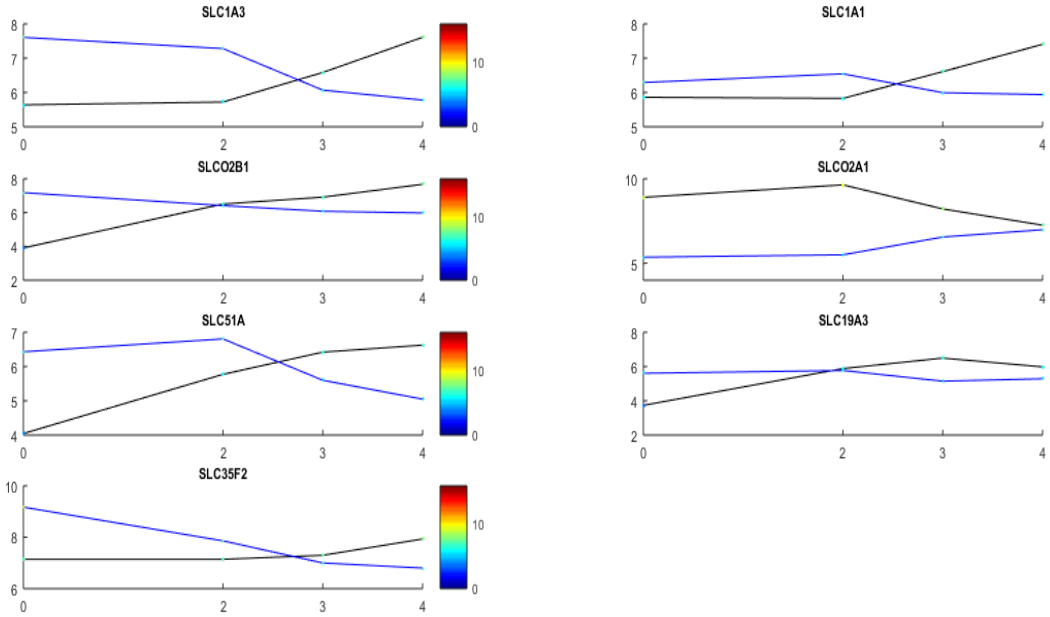
Pathway	Genes
hsa04510:Focal adhesion	COL3A1, IGF1, PDGFC, COL2A1, COL1A1, LAMB1
hsa04512:ECM-receptor interaction	COL3A1, COL2A1, COL1A1, LAMB1
hsa05200:Pathways in cancer	WNT5A, E2F2, BMP2, MMP9, IGF1, BIRC5, LAMB1
hsa00330:Arginine and proline metabolism	ODC1, P4HA2, ALDH4A1

Figure 7-21: Functionally enriched clusters and pathways in the dynamically differentially expressed genes

Transcription Factors



TRANSPORTERS



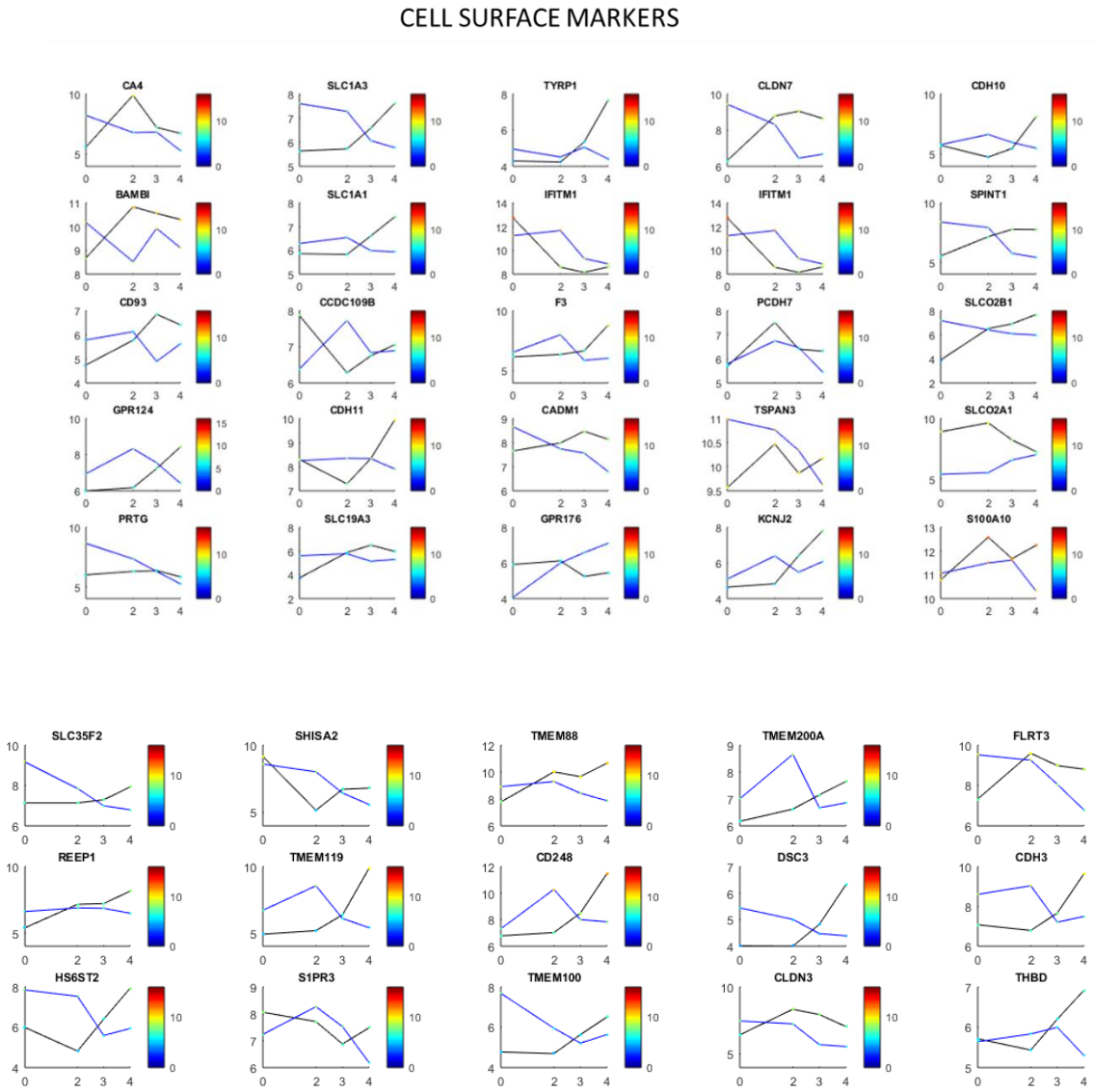
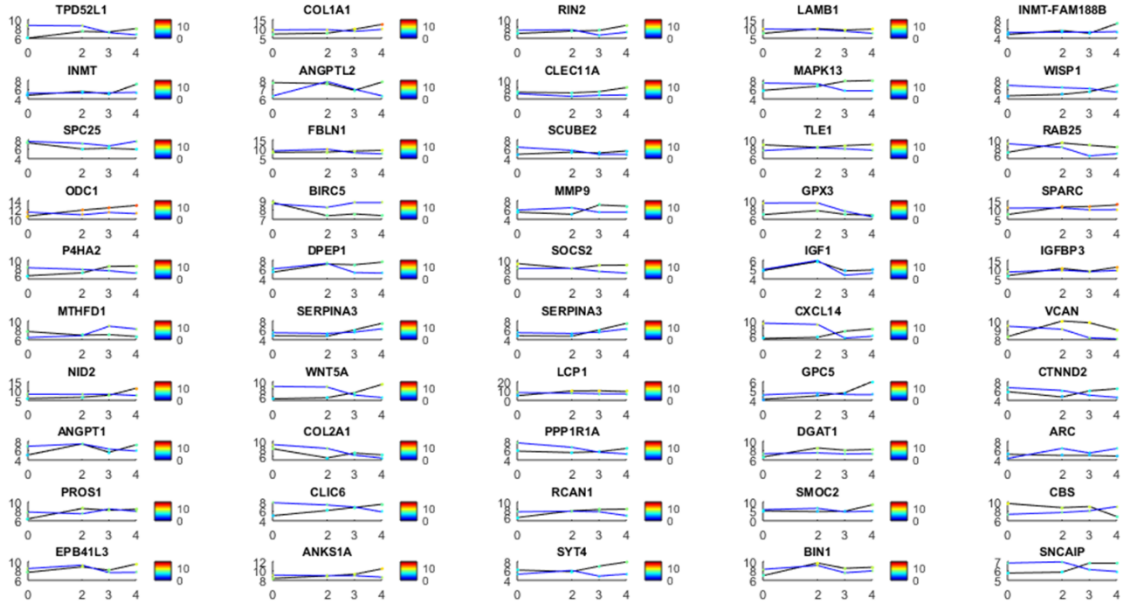
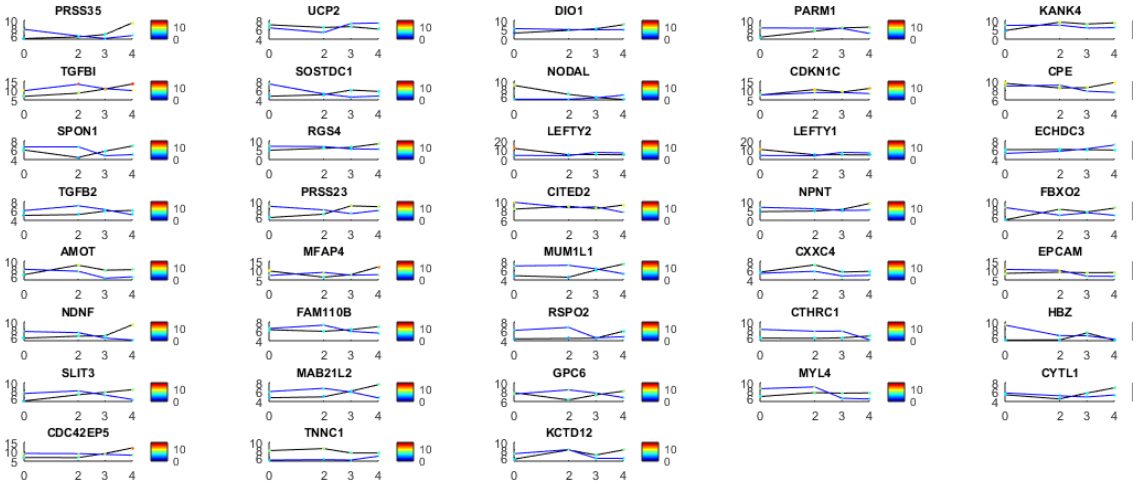


Figure 7-22: Dynamically differentially expressed (A) transcription factors (B) transporters (C) Cell surface markers (D) Other genes between human differentiation (Black : D6_H,D10_H,D14_H,D20_H) and mouse development (Blue :E9.5,E11.5,E13.5,E15.5)

OTHER GENES


TPD52L1
COL1A1
RIN2
LAMB1
INMT-FAM188B
INMT
ANGPTL2
CLEC11A
MAPK13
WISP1
SPC25
FBLN1
SCUBE2
TLE1
RAB25
ODC1
BIRC5
MMP9
GPX3
SPARC
P4HA2
DPEP1
SOCS2
IGF1
IGFBP3
MTHFD1
SERPINA3
SERPINA3
CXCL14
VCAN
NID2
WNT5A
LCP1
GPC5
CTNND2
ANGPT1
COL2A1
PPP1R1A
DGAT1
ARC
PROS1
CLIC6
RCAN1
SMOC2
CBS
EPB41L3
ANKS1A
SYT4
BIN1
SNCAIP


PRSS35
UCP2
DIO1
PARM1
KANK4
TGFBI
SOSTDC1
NODAL
CDKN1C
CPE
SPON1
RGS4
LEFTY2
LEFTY1
ECHDC3
TGFB2
PRSS23
CITED2
NPNT
FBXO2
AMOT
MFAP4
MUM1L1
CXXC4
EPCAM
NDNF
FAM110B
RSPO2
CTHRC1
HBZ
SLIT3
MAB21L2
GPC6
MYL4
CYTL1
CDC42EP5
TNNC1
KCTD12

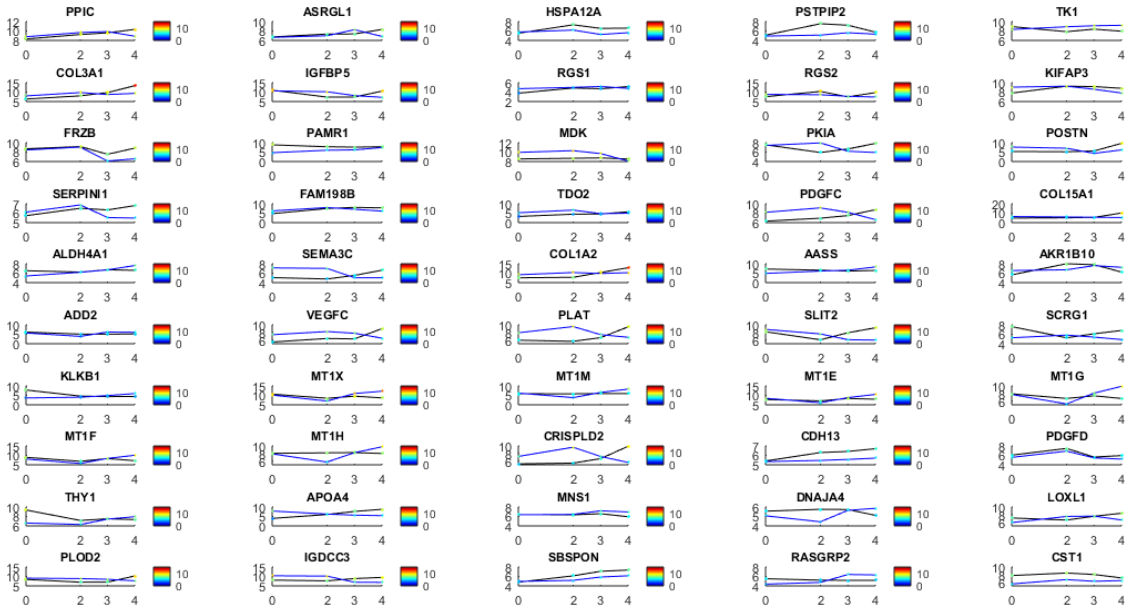


Figure 7-23: Dynamically differentially expressed genes between human differentiation belonging to other functional groups (Black: D6_H, D10_H, D14_H, D20_H) and mouse development (Blue : E9.5, E11.5, E13.5, E15.5)

Table 7-1: Summary of human transcriptome data

Data	Platform	Samples	Brief Description	Abbreviations of important samples
GSE14897	Affymetrix	H9: D0,D20 iPS: Fibroblasts,D20	H9 Differentiation	D20_D
GSE25417	Affymetrix	H9:D5,d10,d15,HNF4 KD D20	HNF4a KD in H9 and differentiated	D5_D, D10_D, D15_D
GSE25744	Illumina	H1: D0, HLC iPS: D0, HLC fibroblasts, fetal liver	H1 and iPS cells differentiated using 2 protocols	D20_A1, D20_A2, D20_A3, D20_A4, D20_A5,D20_A6
Hu et al, to be deposited	Illumina	H9: D0,D6,D9, D10, D12, D14,D20	H9 cells differentiation to HLCs	D6_H, D9_H, D10_H, D12_H, D14_H,D20_H
Hu et al, to be deposited	Illumina	HSF6: D0, D20,D32 spheroids, Adult Liver	HSF6 cells differentiated fro 20 days and cultured as spheroids	D20_H1, D32_H

Hu et al, to be deposited	Affymetrix	HES3: D0, D20	HES3 differentiation to HLCs	D20_H2
GSE42643	Agilent	Fibroblasts, HLCs,	Human fibroblasts reprogrammed to HLC	iHEP
GSE46631	Agilent	Liver bud from iPS	Liver bud: HUVECs,MSCs and iPS derived HLCs	LiverBud

Table 7-2: Summary of mouse transcriptome data

Data	**Platform**	**Brief Description**	**Samples**
GSE46631	Agilent	Mouse liver development	E9.5, E10.5, E11.5, E13.5, E15.5, E17.5, E19.5
GSE13149	Affymetrix	Mouse liver development	E11.5, E12.5, E13.5, E14.5, E15.5, E16.5, E17.5 ,E18.5
www.alexaplatform.org	Illumin HiSeq	Mouse liver development	E8.5, E14.5
GSE23635	Agilent	MEFS reprogrammed to hepatocyes	iHEP1
GSE29725	Agilent	MEFS reprogrammed to hepatocyes	iHEP2
GSE48486	Agilent	MEFS reprogrammed to hepatic progenitor cells	iHEP3

Table 7-3: Transcription factors with overrepresented binding sites in the dynamically differentially expressed genes

Source	Transcription Factor	# Genes with overrepresented TFBS
UCSC_TFBS	GATA	55
UCSC_TFBS	CHX10	51
UCSC_TFBS	FOXO1	50
UCSC_TFBS	NFKAPPAB65	32
UCSC_TFBS	HFH1	52
UCSC_TFBS	MEIS1AHOXA9	43
UCSC_TFBS	RORA2	53
UCSC_TFBS	CREBP1	52
UCSC_TFBS	FREAC4	51
UCSC_TFBS	S8	50
UCSC_TFBS	E47	62
UCSC_TFBS	PBX1	59
UCSC_TFBS	GATA6	24

8. CONCLUSIONS AND FUTURE DIRECTIONS

Stem cells have the ability to propagate in a self-renewing manner, while possessing the potential to differentiate into more specialized cell types in response to molecular cues [14,15]. These special intrinsic characteristics have engendered much interest in using them not only for clinical applications such as regenerative therapy but also for facilitating research by providing better *in vitro* models of diseases for developing novel small molecules of biologics drugs, and drug toxicity testing. The latter applications have gained even more attention with the discovery of iPSCs, as differentiated cells representing a diverse genetic background can now be generated from a large pool of patients, bringing personalized medicine closer to reality. Stem cells for therapeutic applications particularly will entail a robust process that is free from transgene systems or xenogeneic products. Furthermore, the stem cell products will need to meet high quality standards and require thorough characterization to eliminate any undifferentiated or potential tumor forming cells before delivery to patients. Therefore, to harness the full potential of these advances, an enabling technology for consistently and safely producing large quantities of high quality cells is critical.

To address key goals prior to clinical translation of stem cell research, a two-pronged approach was used in this study encompassing both process improvement and a system based analysis for target discovery. The key objectives were a) to increase product cell quantity (Chapter 4) and quality (Chapter 5) from stem cell hepatic differentiations b) to discover the genes that are responsible for a universal gap in maturation of stem cell derived hepatocytes by performing a comparative transcriptome analysis with mouse embryonic liver development (Chapter 6,7)

8.1 Process Improvement

8.1.1 Increasing cell yields through endodermal cell expansion

In Chapter 4, we demonstrated that our hepatic differentiation protocol can be further extended to generate larger quantities of HLCs without utilizing cell sorting or culture on xenogeneic feeder cells. This was implemented by expanding the intermediate

endodermal cell population by at least eight fold by providing a larger culture surface for growth. Stage 2 medium containing FGF2 and BMP4 was chosen to mimic the mesodermal signaling cues occurring in native liver development. These factors play a key role in the transition of the foregut endoderm to the liver bud which is largely dependent on cell migration, proliferation and differentiation [109]. The cell expansion is accompanied with a simultaneous 1000 fold increase in expression of hepatic markers and 1000 fold decrease in endodermal and pluripotent markers in EN2 cells indicates that this phenomenon is selective for cells fated towards the hepatic lineage. The differentiated status of the expanded endodermal cells reduce the contamination of undifferentiated stem cells in the subsequent progeny. Mass cytometry was used to probe the nature of the endodermal to hepatic transition and confirmed that this was occurring in all the cells and not a specific subpopulation. This technique also revealed that the hepatic and endodermal genes are activated and repressed simultaneously for a successful differentiation outcome. Upon continuation of later stages of differentiation, HLCs generated from both protocols exhibited highly similar hepatic transcripts as well as secretory capabilities for albumin and urea. Glycogen synthesis was also seen in both cell types.

Transcriptome analysis further verified hepatic fate acquisition by endodermal expansion through comparison of expression of genes characteristic to pluripotent, endodermal and hepatic fates. The end stage HLCs were also similar in expression of characteristic hepatic genes involved in key functions such as urea cycle, glycolysis/gluconeogenesis and xenobiotic biotransformation. Extensive crosstalk amongst signaling pathways is rampant during embryonic liver development leading cells to embark on different trajectories to attain a similar differentiated status [6]. This is visualized by plotting principal components of samples generated from PCA analysis. While the path to differentiation is different, i.e. the expanded cells are further along in their hepatic phenotype, the HLCs eventually attain the same differentiated state. This suggests that the cells may encounter an upstream barrier to maturation,which will need further evaluation. This approach of endodermal cell expansion using typical

differentiation culture conditions is a simple way to achieve achieving higher cell yields with comparable cell quality with the original protocol.

8.1.2 Enriching hepatic cells from stem cell differentiations to improve product quality

As discussed previously, product purity is critical prior to use of HLCs in clinical applications. The current protocol yields cells with varying hepatic maturity as well as from other lineages. We explored the potential of cell surface marker based sorting and selective medium based enrichment to enrich the desired fraction of cells from Day 14 of differentiation with the hypothesis that these cells may retain the proliferative capabilities of their *in vivo* counterparts.

DLK1 was chosen as a suitable marker and magnetic activated cell sorting was used to enrich hepatic progenitor cells. While the DLK+ population exhibited higher hepatic transcripts, the sorting efficiency was low and cell morphology deteriorated upon resuming culture in Stage 4 conditions. An alternate, gentler method of selection was then pursued using Kubota's medium, a serum free, defined medium which was developed for the ex vivo culture of hepatic progenitor cells isolated from rodent embryonic livers [204]. It was found that cells cultured in Kubota's medium on Collagen 1 were selective for hepatic cells from stem cell differentiations through quantitative PCR and immunohistochemistry. Interestingly, this condition did not support the proliferation of fibroblastic AFP- cells, arising during culture in other combinations of ECM and media. A higher percentage of cells expressing AFP and ALB was seen using flow cytometry. However, these cells did not display the extensive expansive potential of embryonic fetal liver cells. The supplementation of hepatic growth factors commonly used during expansion of *in vivo* fetal liver cells, did not have effects on the phenotype of the *in vitro* cells. Upon differentiation, the hepatic transcripts were comparable to the HLCs generated from the original protocol.

Similar results were also observed in the previous study, where the HLCs from the expanded endodermal cells, which were expressing higher hepatic transcripts, did not show a significant change when compared to the original protocol. This suggests that all differentiating cells encounter a roadblock during the final stages preventing maturation,

which warrants further investigation. Clues from the embryonic liver development, particularly about the progenitor cells and their native microenvironment or 'niche' would be key in deciphering culture conditions for the ex vivo propagation and differentiation of stem cell derived hepatocytes for cell therapies and other clinical applications.

8.1.3 Unveiling the Road Blocks of Stem Cell Differentiation through Cross Species Transcriptome Data Meta-analysis

Comparison of transcriptome data of differentiation to embryonic development can elucidate the genetic roadblocks preventing stem cells from reaching the functional maturity of their tissue counterparts. While transcriptome data for *in vitro* stem cell differentiation is easy to obtain, human *in vivo* developmental transcriptome data is hard to come by. Such *in vivo* data may sometimes be available from other species. Hence, we attempted to combine human *in vitro* data with mouse *in vivo* data with the hypothesis that mouse embryonic development bears a high degree of conservation in gene expression dynamics. The meta-analysis may provide us crucial insights on possible genetic targets that can enhance the maturity in stem cell differentiations.

Briefly, transcriptome data was obtained during the time course of our human hepatic differentiation protocol. This data was augmented with human *in vitro* hepatic differentiation data in the public depository. The data were heterogeneous in their sources, assay methods and platforms. Data were treated with an empirical bayes method based algorithm to remove platform dependent effects and allow for meaningful data comparison. These data of were subjected to hierarchical clustering, principal component analysis (PCA) and non-negative matrix factorization (NMF). Interestingly, in all analyses, data of the same differentiation day from different sources, were grouped together, thereby validating the technique. This same workflow was repeated for mouse *in vitro* stem cell hepatic differentiations. The majority of HLC are similar irrespective of the cell source and protocol. The entire cohort of HLCs clustered separately from the primary hepatocytes and adult liver in both human and mouse differentiations, indicating an inherent roadblock to maturation.

We compiled transcriptome data from various depositories for mouse fetal liver development (from E8.5 to post-natal). We next integrated the transcriptome data of both human and mouse stem cell hepatic differentiations with mouse liver embryonic development data for a comparative analysis. A batch adjustment was performed to integrate the mouse *in vivo* and the different *in vitro* differentiation datasets and validated through the previously described methods. A differential expression analysis between all the HLCs and mature cells revealed that all the major cytochrome P450 enzymes and important metabolic genes were not upregulated in HLCs of both species.

The trajectory of the progression of the developing embryo and differentiating cells were defined as unified "developmental time" with respect to a reference point (E9.5). The unified time scale allowed us to identify that the corresponding embryo development stage at which the *in vitro* stem cell differentiation is blocked is around E15.5. It further allowed the transcriptome dynamics at the critical time points of development to be analyzed systematically. The analysis uncovered a pivotal gene set of about 200 genes with contrasting profiles in hESC differentiation and mouse embryonic liver development and their regulators, that merit combinatorial genetic intervention to enhance maturation of stem cell derived hepatocytes.

8.1.4 Future Outlook

The availability of an expandable source of human hepatocytes has profound implications in the treatment of liver diseases. While these cells may not be suitable for clinical transplantation immediately, this may not limit their use in bio artificial liver devices or as *in vitro* models for drug toxicity and diseases. All these applications will require the reproducible generation of large quantities of high quality functional cells.

With the advent of induced pluripotent stem cells, personalized or population-specific drug screening platforms can be now be established to screen for potential candidates during early stages of drug discovery. This will require the stem cell derived hepatocytes to express the drug metabolizing and detoxification enzymes at levels on par with primary hepatocytes and respond appropriately upon exposure to the respective inducers and inhibitors.

For clinical transplantation where patient safety is paramount, high quality standards will need to be established in addition to requirements of functional maturity. This will entail thorough characterization to confirm cell purity through a combination of cell sorting based on several markers and single cell based techniques. Additionally, the cells will need to undergo routine testing to ensure that they have not acquired any genetic instability through the culture.

The lack of functional maturity in HLCs when compared to their primary counterparts represents a bottleneck in most stem cell applications. The comprehensive transcriptome analysis presented in Chapter 7, shows that this phenomenon is a universal to the state of the art and provides a basis for genetic intervention to overcome this roadblock. The pivotal genes and their regulators could be manipulated genetically using strategies for knock down or overexpression to enhance functional maturity.

Additionally, it was revealed that stem cell differentiation is blocked at a stage corresponding to embryonic day E13-15 in mouse development. Typically, stem cell derived HLCs are transplanted in adult livers of animal models with liver failure and face limited engraftment and restoration of liver function. An alternative strategy will involve transplantation of HLCs in embryonic livers around E13.5 even before the development of the immune system. These HLCs will most likely have a higher chance of engraftment and may even proliferate in the native microenviroment of the developing embryo. Furthermore, these HLCs if tagged with a fluorescent marker could be isolated from mouse livers after birth. A comparative transcriptome and proteomic analysis of both cell states would then reveal critical developmental cues, which can then be recapitulated in *in vitro* systems.

Lastly, stem cell based clinical applications require the development of robust cell production processes. Foremost to developing a robust process is the optimization of culture medium suitable not only for laboratory practice but also for manufacturing and clinical applications. The development of stem cell process technology has the benefit of learning the tremendous expertise accumulated in the past few decades on the manufacturing of therapeutic proteins and viral vaccines. For most applications, the scale up will rely on closed bioreactor systems with stem cells grown on suspended cell

supports such as microcarriers. Untested in those applications is the effect of mechanical and fluid dynamic stresses on selfrenewal and differentiation. For safety reasons and enhanced process robustness, medium used for stem cell culture will increasingly become more chemically defined and animal component free. Innovations in those areas of stem cell bioprocessing will greatly facilitate the translation of discovery to clinical practice.

9. REFERENCES

1. Zorn AM. (2008). Liver Development. In: *StemBook* Stem Cell Research Community.
2. Raju R SSaHW. (2015). Stem Cell Culture Processes. In: *Stem Cells in Regenerative medicine*. Vertes A ed.
3. Jaenisch R and R Young. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132:567-82.
4. Zaret KS. (2002). Regulatory phases of early liver development: paradigms of organogenesis. Nat Rev Genet 3:499-512.
5. Sharma S, R Raju, S Sui and WS Hu. (2011). Stem cell culture engineering - process scale up and beyond. Biotechnol J 6:1317-29.
6. Raju R, D Chau, CM Verfaillie and WS Hu. (2013). The road to regenerative liver therapies: the triumphs, trials and tribulations. Biotechnol Adv 31:1085-93.
7. Lemaigre F and KS Zaret. (2004). Liver development update: new embryo models, cell lineage control, and morphogenesis. Curr Opin Genet Dev 14:582-90.
8. Carpentier B, A Gautier and C Legallais. (2009). Artificial and bioartificial liver devices: present and future. Gut 58:1690-702.
9. Nyberg SL. (2012). Bridging the gap: Advances in artificial liver support. Liver Transpl.
10. Soltys KA, A Soto-Gutierrez, M Nagaya, KM Baskin, M Deutsch, R Ito, BL Shneider, R Squires, J Vockley, C Guha, J Roy-Chowdhury, SC Strom, JL Platt and IJ Fox. (2010). Barriers to the successful treatment of liver disease by hepatocyte transplantation. J Hepatol 53:769-74.
11. Hughes RD, RR Mitry and A Dhawan. (2012). Current status of hepatocyte transplantation. Transplantation 93:342-7.
12. Fukumitsu K, H Yagi and A Soto-Gutierrez. (2011). Bioengineering in organ transplantation: targeting the liver. Transplant Proc 43:2137-8.
13. Raju R, Hu, W.S., Verfaillie, C.M. (2013). Cell-based liver support systems: status and prospect. Current Opinion in Chemical Engineering 2:26-31.
14. Thomson JA, J Itskovitz-Eldor, SS Shapiro, MA Waknitz, JJ Swiergiel, VS Marshall and JM Jones. (1998). Embryonic stem cell lines derived from human blastocysts. Science 282:1145-7.
15. Martin GR. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78:7634-8.
16. Meirelles Lda S and NB Nardi. (2003). Murine marrow-derived mesenchymal stem cell: isolation, in vitro expansion, and characterization. Br J Haematol 123:702-11.

17. Barker N, JH van Es, J Kuipers, P Kujala, M van den Born, M Cozijnsen, A Haegebarth, J Korving, H Begthel, PJ Peters and H Clevers. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449:1003-7.
18. Itzhaki-Alfia A, J Leor, E Raanani, L Sternik, D Spiegelstein, S Netser, R Holbova, M Pevsner-Fischer, J Lavee and IM Barbash. (2009). Patient characteristics and cell source determine the number of isolated human cardiac progenitor cells. Circulation 120:2559-66.
19. Al-Awqati Q and JA Oliver. (2002). Stem cells in the kidney. Kidney Int 61:387-95.
20. Quesenberry P and L Levitt. (1979). Hematopoietic stem cells. N Engl J Med 301:755-61.
21. Ogawa M. (1993). Differentiation and proliferation of hematopoietic stem cells. Blood 81:2844-53.
22. Goldman SA. (2007). Disease targets and strategies for the therapeutic modulation of endogenous neural stem and progenitor cells. Clin Pharmacol Ther 82:453-60.
23. Ohtsuka S and S Dalton. (2008). Molecular and biological properties of pluripotent embryonic stem cells. Gene Ther 15:74-81.
24. Zhang YW, J Denham and RS Thies. (2006). Oligodendrocyte progenitor cells derived from human embryonic stem cells express neurotrophic factors. Stem Cells Dev 15:943-52.
25. Nistor GI, MO Totoiu, N Haque, MK Carpenter and HS Keirstead. (2005). Human embryonic stem cells differentiate into oligodendrocytes in high purity and myelinate after spinal cord transplantation. Glia 49:385-96.
26. Lu B, C Malcuit, S Wang, S Girman, P Francis, L Lemieux, R Lanza and R Lund. (2009). Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. Stem Cells 27:2126-35.
27. Schwartz SD, JP Hubschman, G Heilwell, V Franco-Cardenas, CK Pan, RM Ostrick, E Mickunas, R Gay, I Klimanskaya and R Lanza. (2012). Embryonic stem cell trials for macular degeneration: a preliminary report. Lancet 379:713-20.
28. da Cruz L, FK Chen, A Ahmado, J Greenwood and P Coffey. (2007). RPE transplantation and its role in retinal disease. Prog Retin Eye Res 26:598-635.
29. Idelson M, R Alper, A Obolensky, E Ben-Shushan, I Hemo, N Yachimovich-Cohen, H Khaner, Y Smith, O Wiser, M Gropp, MA Cohen, S Even-Ram, Y Berman-Zaken, L Matzrafi, G Rechavi, E Banin and B Reubinoff. (2009). Directed differentiation of human embryonic stem cells into functional retinal pigment epithelium cells. Cell Stem Cell 5:396-408.
30. Evans MJ and MH Kaufman. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292:154-6.
31. Petersen BE, WC Bowen, KD Patrene, WM Mars, AK Sullivan, N Murase, SS Boggs, JS Greenberger and JP Goff. (1999). Bone marrow as a potential source of hepatic oval cells. Science 284:1168-70.

32. Pittenger MF, AM Mackay, SC Beck, RK Jaiswal, R Douglas, JD Mosca, MA Moorman, DW Simonetti, S Craig and DR Marshak. (1999). Multilineage potential of adult human mesenchymal stem cells. Science 284:143-7.
33. Kopen GC, DJ Prockop and DG Phinney. (1999). Marrow stromal cells migrate throughout forebrain and cerebellum, and they differentiate into astrocytes after injection into neonatal mouse brains. Proc Natl Acad Sci U S A 96:10711-6.
34. Uccelli A, L Moretta and V Pistoia. (2008). Mesenchymal stem cells in health and disease. Nat Rev Immunol.
35. Lasala GP, JA Silva, PA Gardner and JJ Minguell. Combination stem cell therapy for the treatment of severe limb ischemia: safety and efficacy analysis. Angiology 61:551-6.
36. Miyahara Y, N Nagaya, M Kataoka, B Yanagawa, K Tanaka, H Hao, K Ishino, H Ishida, T Shimizu, K Kangawa, S Sano, T Okano, S Kitamura and H Mori. (2006). Monolayered mesenchymal stem cells repair scarred myocardium after myocardial infarction. Nat Med 12:459-65.
37. Boyle AJ, IK McNiece and JM Hare. Mesenchymal stem cell therapy for cardiac repair. Methods Mol Biol 660:65-84.
38. Mannon PJ. (2011). Remestemcel-L: human mesenchymal stem cells as an emerging therapy for Crohn's disease. Expert Opin Biol Ther 11:1249-56.
39. Macmillan ML, BR Blazar, TE DeFor and JE Wagner. (2009). Transplantation of ex-vivo culture-expanded parental haploidentical mesenchymal stem cells to promote engraftment in pediatric recipients of unrelated donor umbilical cord blood: results of a phase I-II clinical trial. Bone Marrow Transplant 43:447-54.
40. Kebriaei P, L Isola, E Bahceci, K Holland, S Rowley, J McGuirk, M Devetten, J Jansen, R Herzig, M Schuster, R Monroy and J Uberti. (2009). Adult human mesenchymal stem cells added to corticosteroid therapy for the treatment of acute graft-versus-host disease. Biol Blood Marrow Transplant 15:804-11.
41. Ji Bao JF, Scott L. Nyberg. (2011). Liver Regeneration and Tissue Engineering. In: *Tissue Engineering in Regenerative Medicine*. Bernstein HS ed.
42. Demetriou AA, RS Brown, Jr., RW Busuttil, J Fair, BM McGuire, P Rosenthal, JS Am Esch, 2nd, J Lerut, SL Nyberg, M Salizzoni, EA Fagan, B de Hemptinne, CE Broelsch, M Muraca, JM Salmeron, JM Rabkin, HJ Metselaar, D Pratt, M De La Mata, LP McChesney, GT Everson, PT Lavin, AC Stevens, Z Pitkin and BA Solomon. (2004). Prospective, randomized, multicenter, controlled trial of a bioartificial liver in treating acute liver failure. Ann Surg 239:660-7; discussion 667-70.
43. Millis JM, DC Cronin, R Johnson, H Conjeevaram, C Conlin, S Trevino and P Maguire. (2002). Initial experience with the modified extracorporeal liver-assist device for patients with fulminant hepatic failure: system modifications and clinical impact. Transplantation 74:1735-46.
44. Carpentier B and SR Ash. (2007). Sorbent-based artificial liver devices: principles of operation, chemical effects and clinical results. Expert Rev Med Devices 4:839-61.

45. Azuma H, N Paulk, A Ranade, C Dorrell, M Al-Dhalimy, E Ellis, S Strom, MA Kay, M Finegold and M Grompe. (2007). Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice. Nat Biotechnol 25:903-10.
46. Hickey RD, JB Lillegard, JE Fisher, TJ McKenzie, SE Hofherr, MJ Finegold, SL Nyberg and M Grompe. (2011). Efficient production of Fah-null heterozygote pigs by chimeric adeno-associated virus-mediated gene knockout and somatic cell nuclear transfer. Hepatology 54:1351-9.
47. Sielaff TD, MY Hu, MD Rollins, JR Bloomer, B Amiot, WS Hu and FB Cerra. (1995). An anesthetized model of lethal canine galactosamine fulminant hepatic failure. Hepatology 21:796-804.
48. Wu FJ, MV Peshwa, FB Cerra and WS Hu. (1995). Entrapment of hepatocyte spheroids in a hollow fiber bioreactor as a potential bioartificial liver. Tissue Eng 1:29-40.
49. Abu-Absi SF, JR Friend, LK Hansen and WS Hu. (2002). Structural polarity and functional bile canaliculi in rat hepatocyte spheroids. Exp Cell Res 274:56-67.
50. Roelandt P, S Obeid, J Paeshuyse, J Vanhove, A Van Lommel, Y Nahmias, F Nevens, J Neyts and CM Verfaillie. (2012). Human pluripotent stem cell-derived hepatocytes support complete replication of hepatitis C virus. J Hepatol 57:246-51.
51. Schwartz RE, K Trehan, L Andrus, TP Sheahan, A Ploss, SA Duncan, CM Rice and SN Bhatia. (2012). Modeling hepatitis C virus infection using human induced pluripotent stem cells. Proc Natl Acad Sci U S A 109:2544-8.
52. Wu X, JM Robotham, E Lee, S Dalton, NM Kneteman, DM Gilbert and H Tang. (2012). Productive hepatitis C virus infection of stem cell-derived hepatocytes reveals a critical transition to viral permissiveness during differentiation. PLoS Pathog 8:e1002617.
53. Yoshida T, K Takayama, M Kondoh, F Sakurai, H Tani, N Sakamoto, Y Matsuura, H Mizuguchi and K Yagi. (2011). Use of human hepatocyte-like cells derived from induced pluripotent stem cells as a model for hepatocytes in hepatitis C virus infection. Biochem Biophys Res Commun 416:119-24.
54. Cayo MA, J Cai, A Delaforest, FK Noto, M Nagaoka, BS Clark, RF Collery, K Si-Tayeb and SA Duncan. (2012). JD induced pluripotent stem cell-derived hepatocytes faithfully recapitulate the pathophysiology of familial hypercholesterolemia. Hepatology.
55. Johnson MH, JC Chisholm, TP Fleming and E Houliston. (1986). A role for cytoplasmic determinants in the development of the mouse early embryo? J Embryol Exp Morphol 97 Suppl:97-121.
56. Yamanaka Y, F Lanner and J Rossant. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development 137:715-24.
57. Gardner RL and J Rossant. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. J Embryol Exp Morphol 52:141-52.

58. De Miguel MP, F Arnalich Montiel, P Lopez Iglesias, A Blazquez Martinez and M Nistal. (2009). Epiblast-derived stem cells in embryonic and adult tissues. Int J Dev Biol 53:1529-40.
59. Liu P, M Wakamiya, MJ Shea, U Albrecht, RR Behringer and A Bradley. (1999). Requirement for Wnt3 in vertebrate axis formation. Nat Genet 22:361-5.
60. Conlon FL, KM Lyons, N Takaesu, KS Barth, A Kispert, B Herrmann and EJ Robertson. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development 120:1919-28.
61. Feldman B, ST Dougan, AF Schier and WS Talbot. (2000). Nodal-related signals establish mesendodermal fate and trunk neural identity in zebrafish. Curr Biol 10:531-4.
62. Gritsman K, J Zhang, S Cheng, E Heckscher, WS Talbot and AF Schier. (1999). The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell 97:121-32.
63. Kim SW, SJ Yoon, E Chuong, C Oyolu, AE Wills, R Gupta and J Baker. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Dev Biol 357:492-504.
64. David NB and FM Rosa. (2001). Cell autonomous commitment to an endodermal fate and behavior by activation of Nodal signalling. Development 128:3937-3947.
65. Schier AF. (2003). Nodal signaling in vertebrate development. Annu Rev Cell Dev Biol 19:589-621.
66. Ciruna B and J Rossant. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. Dev Cell 1:37-49.
67. Rossant J, B Ciruna and J Partanen. (1997). FGF signaling in mouse gastrulation and anteroposterior patterning. Cold Spring Harb Symp Quant Biol 62:127-33.
68. Jung J. (1999). Initiation of Mammalian Liver Development from Endoderm by Fibroblast Growth Factors. Science 284:1998-2003.
69. Douarin N. (1975). An experimental analysis of liver development. Medical biology 53:427.
70. Fukuda-Taira S. (1981). Hepatic induction of in the avian embryo: Specificity of reactive endoderm and inductive mesoderm. Journal of embryology 63.
71. Rossi JM, NR Dunn, BL Hogan and KS Zaret. (2001). Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. Genes Dev 15:1998-2009.
72. Zhang W, TA Yatskievych, RK Baker and PB Antin. (2004). Regulation of Hex gene expression and initial stages of avian hepatogenesis by Bmp and Fgf signaling. Dev Biol 268:312-26.
73. Matsumoto K, H Yoshitomi, J Rossant and KS Zaret. (2001). Liver organogenesis promoted by endothelial cells prior to vascular function. Science 294:559-63.
74. Ishikawa KS, T Masui, K Ishikawa and N Shiojiri. (2001). Immunolocalization of hepatocyte growth factor and its receptor (c-Met) during mouse liver development. Histochem Cell Biol 116:453-62.

75. Hussain SZ, T Sneddon, X Tan, A Micsenyi, GK Michalopoulos and SP Monga. (2004). Wnt impacts growth and differentiation in ex vivo liver development. Exp Cell Res 292:157-69.
76. Kamiya AK, T. Ito, Y. Matsui, T. Morikawa, Y. Senba, E. Nakashima, K. Taga, T. Yoshida, Kanji. Kishimoto, T. Miyajima, A. (1999). Fetal liver development requires a paracrine action of oncostatin M through the gp130 signal transducer. The EMBO Journal 18.
77. Zorn AM. (2008). Liver Development. In: *StemBook* Stem Cell Research Community.
78. Li T, J Huang, Y Jiang, Y Zeng, F He, MQ Zhang, Z Han and X Zhang. (2009). Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. Genomics 93:235-42.
79. Reynolds BA and S Weiss. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science 255:1707-10.
80. Stappenbeck TS and H Miyoshi. (2009). The role of stromal stem cells in tissue regeneration and wound repair. Science 324:1666-9.
81. Takahashi K and S Yamanaka. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663-76.
82. Takahashi K, K Tanabe, M Ohnuki, M Narita, T Ichisaka, K Tomoda and S Yamanaka. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861-72.
83. Yu J, MA Vodyanik, K Smuga-Otto, J Antosiewicz-Bourget, JL Frane, S Tian, J Nie, GA Jonsdottir, V Ruotti, R Stewart, Slukvin, II and JA Thomson. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917-20.
84. Hanna J, M Wernig, S Markoulaki, CW Sun, A Meissner, JP Cassady, C Beard, T Brambrink, LC Wu, TM Townes and R Jaenisch. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science 318:1920-3.
85. Kung JW and SJ Forbes. (2009). Stem cells and liver repair. Curr Opin Biotechnol 20:568-74.
86. Minguet S, I Cortegano, P Gonzalo, JA Martinez-Marin, B de Andres, C Salas, D Melero, ML Gaspar and MA Marcos. (2003). A population of c-Kit(low)(CD45/TER119)- hepatic cell progenitors of 11-day postcoitus mouse embryo liver reconstitutes cell-depleted liver organoids. J Clin Invest 112:1152-63.
87. Suzuki A, Y Zheng, R Kondo, M Kusakabe, Y Takada, K Fukao, H Nakauchi and H Taniguchi. (2000). Flow-cytometric separation and enrichment of hepatic progenitor cells in the developing mouse liver. Hepatology 32:1230-9.
88. Oertel M, A Menthena, YQ Chen, B Teisner, CH Jensen and DA Shafritz. (2008). Purification of fetal liver stem/progenitor cells containing all the repopulation potential for normal adult rat liver. Gastroenterology 134:823-32.

89. Tanimizu N, H Saito, K Mostov and A Miyajima. (2004). Long-term culture of hepatic progenitors derived from mouse Dlk+ hepatoblasts. J Cell Sci 117:6425-34.
90. Okada K, A Kamiya, K Ito, A Yanagida, H Ito, H Kondou, H Nishina and H Nakauchi. (2012). Prospective isolation and characterization of bipotent progenitor cells in early mouse liver development. Stem Cells Dev 21:1124-33.
91. Miki R, N Tatsumi, K Matsumoto and Y Yokouchi. (2008). New primary culture systems to study the differentiation and proliferation of mouse fetal hepatoblasts. Am J Physiol Gastrointest Liver Physiol 294:G529-39.
92. Tanaka M, M Okabe, K Suzuki, Y Kamiya, Y Tsukahara, S Saito and A Miyajima. (2009). Mouse hepatoblasts at distinct developmental stages are characterized by expression of EpCAM and DLK1: drastic change of EpCAM expression during liver development. Mech Dev 126:665-76.
93. Dan YY, KJ Riehle, C Lazaro, N Teoh, J Haque, JS Campbell and N Fausto. (2006). Isolation of multipotent progenitor cells from human fetal liver capable of differentiating into liver and mesenchymal lineages. Proc Natl Acad Sci U S A 103:9912-7.
94. Schmelzer E, L Zhang, A Bruce, E Wauthier, J Ludlow, HL Yao, N Moss, A Melhem, R McClelland, W Turner, M Kulik, S Sherwood, T Tallheden, N Cheng, ME Furth and LM Reid. (2007). Human hepatic stem cells from fetal and postnatal donors. J Exp Med 204:1973-87.
95. Turner R, O Lozoya, Y Wang, V Cardinale, E Gaudio, G Alpini, G Mendel, E Wauthier, C Barbier, D Alvaro and LM Reid. (2011). Human hepatic stem cell and maturational liver lineage biology. Hepatology 53:1035-45.
96. Overturf K, M al-Dhalimy, CN Ou, M Finegold and M Grompe. (1997). Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes. Am J Pathol 151:1273-80.
97. Tanaka M, T Itoh, N Tanimizu and A Miyajima. (2011). Liver stem/progenitor cells: their characteristics and regulatory mechanisms. J Biochem 149:231-9.
98. Kuwahara R, AV Kofman, CS Landis, ES Swenson, E Barendswaard and ND Theise. (2008). The hepatic stem cell niche: identification by label-retaining cell assay. Hepatology 47:1994-2002.
99. Tanaka M and A Miyajima. (2012). Identification and isolation of adult liver stem/progenitor cells. Methods Mol Biol 826:25-32.
100. Dolle L, J Best, C Empsen, J Mei, E Van Rossen, P Roelandt, S Snykers, M Najimi, F Al Battah, ND Theise, K Streetz, E Sokal, IA Leclercq, C Verfaillie, V Rogiers, A Geerts and LA van Grunsven. (2012). Successful isolation of liver progenitor cells by aldehyde dehydrogenase activity in naive mice. Hepatology 55:540-52.
101. Zhang L, N Theise, M Chua and LM Reid. (2008). The stem cell niche of human livers: symmetry between development and regeneration. Hepatology 48:1598-607.
102. Wang Y, CB Cui, M Yamauchi, P Miguez, M Roach, R Malavarca, MJ Costello, V Cardinale, E Wauthier, C Barbier, DA Gerber, D Alvaro and LM Reid. (2011).

Lineage restriction of human hepatic stem cells to mature fates is made efficient by tissue-specific biomatrix scaffolds. Hepatology 53:293-305.
103. D'Amour KA, AD Agulnick, S Eliazer, OG Kelly, E Kroon and EE Baetge. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol 23:1534-41.
104. Agius E, M Oelgeschlager, O Wessely, C Kemp and EM De Robertis. (2000). Endodermal Nodal-related signals and mesoderm induction in Xenopus. Development 127:1173-83.
105. Duboc V, F Lapraz, A Saudemont, N Bessodes, F Mekpoh, E Haillot, M Quirin and T Lepage. (2010). Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development 137:223-35.
106. Gadue P, TL Huber, PJ Paddison and GM Keller. (2006). Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. Proc Natl Acad Sci U S A 103:16806-11.
107. Lowe LA, S Yamada and MR Kuehn. (2001). Genetic dissection of nodal function in patterning the mouse embryo. Development 128:1831-43.
108. Bakre MM, A Hoi, JC Mong, YY Koh, KY Wong and LW Stanton. (2007). Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation. J Biol Chem 282:31703-12.
109. Zorn AM and JM Wells. (2009). Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol 25:221-51.
110. Robertson EJ, DP Norris, J Brennan and EK Bikoff. (2003). Control of early anterior-posterior patterning in the mouse embryo by TGF-beta signalling. Philos Trans R Soc Lond B Biol Sci 358:1351-7; discussion 1357.
111. Cheng X, L Ying, L Lu, AM Galvao, JA Mills, HC Lin, DN Kotton, SS Shen, MC Nostro, JK Choi, MJ Weiss, DL French and P Gadue. (2012). Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. Cell Stem Cell 10:371-84.
112. Zaret KS and M Grompe. (2008). Generation and regeneration of cells of the liver and pancreas. Science 322:1490-4.
113. Sekhon SS, X Tan, A Micsenyi, WC Bowen and SP Monga. (2004). Fibroblast growth factor enriches the embryonic liver cultures for hepatic progenitors. Am J Pathol 164:2229-40.
114. Kamiya AK, T. Miyajima, A. (2001). Oncostatin M and hepatocyte growth factor induce hepatic maturation via distinct signaling pathways. FEBS J 492.
115. Clotman F, P Jacquemin, N Plumb-Rudewiez, CE Pierreux, P Van der Smissen, HC Dietz, PJ Courtoy, GG Rousseau and FP Lemaigre. (2005). Control of liver cell fate decision by a gradient of TGF beta signaling modulated by Onecut transcription factors. Genes Dev 19:1849-54.
116. Schoneveld OJ, IC Gaemers and WH Lamers. (2004). Mechanisms of glucocorticoid signalling. Biochim Biophys Acta 1680:114-28.
117. Aurich I, LP Mueller, H Aurich, J Luetzkendorf, K Tisljar, MM Dollinger, W Schormann, J Walldorf, JG Hengstler, WE Fleig and B Christ. (2007). Functional

integration of hepatocytes derived from human mesenchymal stem cells into mouse livers. Gut 56:405-15.
118. Stock P, S Bruckner, S Ebensing, M Hempel, MM Dollinger and B Christ. (2010). The generation of hepatocytes from mesenchymal stem cells and engraftment into murine liver. Nat Protoc 5:617-27.
119. Aurich H, M Sgodda, P Kaltwasser, M Vetter, A Weise, T Liehr, M Brulport, JG Hengstler, MM Dollinger, WE Fleig and B Christ. (2009). Hepatocyte differentiation of mesenchymal stem cells from human adipose tissue in vitro promotes hepatic integration in vivo. Gut 58:570-81.
120. Banas A. (2012). Purification of adipose tissue mesenchymal stem cells and differentiation toward hepatic-like cells. Methods Mol Biol 826:61-72.
121. Agarwal S, KL Holton and R Lanza. (2008). Efficient differentiation of functional hepatocytes from human embryonic stem cells. Stem Cells 26:1117-27.
122. Basma H, A Soto-Gutierrez, GR Yannam, L Liu, R Ito, T Yamamoto, E Ellis, SD Carson, S Sato, Y Chen, D Muirhead, N Navarro-Alvarez, RJ Wong, J Roy-Chowdhury, JL Platt, DF Mercer, JD Miller, SC Strom, N Kobayashi and IJ Fox. (2009). Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. Gastroenterology 136:990-9.
123. Si-Tayeb K, FK Noto, M Nagaoka, J Li, MA Battle, C Duris, PE North, S Dalton and SA Duncan. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology 51:297-305.
124. Sekiya S and A Suzuki. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature 475:390-3.
125. Huang P, Z He, S Ji, H Sun, D Xiang, C Liu, Y Hu, X Wang and L Hui. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature 475:386-9.
126. Roelandt P, S Obeid, J Paeshuyse, J Vanhove, A Van Lommel, Y Nahmias, F Nevens, J Neyts and CM Verfaillie. (2012). Human pluripotent stem cell-derived hepatocytes support complete replication of hepatitis C virus. J Hepatol 57:246-251.
127. Zhao D, S Chen, J Cai, Y Guo, Z Song, J Che, C Liu, C Wu, M Ding and H Deng. (2009). Derivation and characterization of hepatic progenitor cells from human embryonic stem cells. PLoS One 4:e6468.
128. Asgari S, M Moslem, K Bagheri-Lankarani, B Pournasr, M Miryounesi and H Baharvand. (2011). Differentiation and Transplantation of Human Induced Pluripotent Stem Cell-derived Hepatocyte-like Cells. Stem Cell Rev.
129. Iacob R, U Rudrich, M Rothe, S Kirsch, B Maasoumy, N Narain, CM Verfaillie, P Sancho-Bru, M Iken, I Popescu, A Schambach, MP Manns and M Bock. (2011). Induction of a mature hepatocyte phenotype in adult liver derived progenitor cells by ectopic expression of transcription factors. Stem Cell Res 6:251-61.
130. Takayama K, M Inamura, K Kawabata, M Sugawara, K Kikuchi, M Higuchi, Y Nagamoto, H Watanabe, K Tashiro, F Sakurai, T Hayakawa, MK Furue and H

Mizuguchi. (2012). Generation of metabolically functioning hepatocytes from human pluripotent stem cells by FOXA2 and HNF1alpha transduction. J Hepatol.
131. Takayama K, M Inamura, K Kawabata, K Katayama, M Higuchi, K Tashiro, A Nonaka, F Sakurai, T Hayakawa, MK Furue and H Mizuguchi. (2012). Efficient generation of functional hepatocytes from human embryonic stem cells and induced pluripotent stem cells by HNF4alpha transduction. Mol Ther 20:127-37.
132. van Noort D, SM Ong, C Zhang, S Zhang, T Arooz and H Yu. (2009). Stem cells in microfluidics. Biotechnol Prog 25:52-60.
133. Toh YC, C Zhang, J Zhang, YM Khong, S Chang, VD Samper, D van Noort, DW Hutmacher and H Yu. (2007). A novel 3D mammalian cell perfusion-culture system in microfluidic channels. Lab Chip 7:302-9.
134. Munos B. (2009). Lessons from 60 years of pharmaceutical innovation. Nat Rev Drug Discov 8:959-68.
135. Seth G. HRW, Stapp T.R., Zheng, L., Meier A., Petty K., Leung S. and Chary S. (2013). Development of a new bioprocess scheme using frozen seed train intermediates to initiate CHO cell culture manufacturing campaigns. Biotechnology and Bioengineering 110:1376–1385.
136. Ausubel LJ, PM Lopez and LA Couture. (2011). GMP scale-up and banking of pluripotent stem cells for cellular therapy applications. Methods in Molecular Biology 767:147-59.
137. Hu W-S. *Cell Culture Bioprocess Engineering*. (2012).
138. Wurm FM. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393-8.
139. Hu WS. *Cell Culture Bioprocess Engineering*. (2012). Wei-Shou Hu.
140. Tzanakakis ES, DJ Hess, TD Sielaff and WS Hu. (2000). Extracorporeal tissue engineered liver-assist devices. Annu Rev Biomed Eng 2:607-32.
141. Emamaullee JA and AM Shapiro. (2007). Factors influencing the loss of beta-cell mass in islet transplantation. Cell Transplant 16:1-8.
142. Zweigerdt R. (2009). Large scale production of stem cells and their derivatives. Adv Biochem Eng Biotechnol 114:201-35.
143. Choi JH, KC Keum and SY Lee. (2006). Production of recombinant proteins by high cell density culture of< i> Escherichia coli</i>. Chemical engineering science 61:876-885.
144. Seth G, S Charaniya, KF Wlaschin and WS Hu. (2007). In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. Curr Opin Biotechnol 18:557-64.
145. Giordano A, U Galderisi and IR Marino. (2007). From the laboratory bench to the patient's bedside: an update on clinical trials with mesenchymal stem cells. J Cell Physiol 211:27-35.
146. Kiskinis E and K Eggan. (2010). Progress toward the clinical application of patient-specific pluripotent stem cells. J Clin Invest 120:51-9.
147. Discher DE, DJ Mooney and PW Zandstra. (2009). Growth factors, matrices, and forces combine and control stem cells. Science 324:1673-7.

148. Walker MR, KK Patel and TS Stappenbeck. (2009). The stem cell niche. J Pathol 217:169-80.
149. Pei Wen PS, and Rongwen Xi. (2011). Stem Cell Niche. In: *Regenerative Medicine: From Protocol to Patient*. G.Steinhoff ed. Springer.
150. Burt RK, Y Loh, W Pearce, N Beohar, WG Barr, R Craig, Y Wen, JA Rapp and J Kessler. (2008). Clinical applications of blood-derived and marrow-derived stem cells for nonmalignant diseases. JAMA 299:925-36.
151. Domen J and IL Weissman. (1999). Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate. Mol Med Today 5:201-8.
152. Sorrentino BP. (2004). Clinical strategies for expansion of haematopoietic stem cells. Nat Rev Immunol 4:878-88.
153. Underhill GH and SN Bhatia. (2007). High-throughput analysis of signals regulating stem cell fate and function. Curr Opin Chem Biol 11:357-66.
154. Okita K and S Yamanaka. (2006). Intracellular signaling pathways regulating pluripotency of embryonic stem cells. Curr Stem Cell Res Ther 1:103-11.
155. Xu RH, RM Peck, DS Li, X Feng, T Ludwig and JA Thomson. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods 2:185-90.
156. Chen G, DR Gulbranson, Z Hou, JM Bolin, V Ruotti, MD Probasco, K Smuga-Otto, SE Howden, NR Diol, NE Propson, R Wagner, GO Lee, J Antosiewicz-Bourget, JM Teng and JA Thomson. (2011). Chemically defined conditions for human iPSC derivation and culture. Nat Methods 8:424-9.
157. Levenstein ME, TE Ludwig, RH Xu, RA Llanas, K VanDenHeuvel-Kramer, D Manning and JA Thomson. (2006). Basic fibroblast growth factor support of human embryonic stem cell self-renewal. Stem Cells 24:568-74.
158. Ludwig TE, ME Levenstein, JM Jones, WT Berggren, ER Mitchen, JL Frane, LJ Crandall, CA Daigh, KR Conard, MS Piekarczyk, RA Llanas and JA Thomson. (2006). Derivation of human embryonic stem cells in defined conditions. Nat Biotechnol 24:185-7.
159. Grillberger L, TR Kreil, S Nasr and M Reiter. (2009). Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells. Biotechnol J 4:186-201.
160. Crook JM, TT Peura, L Kravets, AG Bosman, JJ Buzzard, R Horne, H Hentze, NR Dunn, R Zweigerdt, F Chua, A Upshall and A Colman. (2007). The generation of six clinical-grade human embryonic stem cell lines. Cell Stem Cell 1:490-4.
161. Nagaoka M, K Si-Tayeb, T Akaike and SA Duncan. (2010). Culture of human pluripotent stem cells using completely defined conditions on a recombinant E-cadherin substratum. BMC Dev Biol 10:60.
162. Meng G, S Liu, X Li, R Krawetz and DE Rancourt. Extracellular matrix isolated from foreskin fibroblasts supports long-term xeno-free human embryonic stem cell culture. Stem Cells Dev 19:547-56.

163. Burton P, DR Adams, A Abraham, RW Allcock, Z Jiang, A McCahill, J Gilmour, J McAbney, NM Kane, GS Baillie, FR McKenzie, AH Baker, MD Houslay, JC Mountford and G Milligan. Identification and characterization of small-molecule ligands that maintain pluripotency of human embryonic stem cells. Biochem Soc Trans 38:1058-61.
164. Serra M, C Brito, C Correia and PM Alves. (2012). Process engineering of human pluripotent stem cells for clinical application. Trends Biotechnol 30:350-9.
165. Cameron CM, WS Hu and DS Kaufman. (2006). Improved development of human embryonic stem cell-derived embryoid bodies by stirred vessel cultivation. Biotechnol Bioeng 94:938-48.
166. Subramanian K, DJ Owens, TD O'Brien, CM Verfaillie and WS Hu. (2011). Enhanced differentiation of adult bone marrow-derived stem cells to liver lineage in aggregate culture. Tissue Eng Part A 17:2331-41.
167. Cerdan C, SH Hong and M Bhatia. (2007). Formation and hematopoietic differentiation of human embryoid bodies by suspension and hanging drop cultures. Curr Protoc Stem Cell Biol Chapter 1:Unit 1D 2.
168. Niebruegge S, CL Bauwens, R Peerani, N Thavandiran, S Masse, E Sevaptisidis, K Nanthakumar, K Woodhouse, M Husain, E Kumacheva and PW Zandstra. (2009). Generation of human embryonic stem cell-derived mesoderm and cardiac cells using size-specified aggregates in an oxygen-controlled bioreactor. Biotechnol Bioeng 102:493-507.
169. Watanabe K, M Ueno, D Kamiya, A Nishiyama, M Matsumura, T Wataya, JB Takahashi, S Nishikawa, K Muguruma and Y Sasai. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat Biotechnol 25:681-6.
170. Olmer R, A Haase, S Merkert, W Cui, J Palecek, C Ran, A Kirschning, T Scheper, S Glage, K Miller, EC Curnow, ES Hayes and U Martin. (2010). Long term expansion of undifferentiated human iPS and ES cells in suspension culture using a defined medium. Stem Cell Res 5:51-64.
171. Carlsson J, CG Stalnacke, H Acker, M Haji-Karim, S Nilsson and B Larsson. (1979). The influence of oxygen on viability and proliferation in cellular spheroids. Int J Radiat Oncol Biol Phys 5:2011-20.
172. Levine DW, DI Wang and WG Thilly. (2004). Optimization of growth surface parameters in microcarrier cell culture. Biotechnol Bioeng 21:821-845.
173. Ito Y. (1999). Surface micropatterning to regulate cell functions. Biomaterials 20:2333-42.
174. Lock LT and ES Tzanakakis. (2009). Expansion and differentiation of human embryonic stem cells to endoderm progeny in a microcarrier stirred-suspension culture. Tissue Eng Part A 15:2051-63.
175. Lecina M, S Ting, A Choo, S Reuveny and S Oh. (2010). Scalable platform for human embryonic stem cell differentiation to cardiomyocytes in suspended microcarrier cultures. Tissue Eng Part C Methods 16:1609-19.

176. Park Y, K Subramanian, CM Verfaillie and WS Hu. Expansion and hepatic differentiation of rat multipotent adult progenitor cells in microcarrier suspension culture. J Biotechnol 150:131-9.
177. Subramanian K, DJ Owens, R Raju, M Firpo, TD O'Brien, CM Verfaillie and WS Hu. (2014). Spheroid culture for enhanced differentiation of human embryonic stem cells to hepatocyte-like cells. Stem Cells Dev 23:124-31.
178. Du P, WA Kibbe and SM Lin. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics 24:1547-8.
179. Subramanian A, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub, ES Lander and JP Mesirov. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-50.
180. Mootha VK, CM Lindgren, KF Eriksson, A Subramanian, S Sihag, J Lehar, P Puigserver, E Carlsson, M Ridderstrale, E Laurila, N Houstis, MJ Daly, N Patterson, JP Mesirov, TR Golub, P Tamayo, B Spiegelman, ES Lander, JN Hirschhorn, D Altshuler and LC Groop. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267-73.
181. Haridass D, N Narain and M Ott. (2008). Hepatocyte transplantation: waiting for stem cells. Curr Opin Organ Transplant 13:627-32.
182. Roelandt P, KA Pauwelyn, P Sancho-Bru, K Subramanian, B Bose, L Ordovas, K Vanuytsel, M Geraerts, M Firpo, R De Vos, J Fevery, F Nevens, WS Hu and CM Verfaillie. (2010). Human embryonic and rat adult stem cells with primitive endoderm-like phenotype can be fated to definitive endoderm, and finally hepatocyte-like cells. PLoS One 5:e12101.
183. Yu Y, H Liu, Y Ikeda, BP Amiot, P Rinaldo, SA Duncan and SL Nyberg. (2012). Hepatocyte-like cells differentiated from human induced pluripotent stem cells: relevance to cellular therapies. Stem Cell Res 9:196-207.
184. Sharma AD, T Cantz, A Vogel, A Schambach, D Haridass, M Iken, M Bleidissel, MP Manns, HR Scholer and M Ott. (2008). Murine embryonic stem cell-derived hepatic progenitor cells engraft in recipient livers with limited capacity of liver tissue formation. Cell Transplant 17:313-23.
185. Li F, P Liu, C Liu, D Xiang, L Deng, W Li, K Wangensteen, J Song, Y Ma, L Hui, L Wei, L Li, X Ding, Y Hu, Z He and X Wang. (2010). Hepatoblast-like progenitor cells derived from embryonic stem cells can repopulate livers of mice. Gastroenterology 139:2158-2169 e8.
186. Radzisheuskaya A, B Chia Gle, RL dos Santos, TW Theunissen, LF Castro, J Nichols and JC Silva. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol 15:579-90.
187. Yasunaga M, S Tada, S Torikai-Nishikawa, Y Nakano, M Okada, LM Jakt, S Nishikawa, T Chiba, T Era and S Nishikawa. (2005). Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. Nat Biotechnol 23:1542-50.

188. Nava S, M Westgren, M Jaksch, A Tibell, U Broome, BG Ericzon and S Sumitran-Holgersson. (2005). Characterization of cells in the developing human liver. Differentiation 73:249-60.
189. Bendall SC, EF Simonds, P Qiu, AD Amir el, PO Krutzik, R Finck, RV Bruggner, R Melamed, A Trejo, OI Ornatsky, RS Balderas, SK Plevritis, K Sachs, D Pe'er, SD Tanner and GP Nolan. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332:687-96.
190. Orian-Rousseau V, L Chen, JP Sleeman, P Herrlich and H Ponta. (2002). CD44 is required for two consecutive steps in HGF/c-Met signaling. Genes Dev 16:3074-86.
191. Stark R, F Guebre-Egziabher, X Zhao, C Feriod, J Dong, TC Alves, S Ioja, RL Pongratz, S Bhanot, M Roden, GW Cline, GI Shulman and RG Kibbey. (2014). A role for mitochondrial phosphoenolpyruvate carboxykinase (PEPCK-M) in the regulation of hepatic gluconeogenesis. J Biol Chem 289:7257-63.
192. Maurice N and DH Perlmutter. (2012). Novel treatment strategies for liver disease due to alpha1-antitrypsin deficiency. Clin Transl Sci 5:289-94.
193. Bibi Z. (2008). Role of cytochrome P450 in drug interactions. Nutr Metab (Lond) 5:27.
194. Chen J and K Raymond. (2006). Roles of rifampicin in drug-drug interactions: underlying molecular mechanisms involving the nuclear pregnane X receptor. Ann Clin Microbiol Antimicrob 5:3.
195. Yu HB, G Kunarso, FH Hong and LW Stanton. (2009). Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. J Biol Chem 284:31327-35.
196. Seo T, PM Oelkers, MR Giattina, TS Worgall, SL Sturley and RJ Deckelbaum. (2001). Differential modulation of ACAT1 and ACAT2 transcription and activity by long chain free fatty acids in cultured cells. Biochemistry 40:4756-62.
197. Kohjima M, M Enjoji, N Higuchi, M Kato, K Kotoh, T Yoshimoto, T Fujino, M Yada, R Yada, N Harada, R Takayanagi and M Nakamuta. (2007). Re-evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. Int J Mol Med 20:351-8.
198. David NB and FM Rosa. (2001). Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. Development 128:3937-47.
199. Bjornson ZB, GP Nolan and WJ Fantl. (2013). Single-cell mass cytometry for analysis of immune system functional states. Curr Opin Immunol 25:484-94.
200. Bendall SC and GP Nolan. (2012). From single cells to deep phenotypes in cancer. Nat Biotechnol 30:639-47.
201. Huang P, L Zhang, Y Gao, Z He, D Yao, Z Wu, J Cen, X Chen, C Liu, Y Hu, D Lai, Z Hu, L Chen, Y Zhang, X Cheng, X Ma, G Pan, X Wang and L Hui. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell 14:370-84.

202. Jorns C, EC Ellis, G Nowak, B Fischler, A Nemeth, SC Strom and BG Ericzon. (2012). Hepatocyte transplantation for inherited metabolic diseases of the liver. J Intern Med 272:201-23.
203. Oertel M. (2011). Fetal liver cell transplantation as a potential alternative to whole liver transplantation? Journal of Gastroenterology 46:953-65.
204. Kubota H and LM Reid. (2000). Clonogenic hepatoblasts, common precursors for hepatocytic and biliary lineages, are lacking classical major histocompatibility complex class I antigen. Proc Natl Acad Sci U S A 97:12132-7.
205. Turner WS, E Schmelzer, R McClelland, E Wauthier, W Chen and LM Reid. (2007). Human hepatoblast phenotype maintained by hyaluronan hydrogels. J Biomed Mater Res B Appl Biomater 82:156-68.
206. Chen Q, J Kon, H Ooe, K Sasaki and T Mitaka. (2007). Selective proliferation of rat hepatocyte progenitor cells in serum-free culture. Nat Protoc 2:1197-205.
207. Takashimizu I, Y Tanaka, S Yoshie, Y Kano, H Ichikawa, L Cui, N Ogiwara, K Johkura and K Sasaki. (2009). Localization of Liv2 as an immature hepatocyte marker in EB outgrowth. ScientificWorldJournal 9:190-9.
208. Lee NP and JM Luk. (2010). Hepatic tight junctions: from viral entry to cancer metastasis. World J Gastroenterol 16:289-95.
209. Turner R, G Mendel, E Wauthier, C Barbier and LM Reid. (2012). Hyaluronan-Supplemented Buffers Preserve Adhesion Mechanisms Facilitating Cryopreservation of Human Hepatic Stem/Progenitor Cells. Cell Transplant.
210. Brill S, I Zvibel, Z Halpern and R Oren. (2002). The role of fetal and adult hepatocyte extracellular matrix in the regulation of tissue-specific gene expression in fetal and adult hepatocytes. Eur J Cell Biol 81:43-50.
211. Hughes CS, LM Postovit and GA Lajoie. (2010). Matrigel: a complex protein mixture required for optimal growth of cell culture. Proteomics 10:1886-90.
212. Moumen A, S Patane, A Porras, R Dono and F Maina. (2007). Met acts on Mdm2 via mTOR to signal cell survival during development. Development 134:1443-51.
213. Sasaki K, J Kon, T Mizuguchi, Q Chen, H Ooe, H Oshima, K Hirata and T Mitaka. (2008). Proliferation of hepatocyte progenitor cells isolated from adult human livers in serum-free medium. Cell Transplant 17:1221-30.
214. Fong CY, K Gauthaman and A Bongso. (2010). Teratomas from pluripotent stem cells: A clinical hurdle. J Cell Biochem 111:769-81.
215. Choo AB, HL Tan, SN Ang, WJ Fong, A Chin, J Lo, L Zheng, H Hentze, RJ Philp, SK Oh and M Yap. (2008). Selection against undifferentiated human embryonic stem cells by a cytotoxic antibody recognizing podocalyxin-like protein-1. Stem Cells 26:1454-63.
216. Gronthos S, AC Zannettino, SJ Hay, S Shi, SE Graves, A Kortesidis and PJ Simmons. (2003). Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow. J Cell Sci 116:1827-35.
217. Gronthos S, S Fitter, P Diamond, PJ Simmons, S Itescu and AC Zannettino. (2007). A novel monoclonal antibody (STRO-3) identifies an isoform of tissue nonspecific alkaline phosphatase expressed by multipotent bone marrow stromal stem cells. Stem Cells Dev 16:953-63.

218. Klug MG, MH Soonpaa, GY Koh and LJ Field. (1996). Genetically selected cardiomyocytes from differentiating embronic stem cells form stable intracardiac grafts. J Clin Invest 98:216-24.
219. Billon N, C Jolicoeur, QL Ying, A Smith and M Raff. (2002). Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells. J Cell Sci 115:3657-65.
220. Ishizaka S, A Shiroi, S Kanda, M Yoshikawa, H Tsujinoue, S Kuriyama, T Hasuma, K Nakatani and K Takahashi. (2002). Development of hepatocytes from ES cells after transfection with the HNF-3beta gene. FASEB J 16:1444-6.
221. Kouskoff V, G Lacaud, S Schwantz, HJ Fehling and G Keller. (2005). Sequential development of hematopoietic and cardiac mesoderm during embryonic stem cell differentiation. Proc Natl Acad Sci U S A 102:13170-5.
222. Miltenyi S, W Muller, W Weichel and A Radbruch. (1990). High gradient magnetic cell separation with MACS. Cytometry 11:231-8.
223. Grutzkau A and A Radbruch. (2010). Small but mighty: how the MACS-technology based on nanosized superparamagnetic particles has helped to analyze the immune system within the last 20 years. Cytometry A 77:643-7.
224. Keever-Taylor CA, SM Devine, RJ Soiffer, A Mendizabal, S Carter, MC Pasquini, PN Hari, A Stein, HM Lazarus, C Linker, SC Goldstein, EA Stadtmauer and RJ O'Reilly. (2012). Characteristics of CliniMACS(R) System CD34-enriched T cell-depleted grafts in a multicenter trial for acute myeloid leukemia-Blood and Marrow Transplant Clinical Trials Network (BMT CTN) protocol 0303. Biol Blood Marrow Transplant 18:690-7.
225. Adewumi O, B Aflatoonian, L Ahrlund-Richter, M Amit, PW Andrews, G Beighton, PA Bello, N Benvenisty, LS Berry, S Bevan, B Blum, J Brooking, KG Chen, AB Choo, GA Churchill, M Corbel, I Damjanov, JS Draper, P Dvorak, K Emanuelsson, RA Fleck, A Ford, K Gertow, M Gertsenstein, PJ Gokhale, RS Hamilton, A Hampl, LE Healy, O Hovatta, J Hyllner, MP Imreh, J Itskovitz-Eldor, J Jackson, JL Johnson, M Jones, K Kee, BL King, BB Knowles, M Lako, F Lebrin, BS Mallon, D Manning, Y Mayshar, RD McKay, AE Michalska, M Mikkola, M Mileikovsky, SL Minger, HD Moore, CL Mummery, A Nagy, N Nakatsuji, CM O'Brien, SK Oh, C Olsson, T Otonkoski, KY Park, R Passier, H Patel, M Patel, R Pedersen, MF Pera, MS Piekarczyk, RA Pera, BE Reubinoff, AJ Robins, J Rossant, P Rugg-Gunn, TC Schulz, H Semb, ES Sherrer, H Siemen, GN Stacey, M Stojkovic, H Suemori, J Szatkiewicz, T Turetsky, T Tuuri, S van den Brink, K Vintersten, S Vuoristo, D Ward, TA Weaver, LA Young and W Zhang. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol 25:803-16.
226. Farrell J, C Kelly, J Rauch, K Kida, A Garcia-Munoz, N Monsefi, B Turriziani, C Doherty, JP Mehta, D Matallanas, JC Simpson, W Kolch and A von Kriegsheim. (2014). HGF induces epithelial-to-mesenchymal transition by modulating the mammalian hippo/MST2 and ISG15 pathways. J Proteome Res 13:2874-86.
227. Jozefczuk J, A Prigione, L Chavez and J Adjaye. (2011). Comparative analysis of human embryonic stem cell and induced pluripotent stem cell-derived hepatocyte-

like cells reveals current drawbacks and possible strategies for improved differentiation. Stem Cells Dev 20:1259-75.
228. DeLaForest A, M Nagaoka, K Si-Tayeb, FK Noto, G Konopka, MA Battle and SA Duncan. (2011). HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development 138:4143-53.
229. Takebe T, K Sekine, M Enomura, H Koike, M Kimura, T Ogaeri, RR Zhang, Y Ueno, YW Zheng, N Koike, S Aoyama, Y Adachi and H Taniguchi. (2013). Vascularized and functional human liver from an iPSC-derived organ bud transplant. Nature 499:481-4.
230. Wei W, J Hou, O Alder, X Ye, S Lee, R Cullum, A Chu, Y Zhao, SM Warner, DA Knight, D Yang, SJ Jones, MA Marra and PA Hoodless. (2013). Genome-wide microRNA and messenger RNA profiling in rodent liver development implicates mir302b and mir20a in repressing transforming growth factor-beta signaling. Hepatology 57:2491-501.
231. Heider A and R Alt. (2013). virtualArray: a R/bioconductor package to merge raw data from different microarray platforms. BMC Bioinformatics 14:75.
232. Johnson WE, C Li and A Rabinovic. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118-27.
233. Durinck S, PT Spellman, E Birney and W Huber. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4:1184-91.
234. Hollern DP and ER Andrechek. (2014). A genomic analysis of mouse models of breast cancer reveals molecular features of mouse models and relationships to human breast cancer. Breast Cancer Res 16:R59.
235. Alter O, PO Brown and D Botstein. (2000). Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A 97:10101-6.
236. Brunet JP, P Tamayo, TR Golub and JP Mesirov. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 101:4164-9.
237. Gaujoux R and C Seoighe. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11:367.
238. Kim H and H Park. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23:1495-502.
239. Gadgil M, S Mehra, V Kapur and WS Hu. (2006). TimeView: for comparative gene expression analysis. Appl Bioinformatics 5:41-4.
240. Vaquerizas JM, SK Kummerfeld, SA Teichmann and NM Luscombe. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10:252-63.
241. Tamas Hegedus TM. (2013). SLC Tables. Journal Molecular Aspects of Medicine. .

242. da Cunha JP, PA Galante, JE de Souza, RF de Souza, PM Carvalho, DT Ohara, RP Moura, SM Oba-Shinja, SK Marie, WA Silva, Jr., RO Perez, B Stransky, M Pieprzyk, J Moore, O Caballero, J Gama-Rodrigues, A Habr-Gama, WP Kuo, AJ Simpson, AA Camargo, LJ Old and SJ de Souza. (2009). Bioinformatics construction of the human cell surfaceome. Proc Natl Acad Sci U S A 106:16752-7.
243. Bolstad BM, RA Irizarry, M Astrand and TP Speed. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185-93.
244. Hsu SD, YT Tseng, S Shrestha, YL Lin, A Khaleel, CH Chou, CF Chu, HY Huang, CM Lin, SY Ho, TY Jian, FM Lin, TH Chang, SL Weng, KW Liao, IE Liao, CC Liu and HD Huang. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 42:D78-85.
245. Wong N and X Wang. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 43:D146-52.
246. Liu H, Y Kim, S Sharkis, L Marchionni and YY Jang. (2011). In vivo liver regeneration potential of human induced pluripotent stem cells from diverse origins. Sci Transl Med 3:82ra39.
247. Shan J, RE Schwartz, NT Ross, DJ Logan, D Thomas, SA Duncan, TE North, W Goessling, AE Carpenter and SN Bhatia. (2013). Identification of small molecules for human hepatocyte expansion and iPS differentiation. Nat Chem Biol 9:514-20.
248. Takayama K, M Inamura, K Kawabata, M Sugawara, K Kikuchi, M Higuchi, Y Nagamoto, H Watanabe, K Tashiro, F Sakurai, T Hayakawa, MK Furue and H Mizuguchi. (2012). Generation of metabolically functioning hepatocytes from human pluripotent stem cells by FOXA2 and HNF1alpha transduction. J Hepatol 57:628-36.
249. Du Y, J Wang, J Jia, N Song, C Xiang, J Xu, Z Hou, X Su, B Liu, T Jiang, D Zhao, Y Sun, J Shu, Q Guo, M Yin, D Sun, S Lu, Y Shi and H Deng. (2014). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. Cell Stem Cell 14:394-403.
250. Sancho-Bru P, P Roelandt, N Narain, K Pauwelyn, T Notelaers, T Shimizu, M Ott and C Verfaillie. (2011). Directed differentiation of murine-induced pluripotent stem cells to functional hepatocyte-like cells. J Hepatol 54:98-107.
251. Roelandt P, J Vanhove and C Verfaillie. (2013). Directed differentiation of pluripotent stem cells to functional hepatocytes. Methods Mol Biol 997:141-7.
252. Benito M, J Parker, Q Du, J Wu, D Xiang, CM Perou and JS Marron. (2004). Adjustment of systematic microarray data biases. Bioinformatics 20:105-14.
253. Sims AH, GJ Smethurst, Y Hey, MJ Okoniewski, SD Pepper, A Howell, CJ Miller and RB Clarke. (2008). The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets - improving meta-analysis and prediction of prognosis. BMC Med Genomics 1:42.

254. Chen C, K Grennan, J Badner, D Zhang, E Gershon, L Jin and C Liu. (2011). Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. PLoS One 6:e17238.
255. Luo J, M Schumacher, A Scherer, D Sanoudou, D Megherbi, T Davison, T Shi, W Tong, L Shi, H Hong, C Zhao, F Elloumi, W Shi, R Thomas, S Lin, G Tillinghast, G Liu, Y Zhou, D Herman, Y Li, Y Deng, H Fang, P Bushel, M Woods and J Zhang. (2010). A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. Pharmacogenomics J 10:278-91.
256. Kupfer P, R Guthke, D Pohlers, R Huber, D Koczan and RW Kinne. (2012). Batch correction of microarray data substantially improves the identification of genes differentially expressed in rheumatoid arthritis and osteoarthritis. BMC Med Genomics 5:23.
257. Leek JT and JD Storey. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 3:1724-35.
258. Nelms BL and PA Labosky. (2010). In: *Transcriptional Control of Neural Crest Development*. San Rafael (CA).
259. Strumpf D, CA Mao, Y Yamanaka, A Ralston, K Chawengsaksophak, F Beck and J Rossant. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132:2093-102.
260. Niwa H, Y Toyooka, D Shimosato, D Strumpf, K Takahashi, R Yagi and J Rossant. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123:917-29.
261. Gao N, P White and KH Kaestner. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell 16:588-99.
262. McFadden DG, AC Barbosa, JA Richardson, MD Schneider, D Srivastava and EN Olson. (2005). The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. Development 132:189-201.
263. Duboc V and MP Logan. (2011). Pitx1 is necessary for normal initiation of hindlimb outgrowth through regulation of Tbx4 expression and shapes hindlimb morphologies via targeted growth control. Development 138:5301-9.
264. Bensoussan-Trigano V, Y Lallemand, C Saint Cloment and B Robert. (2011). Msx1 and Msx2 in limb mesenchyme modulate digit number and identity. Dev Dyn 240:1190-202.
265. Belo J, M Krishnamurthy, A Oakie and R Wang. (2013). The role of SOX9 transcription factor in pancreatic and duodenal development. Stem Cells Dev 22:2935-43.
266. Casas E, J Kim, A Bendesky, L Ohno-Machado, CJ Wolfe and J Yang. (2011). Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. Cancer Res 71:245-54.
267. Fenouille N, M Tichet, M Dufies, A Pottier, A Mogha, JK Soo, S Rocchi, A Mallavialle, MD Galibert, A Khammari, JP Lacour, R Ballotti, M Deckert and S

Tartare-Deckert. (2012). The epithelial-mesenchymal transition (EMT) regulatory factor SLUG (SNAI2) is a downstream target of SPARC and AKT in promoting melanoma cell invasion. PLoS One 7:e40378.
268. Bernardo GM and RA Keri. (2012). FOXA1: a transcription factor with parallel functions in development and cancer. Biosci Rep 32:113-30.
269. Raum JC, SA Soleimanpour, DN Groff, N Core, L Fasano, AN Garratt, C Dai, AC Powers and DA Stoffers. (2015). Tshz1 Regulates Pancreatic beta-Cell Maturation. Diabetes 64:2905-14.
270. Smith SM and DW Murray. (2012). An overview of microRNA methods: expression profiling and target identification. Methods Mol Biol 823:119-38.
271. Tzur G, A Israel, A Levy, H Benjamin, E Meiri, Y Shufaro, K Meir, E Khvalevsky, Y Spector, N Rojansky, Z Bentwich, BE Reubinoff and E Galun. (2009). Comprehensive gene and microRNA expression profiling reveals a role for microRNAs in human liver development. PLoS One 4:e7511.
272. Haussecker D and MA Kay. (2010). miR-122 continues to blaze the trail for microRNA therapeutics. Mol Ther 18:240-2.
273. Khella HW, M Bakhet, G Allo, MA Jewett, AH Girgis, A Latif, H Girgis, I Von Both, GA Bjarnason and GM Yousef. (2013). miR-192, miR-194 and miR-215: a convergent microRNA network suppressing tumor progression in renal cell carcinoma. Carcinogenesis 34:2231-9.
274. Meng Z FX, Chen X, Zeng S, Tian Y, Jove R, Xu R, Huang W. (2010). miR-194 is a marker of hepatic epithelial cells and suppresses metastasis of liver cancer cells in mice. Hepatology 52:2148-57.
275. Kuo CH, JH Deng, Q Deng and SY Ying. (2012). A novel role of miR-302/367 in reprogramming. Biochem Biophys Res Commun 417:11-6.
276. Xue L, JY Cai, J Ma, Z Huang, MX Guo, LZ Fu, YB Shi and WX Li. (2013). Global expression profiling reveals genetic programs underlying the developmental divergence between mouse and human embryogenesis. BMC Genomics 14:568.
277. Clancy B, RB Darlington and BL Finlay. (2001). Translating developmental time across mammalian species. Neuroscience 105:7-17.
278. Chitwood DH, JN Maloof and NR Sinha. (2013). Dynamic transcriptomic profiles between tomato and a wild relative reflect distinct developmental architectures. Plant Physiol 162:537-52.
279. Levin M, T Hashimshony, F Wagner and I Yanai. (2012). Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev Cell 22:1101-8.
280. Anavy L, M Levin, S Khair, N Nakanishi, SL Fernandez-Valverde, BM Degnan and I Yanai. (2014). BLIND ordering of large-scale transcriptomic developmental timecourses. Development 141:1161-6.
281. Baxter M, S Withey, S Harrison, CP Segeritz, F Zhang, R Atkinson-Dell, C Rowe, DT Gerrard, R Sison-Young, R Jenkins, J Henry, AA Berry, L Mohamet, M Best, SW Fenwick, H Malik, NR Kitteringham, CE Goldring, K Piper Hanley, L Vallier and NA Hanley. (2015). Phenotypic and functional analyses show stem

cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. J Hepatol 62:581-9.

282. Morris SA, P Cahan, H Li, AM Zhao, AK San Roman, RA Shivdasani, JJ Collins and GQ Daley. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell 158:889-902.
283. Cahan P, H Li, SA Morris, E Lummertz da Rocha, GQ Daley and JJ Collins. (2014). CellNet: network biology applied to stem cell engineering. Cell 158:903-15.
284. Godoy P, W Schmidt-Heck, K Natarajan, B Lucendo-Villarin, D Szkolnicka, A Asplund, P Bjorquist, A Widera, R Stoeber, G Campos, S Hammad, A Sachinidis, U Chaudhari, G Damm, TS Weiss, A Nussler, J Synnergren, K Edlund, B Kuppers-Munther, D Hay and JG Hengstler. (2015). Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. J Hepatol.
285. Arterbery AS and CW Bogue. (2014). Endodermal and mesenchymal cross talk: a crossroad for the maturation of foregut organs. Pediatr Res 75:120-6.
286. Bryant PJ and P Simpson. (1984). Intrinsic and extrinsic control of growth in developing organs. Q Rev Biol 59:387-415.

10. APPENDIX A: GENE LISTS FROM META ANALYSIS

Table 10-1: Metagenes from NMF analysis of human differentiation data

Ensembl_ID	SYMBOL	Metagene
ENSG00000165495	PKNOX2	1
ENSG00000157827	FMNL2	1
ENSG00000160326	SLC2A6	1
ENSG00000158270	COLEC12	1
ENSG00000171208	NETO2	1
ENSG00000134323	MYCN	1
ENSG00000090530	LEPREL1	1
ENSG00000147869	CER1	1
ENSG00000169169	CPT1C	1
ENSG00000196376	SLC35F1	1
ENSG00000143768	LEFTY2	1
ENSG00000243709	LEFTY1	1
ENSG00000149201	CCDC81	1
ENSG00000174469	CNTNAP2	1
ENSG00000167157	PRRX2	1
ENSG00000172348	RCAN2	1
ENSG00000101144	BMP7	1
ENSG00000137285	TUBB2B	1
ENSG00000121966	CXCR4	1
ENSG00000240694	PNMA2	1
ENSG00000158246	FAM46B	1
ENSG00000272398	CD24	1
ENSG00000137821	LRRC49	1
ENSG00000019144	PHLDB1	1
ENSG00000178531	CTXN1	1
ENSG00000105270	CLIP3	1
ENSG00000138650	PCDH10	1
ENSG00000109705	NKX3-2	1
ENSG00000128045	RASL11B	1
ENSG00000184261	KCNK12	1
ENSG00000130294	KIF1A	1
ENSG00000131914	LIN28A	1
ENSG00000159167	STC1	1
ENSG00000169992	NLGN2	1
ENSG00000131711	MAP1B	1
ENSG00000213190	MLLT11	1

ENSG00000164659	KIAA1324L	1
ENSG00000152377	SPOCK1	1
ENSG00000082497	SERTAD4	1
ENSG00000121570	DPPA4	1
ENSG00000148143	ZNF462	1
ENSG00000161249	DMKN	1
ENSG00000125285	SOX21	1
ENSG00000043355	ZIC2	1
ENSG00000258947	TUBB3	1
ENSG00000176887	SOX11	1
ENSG00000187772	LIN28B	1
ENSG00000143768	LEFTY2	1
ENSG00000243709	LEFTY1	1
ENSG00000156925	ZIC3	1
ENSG00000105655	ISYNA1	1
ENSG00000144810	COL8A1	1
ENSG00000106278	PTPRZ1	1
ENSG00000128683	GAD1	1
ENSG00000104728	ARHGEF10	1
ENSG00000168268	NT5DC2	1
ENSG00000164434	FABP7	1
ENSG00000133636	NTS	1
ENSG00000181449	SOX2	1
ENSG00000179431	FJX1	1
ENSG00000204335	SP5	1
ENSG00000177519	RPRM	1
ENSG00000124766	SOX4	1
ENSG00000106631	MYL7	1
ENSG00000033170	FUT8	1
ENSG00000136943	CTSV	1
ENSG00000126878	AIF1L	1
ENSG00000122861	PLAU	1
ENSG00000139946	PELI2	1
ENSG00000165588	OTX2	1
ENSG00000104722	NEFM	1
ENSG00000154188	ANGPT1	1
ENSG00000184613	NELL2	1
ENSG00000144834	TAGLN3	1
ENSG00000183145	RIPPLY3	1
ENSG00000088756	ARHGAP28	1

ENSG00000082397	EPB41L3	1
ENSG00000198795	ZNF521	1
ENSG00000138829	FBN2	1
ENSG00000184304	PRKD1	1
ENSG00000006118	TMEM132A	1
ENSG00000095596	CYP26A1	1
ENSG00000138180	CEP55	1
ENSG00000166503		1
ENSG00000067445	TRO	1
ENSG00000151640	DPYSL4	1
ENSG00000128567	PODXL	1
ENSG00000114948	ADAM23	1
ENSG00000144339	TMEFF2	1
ENSG00000150551	LYPD1	1
ENSG00000026025	VIM	1
ENSG00000168280	KIF5C	1
ENSG00000092445	TYRO3	1
ENSG00000088882	CPXM1	1
ENSG00000138795	LEF1	1
ENSG00000116729	WLS	1
ENSG00000041982	TNC	1
ENSG00000153707	PTPRD	1
ENSG00000198467	TPM2	1
ENSG00000162551	ALPL	1
ENSG00000075213	SEMA3A	1
ENSG00000196526	AFAP1	1
ENSG00000154277	UCHL1	1
ENSG00000109255	NMU	1
ENSG00000128596	CCDC136	1
ENSG00000136231	IGF2BP3	1
ENSG00000143320	CRABP2	1
ENSG00000132688	NES	1
ENSG00000134531	EMP1	1
ENSG00000111716	LDHB	1
ENSG00000050165	DKK3	1
ENSG00000076716	GPC4	1
ENSG00000165349	SLC7A3	1
ENSG00000046653	GPM6B	1
ENSG00000123560	PLP1	1
ENSG00000104332	SFRP1	1

ENSG00000103449	SALL1	1
ENSG00000141971	MVB12A	1
ENSG00000143632	ACTA1	1
ENSG00000166086	JAM3	1
ENSG00000149591	TAGLN	1
ENSG00000166426	CRABP1	1
ENSG00000163508	EOMES	1
ENSG00000204711	C9orf135	1
ENSG00000170373	CST1	1
ENSG00000184867	ARMCX2	1
ENSG00000134775	FHOD3	1
ENSG00000163629	PTPN13	1
ENSG00000175928	LRRN1	1
ENSG00000113758	DBN1	1
ENSG00000167657	DAPK3	1
ENSG00000114854	TNNC1	1
ENSG00000136011	STAB2	2
ENSG00000145321	GC	2
ENSG00000151365	THRSP	2
ENSG00000163959	SLC51A	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000213424	KRT222	2
ENSG00000264058	KRT222	2
ENSG00000162267	ITIH3	2
ENSG00000055957	ITIH1	2
ENSG00000112337	SLC17A2	2
ENSG00000164406	LEAP2	2
ENSG00000266964	FXYD1	2
ENSG00000250799	PRODH2	2
ENSG00000159189	C1QC	2
ENSG00000173369	C1QB	2
ENSG00000160862	AZGP1	2
ENSG00000171236	LRG1	2
ENSG00000151655	ITIH2	2
ENSG00000112299	VNN1	2
ENSG00000118514	ALDH8A1	2
ENSG00000131910	NR0B2	2

ENSG00000198099	ADH4	2
ENSG00000151224	MAT1A	2
ENSG00000132693	CRP	2
ENSG00000167711	SERPINF2	2
ENSG00000168509	HFE2	2
ENSG00000182326	C1S	2
ENSG00000160870	CYP3A7	2
ENSG00000160868	CYP3A4	2
ENSG00000143819	EPHX1	2
ENSG00000229314	ORM1	2
ENSG00000121858	TNFSF10	2
ENSG00000136881	BAAT	2
ENSG00000117009	KMO	2
ENSG00000127951	FGL2	2
ENSG00000139547	RDH16	2
ENSG00000069535	MAOB	2
ENSG00000177575	CD163	2
ENSG00000138075	ABCG5	2
ENSG00000130208	APOC1	2
ENSG00000170608	FOXA3	2
ENSG00000161944	ASGR2	2
ENSG00000171840	NINJ2	2
ENSG00000273259	SERPINA3	2
ENSG00000196136	SERPINA3	2
ENSG00000188488	SERPINA5	2
ENSG00000106538	RARRES2	2
ENSG00000129151	BBOX1	2
ENSG00000110887	DAO	2
ENSG00000132840	BHMT2	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000157873	TNFRSF14	2
ENSG00000005471	ABCB4	2
ENSG00000138308	PLA2G12B	2
ENSG00000111249	CUX2	2
ENSG00000169856	ONECUT1	2
ENSG00000117707	PROX1	2
ENSG00000135423	GLS2	2
ENSG00000155659	VSIG4	2
ENSG00000169174	PCSK9	2

ENSG00000154262	ABCA6	2
ENSG00000099769	IGFALS	2
ENSG00000175336	APOF	2
ENSG00000012504	NR1H4	2
ENSG00000169562	GJB1	2
ENSG00000214274	ANG	2
ENSG00000140107	SLC25A47	2
ENSG00000144891	AGTR1	2
ENSG00000105697	HAMP	2
ENSG00000180432	CYP8B1	2
ENSG00000180745	CLRN3	2
ENSG00000144035	NAT8	2
ENSG00000170458	CD14	2
ENSG00000176919	C8G	2
ENSG00000244734	HBB	2
ENSG00000134716	CYP2J2	2
ENSG00000148584	A1CF	2
ENSG00000134716	CYP2J2	2
ENSG00000134716	CYP2J2	2
ENSG00000172955	ADH6	2
ENSG00000165092	ALDH1A1	2
ENSG00000196600	SLC22A25	2
ENSG00000149742	SLC22A9	2
ENSG00000171747	LGALS4	2
ENSG00000160870	CYP3A7	2
ENSG00000160868	CYP3A4	2
ENSG00000133574	GIMAP4	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000144891	AGTR1	2
ENSG00000145850	TIMD4	2
ENSG00000160870	CYP3A7	2
ENSG00000160868	CYP3A4	2
ENSG00000143546	S100A8	2
ENSG00000139547	RDH16	2
ENSG00000139344	AMDHD1	2
ENSG00000105697	HAMP	2
ENSG00000080910	CFHR2	2

ENSG00000141293	SKAP1	2
ENSG00000144035	NAT8	2
ENSG00000129988	LBP	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000117594	HSD11B1	2
ENSG00000244067	GSTA2	2
ENSG00000273259	SERPINA3	2
ENSG00000196136	SERPINA3	2
ENSG00000205403	CFI	2
ENSG00000084734	GCKR	2
ENSG00000198670	LPA	2
ENSG00000187045	TMPRSS6	2
ENSG00000154274	C4orf19	2
ENSG00000108515	ENO3	2
ENSG00000214274	ANG	2
ENSG00000170099	SERPINA6	2
ENSG00000229314	ORM1	2
ENSG00000179761	PIPOX	2
ENSG00000100197	CYP2D6	2
ENSG00000140093	SERPINA10	2
ENSG00000149124	GLYAT	2
ENSG00000188582	PAQR9	2
ENSG00000151726	ACSL1	2
ENSG00000187048	CYP4A11	2
ENSG00000162365	CYP4A22	2
ENSG00000134716	CYP2J2	2
ENSG00000163825	RTP3	2
ENSG00000273259	SERPINA3	2
ENSG00000196136	SERPINA3	2
ENSG00000142484	TM4SF5	2
ENSG00000137204	SLC22A7	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000100197	CYP2D6	2
ENSG00000160282	FTCD	2
ENSG00000144035	NAT8	2
ENSG00000164089	ETNPPL	2

ENSG00000133488	SEC14L4	2
ENSG00000145692	BHMT	2
ENSG00000139547	RDH16	2
ENSG00000165140	FBP1	2
ENSG00000143845	ETNK2	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000214274	ANG	2
ENSG00000105707	HPN	2
ENSG00000186115	CYP4F2	2
ENSG00000171903	CYP4F11	2
ENSG00000186204	CYP4F12	2
ENSG00000118513	MYB	2
ENSG00000118520	ARG1	2
ENSG00000244734	HBB	2
ENSG00000084110	HAL	2
ENSG00000187758	ADH1A	2
ENSG00000248144	ADH1C	2
ENSG00000171759	PAH	2
ENSG00000122862	SRGN	2
ENSG00000139547	RDH16	2
ENSG00000145692	BHMT	2
ENSG00000132437	DDC	2
ENSG00000160255	ITGB2	2
ENSG00000146678	IGFBP1	2
ENSG00000131482	G6PC	2
ENSG00000084674	APOB	2
ENSG00000273259	SERPINA3	2
ENSG00000196136	SERPINA3	2
ENSG00000110077	MS4A6A	2
ENSG00000144035	NAT8	2
ENSG00000141485	SLC13A5	2
ENSG00000141505	ASGR1	2
ENSG00000197249	SERPINA1	2
ENSG00000273259	SERPINA3	2
ENSG00000196136	SERPINA3	2
ENSG00000100652	SLC10A1	2
ENSG00000196139	AKR1C3	2

ENSG00000198610	AKR1C4	2
ENSG00000198650	TAT	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000124568	SLC17A1	2
ENSG00000131187	F12	2
ENSG00000145826	LECT2	2
ENSG00000172497	ACOT12	2
ENSG00000258818	RNASE4	2
ENSG00000055955	ITIH4	2
ENSG00000080618	CPB2	2
ENSG00000120915	EPHX2	2
ENSG00000113600	C9	2
ENSG00000039537	C6	2
ENSG00000136305	CIDEB	2
ENSG00000147647	DPYS	2
ENSG00000132541	HRSP12	2
ENSG00000100197	CYP2D6	2
ENSG00000135447	PPP1R1A	2
ENSG00000163347	CLDN1	2
ENSG00000099937	SERPIND1	2
ENSG00000113924	HGD	2
ENSG00000113790	EHHADH	2
ENSG00000145192	AHSG	2
ENSG00000090512	FETUB	2
ENSG00000113889	KNG1	2
ENSG00000113905	HRG	2
ENSG00000198300	PEG3	2
ENSG00000130988	RGN	2
ENSG00000178772	CPN2	2
ENSG00000002933	TMEM176A	2
ENSG00000175003	SLC22A1	2
ENSG00000146233	CYP39A1	2
ENSG00000125730	C3	2
ENSG00000186115	CYP4F2	2
ENSG00000171903	CYP4F11	2

ENSG00000186204	CYP4F12	2
ENSG00000166278	C2	2
ENSG00000115718	PROC	2
ENSG00000204444	APOM	2
ENSG00000166927	MS4A7	2
ENSG00000110077	MS4A6A	2
ENSG00000110077	MS4A6A	2
ENSG00000165092	ALDH1A1	2
ENSG00000196600	SLC22A25	2
ENSG00000149742	SLC22A9	2
ENSG00000165471	MBL2	2
ENSG00000124713	GNMT	2
ENSG00000152804	HHEX	2
ENSG00000138207	RBP4	2
ENSG00000108242	CYP2C18	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000189221	MAOA	2
ENSG00000148702	HABP2	2
ENSG00000215644	GCGR	2
ENSG00000241935	HOGA1	2
ENSG00000023839	ABCC2	2
ENSG00000120054	CPN1	2
ENSG00000144837	PLA1A	2
ENSG00000163687	DNASE1L3	2
ENSG00000174990	CA5A	2
ENSG00000170439	METTL7B	2
ENSG00000025423	HSD17B6	2
ENSG00000135218	CD36	2
ENSG00000160255	ITGB2	2
ENSG00000165828	PRAP1	2
ENSG00000125246	CLYBL	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000021826	CPS1	2
ENSG00000172482	AGXT	2
ENSG00000072080	SPP2	2
ENSG00000143278	F13B	2

ENSG00000019169	MARCO	2
ENSG00000123838	C4BPA	2
ENSG00000132703	APCS	2
ENSG00000198734	F5	2
ENSG00000196660	SLC30A10	2
ENSG00000007933	FMO3	2
ENSG00000117601	SERPINC1	2
ENSG00000077420	APBB1IP	2
ENSG00000148346	LCN2	2
ENSG00000115919	KYNU	2
ENSG00000106804	C5	2
ENSG00000073734	ABCB11	2
ENSG00000091583	APOH	2
ENSG00000121691	CAT	2
ENSG00000186115	CYP4F2	2
ENSG00000171903	CYP4F11	2
ENSG00000186204	CYP4F12	2
ENSG00000171766	GATM	2
ENSG00000101323	HAO1	2
ENSG00000140284	SLC27A2	2
ENSG00000047457	CP	2
ENSG00000163581	SLC2A2	2
ENSG00000114771	AADAC	2
ENSG00000114200	BCHE	2
ENSG00000169903	TM4SF4	2
ENSG00000116882	HAO2	2
ENSG00000134240	HMGCS2	2
ENSG00000171560	FGA	2
ENSG00000151790	TDO2	2
ENSG00000131781	FMO5	2
ENSG00000108924	HLF	2
ENSG00000138823	MTTP	2
ENSG00000116791	CRYZ	2
ENSG00000136872	ALDOB	2
ENSG00000106927	AMBP	2
ENSG00000229314	ORM1	2
ENSG00000132855	ANGPTL3	2
ENSG00000134716	CYP2J2	2
ENSG00000134709	HOOK1	2
ENSG00000187048	CYP4A11	2

ENSG00000162365	CYP4A22	2
ENSG00000187048	CYP4A11	2
ENSG00000162365	CYP4A22	2
ENSG00000159423	ALDH4A1	2
ENSG00000187017	ESPN	2
ENSG00000009724	MASP2	2
ENSG00000109758	HGFAC	2
ENSG00000138030	KHK	2
ENSG00000109819	PPARGC1A	2
ENSG00000134962	KLB	2
ENSG00000196620	UGT2B15	2
ENSG00000156096	UGT2B4	2
ENSG00000163631	ALB	2
ENSG00000079557	AFM	2
ENSG00000158104	HPD	2
ENSG00000160870	CYP3A7	2
ENSG00000160868	CYP3A4	2
ENSG00000021852	C8B	2
ENSG00000106327	TFR2	2
ENSG00000106258	CYP3A5	2
ENSG00000105852	PON3	2
ENSG00000106565	TMEM176B	2
ENSG00000144035	NAT8	2
ENSG00000144908	ALDH1L1	2
ENSG00000111181	SLC6A12	2
ENSG00000175899	A2M	2
ENSG00000111796	KLRB1	2
ENSG00000165682	CLEC1B	2
ENSG00000139144	PIK3C2G	2
ENSG00000134538	SLCO1B1	2
ENSG00000111700	SLCO1B3	2
ENSG00000157103	SLC6A1	2
ENSG00000111796	KLRB1	2
ENSG00000111796	KLRB1	2
ENSG00000105398	SULT2A1	2
ENSG00000083807	SLC27A5	2
ENSG00000021488	SLC7A9	2
ENSG00000011600	TYROBP	2
ENSG00000137491	SLCO2B1	2
ENSG00000062282	DGAT2	2

ENSG00000110169	HPX	2
ENSG00000066813	ACSM2B	2
ENSG00000101981	F9	2
ENSG00000036473	OTC	2
ENSG00000123561	SERPINA7	2
ENSG00000126218	F10	2
ENSG00000126231	PROZ	2
ENSG00000009950	MLXIPL	2
ENSG00000104760	FGL1	2
ENSG00000109511	ANXA10	2
ENSG00000164344	KLKB1	2
ENSG00000088926	F11	2
ENSG00000257017	HP	2
ENSG00000261701	HPR	2
ENSG00000187193	MT1X	2
ENSG00000198417	MT1F	2
ENSG00000205364	MT1M	2
ENSG00000125144	MT1G	2
ENSG00000086696	HSD17B2	2
ENSG00000166816	LDHD	2
ENSG00000135744	AGT	2
ENSG00000110243	APOA5	2
ENSG00000110245	APOC3	2
ENSG00000103569	AQP9	2
ENSG00000166035	LIPC	2
ENSG00000158874	APOA2	2
ENSG00000138115	CYP2C8	2
ENSG00000138109	CYP2C9	2
ENSG00000129596	CDO1	2
ENSG00000100024	UPB1	2
ENSG00000144035	NAT8	2
ENSG00000196139	AKR1C3	2
ENSG00000198610	AKR1C4	2
ENSG00000171557	FGG	2
ENSG00000211452	DIO1	2
ENSG00000245848	CEBPA	2
ENSG00000157131	C8A	2
ENSG00000060566	CREB3L3	2
ENSG00000139547	RDH16	2
ENSG00000100197	CYP2D6	2

ENSG00000091513	TF	2
ENSG00000100197	CYP2D6	2
ENSG00000113492	AGXT2	2
ENSG00000174567	GOLT1A	2
ENSG00000100197	CYP2D6	2
ENSG00000125398	SOX9	3
ENSG00000121440	PDZRN3	3
ENSG00000132854	KANK4	3
ENSG00000129514	FOXA1	3
ENSG00000120708	TGFBI	3
ENSG00000165379	LRFN5	3
ENSG00000122691	TWIST1	3
ENSG00000111913	FAM65B	3
ENSG00000139174	PRICKLE1	3
ENSG00000172380	GNG12	3
ENSG00000183876	ARSI	3
ENSG00000139329	LUM	3
ENSG00000163171	CDC42EP3	3
ENSG00000180318	ALX1	3
ENSG00000170955	PRKCDBP	3
ENSG00000108244	KRT23	3
ENSG00000125850	OVOL2	3
ENSG00000141696	LEPREL4	3
ENSG00000113196	HAND1	3
ENSG00000106366	SERPINE1	3
ENSG00000129757	CDKN1C	3
ENSG00000125384	PTGER2	3
ENSG00000109472	CPE	3
ENSG00000203857	HSD3B1	3
ENSG00000262655	SPON1	3
ENSG00000164107	HAND2	3
ENSG00000106031	HOXA13	3
ENSG00000141744	PNMT	3
ENSG00000211448	DIO2	3
ENSG00000117152	RGS4	3
ENSG00000198121	LPAR1	3
ENSG00000146197	SCUBE3	3
ENSG00000260027	HOXB7	3
ENSG00000130812	ANGPTL6	3
ENSG00000150687	PRSS23	3

ENSG00000104723	TUSC3	3
ENSG00000144642	RBMS3	3
ENSG00000143036	SLC44A3	3
ENSG00000132698	RAB25	3
ENSG00000010932	FMO1	3
ENSG00000142227	EMP3	3
ENSG00000164683	HEY1	3
ENSG00000171729	TMEM51	3
ENSG00000084636	COL16A1	3
ENSG00000185559	DLK1	3
ENSG00000144645	OSBPL10	3
ENSG00000168743	NPNT	3
ENSG00000127472	PLA2G5	3
ENSG00000001626	CFTR	3
ENSG00000100739	BDKRB1	3
ENSG00000166900	STX3	3
ENSG00000152137	HSPB8	3
ENSG00000122176	FMOD	3
ENSG00000163362	C1orf106	3
ENSG00000168032	ENTPD3	3
ENSG00000148942	SLC5A12	3
ENSG00000126016	AMOT	3
ENSG00000123700	KCNJ2	3
ENSG00000183853	KIRREL	3
ENSG00000174307	PHLDA3	3
ENSG00000134369	NAV1	3
ENSG00000110660	SLC35F2	3
ENSG00000138615	CILP	3
ENSG00000189334	S100A14	3
ENSG00000166482	MFAP4	3
ENSG00000157502	MUM1L1	3
ENSG00000070019	GUCY2C	3
ENSG00000185222	WBP5	3
ENSG00000103241	FOXF1	3
ENSG00000164318	EGFLAM	3
ENSG00000125872	LRRN4	3
ENSG00000100079	LGALS2	3
ENSG00000077616	NAALAD2	3
ENSG00000132386	SERPINF1	3
ENSG00000046604	DSG2	3

ENSG00000164825	DEFB1	3
ENSG00000105989	WNT2	3
ENSG00000174482	LINGO2	3
ENSG00000100196	KDELR3	3
ENSG00000167874	TMEM88	3
ENSG00000119888	EPCAM	3
ENSG00000188176	SMTNL2	3
ENSG00000176788	BASP1	3
ENSG00000168952	STXBP6	3
ENSG00000013588	GPRC5A	3
ENSG00000136026	CKAP4	3
ENSG00000167434	CA4	3
ENSG00000186377	CYP4X1	3
ENSG00000152049	KCNE4	3
ENSG00000079257	LXN	3
ENSG00000166394	CYB5R2	3
ENSG00000139926	FRMD6	3
ENSG00000183722	LHFP	3
ENSG00000143867	OSR1	3
ENSG00000163132	MSX1	3
ENSG00000138119	MYOF	3
ENSG00000173376	NDNF	3
ENSG00000164484	TMEM200A	3
ENSG00000167755	KLK6	3
ENSG00000172458	IL17D	3
ENSG00000180340	FZD2	3
ENSG00000187123	LYPD6	3
ENSG00000171951	SCG2	3
ENSG00000148344	PTGES	3
ENSG00000179178	TMEM125	3
ENSG00000152661	GJA1	3
ENSG00000134258	VTCN1	3
ENSG00000174099	MSRB3	3
ENSG00000153822	KCNJ16	3
ENSG00000112852	PCDHB2	3
ENSG00000147655	RSPO2	3
ENSG00000107317	PTGDS	3
ENSG00000169583	CLIC3	3
ENSG00000103888	CEMIP	3
ENSG00000184005	ST6GALNAC3	3

ENSG00000151693	ASAP2	3
ENSG00000180447	GAS1	3
ENSG00000147003	TMEM27	3
ENSG00000086548	CEACAM6	3
ENSG00000164932	CTHRC1	3
ENSG00000183160	TMEM119	3
ENSG00000107485	GATA3	3
ENSG00000081051	AFP	3
ENSG00000155465	SLC7A7	3
ENSG00000147257	GPC3	3
ENSG00000112964	GHR	3
ENSG00000171812	COL8A2	3
ENSG00000176907	C8orf4	3
ENSG00000184347	SLIT3	3
ENSG00000174807	CD248	3
ENSG00000168461	RAB31	3
ENSG00000139278	GLIPR1	3
ENSG00000071282	LMCD1	3
ENSG00000181541	MAB21L2	3
ENSG00000129116	PALLD	3
ENSG00000198108	CHSY3	3
ENSG00000164946	FREM1	3
ENSG00000163017	ACTG2	3
ENSG00000137573	SULF1	3
ENSG00000124107	SLPI	3
ENSG00000108797	CNTNAP1	3
ENSG00000084207	GSTP1	3
ENSG00000062038	CDH3	3
ENSG00000185565	LSAMP	3
ENSG00000118271	TTR	3
ENSG00000171004	HS6ST2	3
ENSG00000203857	HSD3B1	3
ENSG00000135736	CCDC102A	3
ENSG00000203857	HSD3B1	3
ENSG00000167617	CDC42EP5	3
ENSG00000163520	FBLN2	3
ENSG00000144619	CNTN4	3
ENSG00000109099	PMP22	3
ENSG00000170425	ADORA2B	3
ENSG00000142156	COL6A1	3

ENSG00000113140	SPARC	3
ENSG00000136193	SCRN1	3
ENSG00000159251	ACTC1	3
ENSG00000015413	DPEP1	3
ENSG00000128591	FLNC	3
ENSG00000167695	FAM57A	3
ENSG00000142552	RCN3	3
ENSG00000116017	ARID3A	3
ENSG00000118495	PLAGL1	3
ENSG00000154856	APCDD1	3
ENSG00000165794	SLC39A2	3
ENSG00000118503	TNFAIP3	3
ENSG00000135549	PKIB	3
ENSG00000146374	RSPO3	3
ENSG00000011465	DCN	3
ENSG00000184838	PRR16	3
ENSG00000049130	KITLG	3
ENSG00000166165	CKB	3
ENSG00000147883	CDKN2B	3
ENSG00000079931	MOXD1	3
ENSG00000167642	SPINT2	3
ENSG00000125266	EFNB2	3
ENSG00000156510	HKDC1	3
ENSG00000166923	GREM1	3
ENSG00000130176	CNN1	3
ENSG00000198832		3
ENSG00000142173	COL6A2	3
ENSG00000120820	GLT8D2	3
ENSG00000007384	RHBDF1	3
ENSG00000113739	STC2	3
ENSG00000070404	FSTL3	3
ENSG00000131459	GFPT2	3
ENSG00000146674	IGFBP3	3
ENSG00000115380	EFEMP1	3
ENSG00000251493	FOXD1	3
ENSG00000132031	MATN3	3
ENSG00000108821	COL1A1	3
ENSG00000130508	PXDN	3
ENSG00000005884	ITGA3	3
ENSG00000182534	MXRA7	3

ENSG00000141756	FKBP10	3
ENSG00000171345	KRT19	3
ENSG00000037965	HOXC8	3
ENSG00000067057	PFKP	3
ENSG00000121297	TSHZ3	3
ENSG00000079689	SCGN	3
ENSG00000078401	EDN1	3
ENSG00000120149	MSX2	3
ENSG00000069011	PITX1	3
ENSG00000145824	CXCL14	3
ENSG00000135052	GOLM1	3
ENSG00000118971	CCND2	3
ENSG00000145681	HAPLN1	3
ENSG00000038427	VCAN	3
ENSG00000164251	F2RL1	3
ENSG00000110195	FOLR1	3
ENSG00000164294	GPX8	3
ENSG00000133661	SFTPD	3
ENSG00000087303	NID2	3
ENSG00000163947	ARHGEF3	3
ENSG00000114251	WNT5A	3
ENSG00000116991	SIPA1L2	3
ENSG00000100842	EFS	3
ENSG00000145555	MYO10	3
ENSG00000040731	CDH10	3
ENSG00000160207	HSF2BP	3
ENSG00000091986	CCDC80	3
ENSG00000019549	SNAI2	3
ENSG00000099960	SLC7A4	3
ENSG00000163430	FSTL1	3
ENSG00000154734	ADAMTS1	3
ENSG00000183844	FAM3B	3
ENSG00000159212	CLIC6	3
ENSG00000050438	SLC4A8	3
ENSG00000135480	KRT7	3
ENSG00000110811	LEPREL2	3
ENSG00000103175	WFDC1	3
ENSG00000112562	SMOC2	3
ENSG00000184697	CLDN6	3
ENSG00000112818	MEP1A	3

ENSG00000146072	TNFRSF21	3
ENSG00000112655	PTK7	3
ENSG00000132205	EMILIN2	3
ENSG00000150938	CRIM1	3
ENSG00000161896	IP6K3	3
ENSG00000006747	SCIN	3
ENSG00000113657	DPYSL3	3
ENSG00000113083	LOX	3
ENSG00000064692	SNCAIP	3
ENSG00000132915	PDE6A	3
ENSG00000135046	ANXA1	3
ENSG00000148677	ANKRD1	3
ENSG00000175315	CST6	3
ENSG00000172638	EFEMP2	3
ENSG00000138193	PLCE1	3
ENSG00000169129	AFAP1L2	3
ENSG00000165868	HSPA12A	3
ENSG00000169594	BNC1	3
ENSG00000141526	SLC16A3	3
ENSG00000091136	LAMB1	3
ENSG00000166863	TAC3	3
ENSG00000114270	COL7A1	3
ENSG00000204262	COL5A2	3
ENSG00000168542	COL3A1	3
ENSG00000115602	IL1RL1	3
ENSG00000115604	IL18R1	3
ENSG00000173559	NABP1	3
ENSG00000119900	OGFRL1	3
ENSG00000135919	SERPINE2	3
ENSG00000145730	PAM	3
ENSG00000118194	TNNT2	3
ENSG00000167535	CACNB3	3
ENSG00000150556	LYPD6B	3
ENSG00000077943	ITGA8	3
ENSG00000130635	COL5A1	3
ENSG00000050555	LAMC3	3
ENSG00000162998	FRZB	3
ENSG00000198648	STK39	3
ENSG00000149090	PAMR1	3
ENSG00000106348	IMPDH1	3

ENSG00000166147	FBN1	3
ENSG00000166963	MAP1A	3
ENSG00000166145	SPINT1	3
ENSG00000115008	IL1A	3
ENSG00000171033	PKIA	3
ENSG00000104435	STMN2	3
ENSG00000101198	NKAIN4	3
ENSG00000133110	POSTN	3
ENSG00000116774	OLFML3	3
ENSG00000203857	HSD3B1	3
ENSG00000164125	FAM198B	3
ENSG00000164124	TMEM144	3
ENSG00000060718	COL11A1	3
ENSG00000164093	PITX2	3
ENSG00000108947	EFNB3	3
ENSG00000117525	F3	3
ENSG00000079112	CDH17	3
ENSG00000204291	COL15A1	3
ENSG00000165124	SVEP1	3
ENSG00000182752	PAPPA	3
ENSG00000173068	BNC2	3
ENSG00000136859	ANGPTL2	3
ENSG00000162493	PDPN	3
ENSG00000116157	GPX7	3
ENSG00000142583	SLC2A5	3
ENSG00000157005	SST	3
ENSG00000095752	IL11	3
ENSG00000162576	MXRA8	3
ENSG00000007062	PROM1	3
ENSG00000168824		3
ENSG00000197712	FAM114A1	3
ENSG00000163288	GABRB1	3
ENSG00000069702	TGFBR3	3
ENSG00000145287	PLAC8	3
ENSG00000163624	CDS1	3
ENSG00000138772	ANXA3	3
ENSG00000240583	AQP1	3
ENSG00000078098	FAP	3
ENSG00000165556	CDX2	3
ENSG00000164692	COL1A2	3

ENSG00000105825	TFPI2	3
ENSG00000122585	NPY	3
ENSG00000105928	DFNA5	3
ENSG00000105894	PTN	3
ENSG00000119630	PGF	3
ENSG00000177875	CCDC184	3
ENSG00000163637	PRICKLE2	3
ENSG00000163638	ADAMTS9	3
ENSG00000156711	MAPK13	3
ENSG00000197614	MFAP5	3
ENSG00000173391	OLR1	3
ENSG00000111339	ART4	3
ENSG00000139055	ERP27	3
ENSG00000198759	EGFL6	3
ENSG00000121361	KCNJ8	3
ENSG00000123096	SSPN	3
ENSG00000111319	SCNN1A	3
ENSG00000104044	OCA2	3
ENSG00000086991	NOX4	3
ENSG00000140545	MFGE8	3
ENSG00000197702	PARVA	3
ENSG00000140682	TGFB1I1	3
ENSG00000052344	PRSS8	3
ENSG00000104415	WISP1	3
ENSG00000102034	ELF4	3
ENSG00000102243	VGLL1	3
ENSG00000089472	HEPH	3
ENSG00000102359	SRPX2	3
ENSG00000188153	COL4A5	3
ENSG00000130224	LRCH2	3
ENSG00000079215	SLC1A3	3
ENSG00000133135	RNF128	3
ENSG00000020181	GPR124	3
ENSG00000187498	COL4A1	3
ENSG00000134871	COL4A2	3
ENSG00000150630	VEGFC	3
ENSG00000104368	PLAT	3
ENSG00000168615	ADAM9	3
ENSG00000145147	SLIT2	3
ENSG00000218336	TENM3	3

ENSG00000104213	PDGFRL	3
ENSG00000164106	SCRG1	3
ENSG00000164120	HPGD	3
ENSG00000140937	CDH11	3
ENSG00000103196	CRISPLD2	3
ENSG00000140945	CDH13	3
ENSG00000103064	SLC7A6	3
ENSG00000154127	UBASH3B	3
ENSG00000166250	CLMP	3
ENSG00000109846	CRYAB	3
ENSG00000182985	CADM1	3
ENSG00000110244	APOA4	3
ENSG00000118137	APOA1	3
ENSG00000065911	MTHFD2	3
ENSG00000112175	BMP5	3
ENSG00000182718	ANXA2	3
ENSG00000140465	CYP1A1	3
ENSG00000111799	COL12A1	3
ENSG00000129038	LOXL1	3
ENSG00000137269	LRRC1	3
ENSG00000174498	IGDCC3	3
ENSG00000158258	CLSTN2	3
ENSG00000114113	RBP2	3
ENSG00000073756	PTGS2	3
ENSG00000153558	FBXL2	3
ENSG00000164764	SBSPON	3
ENSG00000101680	LAMA1	3
ENSG00000103742	IGDCC4	3
ENSG00000173546	CSPG4	3
ENSG00000107165	TYRP1	3
ENSG00000126947	ARMCX1	3
ENSG00000146250	PRSS35	3
ENSG00000182636	NDN	3
ENSG00000171564	FGB	3
ENSG00000164116	GUCY1A3	3
ENSG00000135750	KCNK1	3
ENSG00000099994	SUSD2	3
ENSG00000166173	LARP6	3
ENSG00000118407	FILIP1	3
ENSG00000077942	FBLN1	3

ENSG00000169116	PARM1	3
ENSG00000197635	DPP4	3
ENSG00000101955	SRPX	3
ENSG00000206538	VGLL3	3
ENSG00000183036	PCP4	3

Table 10-2: Metagenes from NMF analysis on mouse liver development data

Ensembl_ID	Symbol	Metagene
ENSMUSG00000035540	Gc	1
ENSMUSG00000036216	Leap2	1
ENSMUSG00000036381	P2ry14	1
ENSMUSG00000037053	Azgp1	1
ENSMUSG00000037798	Mat1a	1
ENSMUSG00000037942	Crp	1
ENSMUSG00000038641	Akr1d1	1
ENSMUSG00000038656	Cyp3a16	1
ENSMUSG00000038656	Cyp3a16	1
ENSMUSG00000038656	Cyp3a16	1
ENSMUSG00000039476	Prrx2	2
ENSMUSG00000040134	Rdh7	1
ENSMUSG00000040809	Chil3	1
ENSMUSG00000041660	Bbox1	1
ENSMUSG00000042118	Bhmt2	1
ENSMUSG00000044206	Vsig4	2
ENSMUSG00000010601	Apol7a	1
ENSMUSG00000010601	Apol7a	1
ENSMUSG00000010601	Apol7a	1
ENSMUSG00000010601	Apol7a	1
ENSMUSG00000044749	Abca6	1
ENSMUSG00000045179	Sox3	2
ENSMUSG00000048217	Nags	1
ENSMUSG00000052131	Akr1b7	1
ENSMUSG00000052595	A1cf	1
ENSMUSG00000052974	Cyp2f2	1
ENSMUSG00000054417	Cyp3a44	1
ENSMUSG00000054417	Cyp3a44	1
ENSMUSG00000054417	Cyp3a44	1
ENSMUSG00000054630	Ugt2b5	1

ENSMUSG00000054630	Ugt2b5	1
ENSMUSG00000055782	Abcd2	1
ENSMUSG00000056035	Cyp3a11	1
ENSMUSG00000056035	Cyp3a11	1
ENSMUSG00000056035	Cyp3a11	1
ENSMUSG00000057037	Cfhr1	1
ENSMUSG00000015970	Chdh	1
ENSMUSG00000057425	Ugt2b37	1
ENSMUSG00000057425	Ugt2b37	1
ENSMUSG00000058207	Serpina3k	1
ENSMUSG00000058207	Serpina3k	1
ENSMUSG00000060407	Cyp2a12	1
ENSMUSG00000060407	Cyp2a12	1
ENSMUSG00000062410	Hsd3b3	1
ENSMUSG00000062410	Hsd3b3	1
ENSMUSG00000019232	Etnppl	1
ENSMUSG00000069805	Fbp1	1
ENSMUSG00000070704	Ugt2b36	1
ENSMUSG00000070704	Ugt2b36	1
ENSMUSG00000072601	Ear1	1
ENSMUSG00000072601	Ear1	1
ENSMUSG00000019987	Arg1	1
ENSMUSG00000020010	Vnn3	1
ENSMUSG00000074207	Adh1	1
ENSMUSG00000074207	Adh1	1
ENSMUSG00000020072	Pbld1	1
ENSMUSG00000074637	Sox2	2
ENSMUSG00000075334	Rprm	2
ENSMUSG00000078650	G6pc	1
ENSMUSG00000079012	Serpina3m	2
ENSMUSG00000079012	Serpina3m	2
ENSMUSG00000020641	Rsad2	1
ENSMUSG00000020884	Asgr1	1
ENSMUSG00000021091	Serpina3n	1
ENSMUSG00000021091	Serpina3n	1
ENSMUSG00000021135	Slc10a1	1
ENSMUSG00000021210	Akr1c6	1
ENSMUSG00000021210	Akr1c6	1
ENSMUSG00000021210	Akr1c6	1
ENSMUSG00000021390	Ogn	1

ENSMUSG00000021506	Pitx1	2
ENSMUSG00000021508	Cxcl14	2
ENSMUSG00000021922	Itih4	1
ENSMUSG00000021999	Cpb2	1
ENSMUSG00000022445	Cyp2d26	1
ENSMUSG00000022512	Cldn1	1
ENSMUSG00000022868	Ahsg	1
ENSMUSG00000022871	Fetub	1
ENSMUSG00000024863	Mbl2	1
ENSMUSG00000025003	Cyp2c39	1
ENSMUSG00000025003	Cyp2c39	1
ENSMUSG00000025003	Cyp2c39	1
ENSMUSG00000025194	Abcc2	1
ENSMUSG00000025478	Dpysl4	2
ENSMUSG00000025479	Cyp2e1	1
ENSMUSG00000003053	Cyp2c29	1
ENSMUSG00000003053	Cyp2c29	1
ENSMUSG00000003053	Cyp2c29	1
ENSMUSG00000025911	Adhfe1	1
ENSMUSG00000025991	Cps1	1
ENSMUSG00000026368	F13b	1
ENSMUSG00000026542	Apcs	1
ENSMUSG00000026822	Lcn2	1
ENSMUSG00000027048	Abcb11	1
ENSMUSG00000000049	Apoh	1
ENSMUSG00000027261	Hao1	1
ENSMUSG00000003617	Cp	1
ENSMUSG00000027761	Aadac	1
ENSMUSG00000027870	Hao2	1
ENSMUSG00000028553	Angptl3	1
ENSMUSG00000028715	Cyp4a14	1
ENSMUSG00000028715	Cyp4a14	1
ENSMUSG00000029368	Alb	1
ENSMUSG00000029445	Hpd	1
ENSMUSG00000029556	Hnf1a	1
ENSMUSG00000029630	Cyp3a25	1
ENSMUSG00000029630	Cyp3a25	1
ENSMUSG00000029630	Cyp3a25	1
ENSMUSG00000029656	C8b	1
ENSMUSG00000029727	Cyp3a13	1

ENSMUSG00000030111	A2m	1
ENSMUSG00000004885	Crabp2	2
ENSMUSG00000030236	Slco1b2	1
ENSMUSG00000030236	Slco1b2	1
ENSMUSG00000030244	Gys2	1
ENSMUSG00000030382	Slc27a5	1
ENSMUSG00000030413	Pglyrp1	1
ENSMUSG00000030909	Anks4b	1
ENSMUSG00000031138	F9	1
ENSMUSG00000031173	Otc	1
ENSMUSG00000031271	Serpina7	1
ENSMUSG00000031594	Fgl1	1
ENSMUSG00000031640	Klkb1	1
ENSMUSG00000031645	F11	1
ENSMUSG00000031722	Hp	1
ENSMUSG00000031722	Hp	1
ENSMUSG00000032081	Apoc3	1
ENSMUSG00000032291	Crabp1	2
ENSMUSG00000033533	Acsm1	1
ENSMUSG00000033715	Akr1c14	1
ENSMUSG00000033715	Akr1c14	1
ENSMUSG00000033715	Akr1c14	1
ENSMUSG00000034435	Tmem30b	1
ENSMUSG00000096852	Cyp2d12	1

Table 10-3: Differential expression analysis using SAM on HLCs and mature cells (E19.5, PHH)

Gene Name	Fold Change	q-value(%)
ARG1	19.476484	0
ADH1A	17.694997	0
APCS	17.035175	0
GC	15.791466	0
FETUB	14.363967	0
ADH1C	14.036458	0
AZGP1	14.035046	0
CYP2C9	13.929656	0
HPR	13.095323	0
ITIH4	12.987485	0

F9	12.696935	0
HAMP	12.495787	0
ITIH3	12.132906	0
RARRES2	11.829739	0
CYP2C9	11.723588	0
CYP2C9	11.567727	0
FBP1	11.267267	0
CYP8B1	11.098447	0
KLKB1	10.890199	0
CFHR2	10.693478	0
SLCO1B1	10.692924	0
CYP2C9	10.240908	0
HAO1	10.194564	0
G6PC	10.074021	0
BAAT	9.9903307	0
C4BPA	9.9514221	0
F11	9.8224018	0
CYP3A4	9.7875974	0
LRG1	9.3539386	0
SLCO1B3	8.9833948	0
C3	8.7114759	0
CYP3A5	8.6479512	0
CRP	8.603505	0
CYP3A4	8.5868513	0
SEC14L4	8.462595	0
MT1G	8.3981339	0
C8A	8.3685927	0
C8G	8.1502165	0
AKR1C4	7.9824454	0
CYP3A4	7.7964322	0
AQP9	7.7606928	0
AKR1C4	7.699717	0
BHMT2	7.66797	0
MBL2	7.6673511	0
ETNPPL	7.6374664	0
CYP2C19	7.5933779	0
CYP3A4	7.5844231	0
THRSP	7.5609523	0
SLC2A2	7.5479197	0
CPB2	7.5173972	0

UGT2B15	7.4916742	0
CYP4A22	7.3322409	0
CYP4A11	7.3322409	0
AGXT	7.3186786	0
HFE2	7.2941923	0
CYP2D6	6.9804188	0
F12	6.977067	0
AKR1C4	6.830319	0
CYP2C19	6.8180732	0
MT1M	6.8036695	0
RDH16	6.7900646	0
SLC27A5	6.5930516	0
CYP4A22	6.4774698	0
CYP4A11	6.4774698	0
PAH	6.4511817	0
MT1F	6.4198348	0
IGFALS	6.37437	0
CYP2C19	6.368176	0
AKR1C1	6.1601392	0
CYP2D6	6.1250017	0
UGT2B15	5.9965232	0
CYP4A11	5.9212385	0
CYP4A22	5.9212385	0
CYP2C19	5.9135061	0
AKR1C1	5.9103627	0
IGFBP1	5.8699256	0
HAO2	5.8604568	0
ABCG5	5.8573298	0
UPB1	5.6787492	0
CYP2D6	5.615751	0
LDHD	5.5948737	0
CYP2D6	5.5853906	0
APOA5	5.5714422	0
AOX1	5.5522439	0
CYP2J2	5.5331258	0
MASP2	5.4683543	0
NAT8	5.3984581	0
SLC17A2	5.3482406	0
UGT2B15	5.2699503	0
AKR1C4	5.1879656	0

NR1I2	5.1743157	0
UGT2B15	5.1706553	0
CYP4F2	5.1631661	0
CYP2C8	5.1569698	0
HGFAC	5.0527051	0
LPA	5.0428147	0
ONECUT1	5.0363388	0
CYP4F12	5.0319708	0
CYP2D6	4.9414974	0
ABCA6	4.9184354	0
AKR1C1	4.8803929	0
GLS2	4.8477262	0
UGT2B15	4.8466351	0
HABP2	4.8361336	0
UGT2B15	4.7937836	0
C2	4.7359731	0
AKR1C4	4.7129475	0
CYP4F2	4.7102656	0
TAT	4.7032425	0
ABCB11	4.6918663	0
VNN1	4.6077168	0
F13B	4.571373	0
ONECUT2	4.5543219	0
CYP2C8	4.5194386	0
RTP3	4.4910449	0
CYP4F12	4.4880117	0
RDH16	4.4414993	0
PEMT	4.402901	0
CYP2C8	4.3383354	0
CSTA	4.3082668	0
CD14	4.2627324	0
SLC25A47	4.2280517	0
ACOT12	4.1566647	0
CYP39A1	4.1458989	0
MLXIPL	4.130333	0
SDS	4.1059429	0
PAQR9	4.0143963	0

Table 10-4: Dynamic differentially expressed genes between human *in vitro* differentiation and mouse development (E9.5 to E15.5)

Gene	Spearman	Pearson	Euclidean	Foldchange_human	Foldchange_mouse
TPD52L1	-1.00	-1.00	8.20	3.29	2.42
SOX9	-1.00	-0.97	6.47	3.67	2.69
CA4	-1.00	-0.99	12.77	5.02	2.93
COL1A1	-1.00	-1.00	10.53	5.88	1.17
FOXM1	-1.00	-0.96	6.18	2.56	0.11
RIN2	-1.00	-1.00	7.20	2.97	1.45
LAMB1	-1.00	-0.98	8.43	3.27	2.53
INMT-FAM188B	-0.88	-0.90	8.72	3.39	0.19
INMT	-0.88	-0.90	8.72	3.39	0.19
ANGPTL2	-1.00	-0.97	6.67	4.64	1.99
CLEC11A	-1.00	-0.99	6.61	2.79	0.74
MAPK13	-1.00	-1.00	7.06	2.11	1.63
WISP1	-1.00	-1.00	7.10	4.69	1.85
SPC25	-1.00	-0.98	6.57	2.61	0.99
SLC1A3	-1.00	-1.00	6.15	2.94	2.01
TYRP1	-1.00	-1.00	10.58	4.49	0.64
FBLN1	-1.00	-1.00	9.12	1.82	3.44
SCUBE2	-1.00	-0.99	7.18	3.65	1.54
TLE1	-1.00	-1.00	6.54	2.19	0.65
RAB25	-1.00	-1.00	9.01	2.63	3.26
ODC1	-1.00	-1.00	8.15	2.95	0.57
GATA5	-1.00	-0.95	8.66	3.40	1.18
BIRC5	-1.00	-1.00	5.88	2.17	0.50
MMP9	-1.00	-0.99	7.25	2.60	0.97
GPX3	-1.00	-0.99	7.09	3.20	2.98
CLDN7	-1.00	-0.99	7.64	2.74	2.99
SPARC	-1.00	-0.98	9.98	6.05	1.12
P4HA2	-1.00	-0.99	7.01	2.84	1.70
E2F2	-1.00	-0.98	6.31	2.49	2.25
DPEP1	-1.00	-1.00	8.17	2.50	2.24
SOCS2	-0.93	-0.90	9.10	2.58	1.06
IGF1	-1.00	-0.99	6.02	2.83	1.58
IGFBP3	-1.00	-0.99	11.78	7.48	2.05
MTHFD1	-1.00	-0.97	8.03	1.58	2.41
MSX2	-0.93	-0.89	6.68	3.94	2.89
CXCL14	-1.00	-1.00	6.81	6.91	3.97

VCAN	-0.88	-0.91	6.07	2.97	1.74
NID2	-1.00	-0.99	9.80	7.74	1.52
WNT5A	-1.00	-1.00	8.46	3.83	2.97
LCP1	-1.00	-0.99	17.44	5.72	1.48
GPC5	-1.00	-1.00	7.51	4.47	0.46
CTNND2	-1.00	-1.00	6.89	2.16	2.12
ZFPM2	-1.00	-1.00	6.43	2.97	2.00
ANGPT1	-1.00	-1.00	7.69	3.27	2.45
CDH10	-1.00	-1.00	9.99	5.28	1.36
COL2A1	-1.00	-0.97	6.11	2.44	4.14
PPP1R1A	-1.00	-1.00	6.18	3.72	2.62
DGAT1	-1.00	-1.00	7.96	2.41	0.45
ARC	-1.00	-0.98	6.23	0.72	2.14
SNAI2	-1.00	-1.00	11.54	5.40	3.32
PROS1	-1.00	-0.97	9.90	3.82	1.11
CLIC6	-1.00	-1.00	7.55	3.35	2.78
RCAN1	-1.00	-1.00	6.99	2.50	1.16
SMOC2	-1.00	-0.98	5.76	3.83	1.81
CBS	-1.00	-1.00	6.81	3.70	2.54
EPB41L3	-1.00	-1.00	6.69	2.42	2.30
ANKS1A	-1.00	-1.00	5.63	3.29	0.49
BAMBI	-0.94	-0.91	7.82	2.69	1.65
SYT4	-1.00	-1.00	7.90	4.13	1.21
BIN1	-1.00	-1.00	5.92	2.70	1.55
SNCAIP	-1.00	-1.00	7.30	4.13	1.47
PPIC	-1.00	-0.99	6.58	2.86	1.56
ASRGL1	-1.00	-0.98	5.95	2.78	1.64
SLC1A1	-1.00	-0.97	5.64	2.74	0.60
HSPA12A	-1.00	-0.99	6.38	2.27	1.03
BNC1	-1.00	-1.00	5.74	3.71	1.79
PSTPIP2	-1.00	-0.98	6.24	2.44	0.94
TK1	-1.00	-1.00	6.95	2.81	1.23
COL3A1	-1.00	-0.98	16.61	7.13	1.79
IGFBP5	-1.00	-0.99	8.86	4.59	4.06
RGS1	-1.00	-0.99	5.67	2.87	0.46
RGS2	-0.95	-0.96	7.16	3.42	1.39
KIFAP3	-1.00	-1.00	7.62	2.14	2.06
FRZB	-1.00	-0.99	10.57	3.12	2.95
PAMR1	-0.93	-0.92	5.72	5.80	3.25
MDK	-1.00	-0.99	7.17	1.84	3.30

SPINT1	-1.00	-1.00	8.26	2.24	2.98
BMP2	-1.00	-1.00	9.25	3.00	0.17
CD93	-1.00	-0.99	7.17	2.58	1.23
PKIA	-1.00	-1.00	7.83	3.79	2.25
POSTN	-1.00	-0.99	9.23	4.44	3.34
SERPINI1	-1.00	-1.00	5.69	2.22	1.62
FAM198B	-1.00	-0.98	9.51	3.78	2.42
CCDC109B	-1.00	-0.99	7.50	2.35	2.38
TDO2	-1.00	-1.00	8.92	4.86	1.87
PDGFC	-1.00	-1.00	7.62	3.27	3.12
PITX2	-1.00	-0.98	11.57	4.54	3.11
F3	-1.00	-0.99	7.14	3.96	2.14
COL15A1	-1.00	-1.00	10.76	9.60	1.29
ALDH4A1	-1.00	-1.00	7.55	1.17	2.87
SEMA3C	-1.00	-0.99	7.56	2.41	2.10
PCDH7	-1.00	-1.00	6.06	2.39	1.86
CDX2	-1.00	-1.00	7.90	3.48	2.92
COL1A2	-1.00	-0.98	11.01	6.08	1.53
AASS	-1.00	-1.00	8.45	1.22	3.42
ADD2	-1.00	-1.00	6.63	1.92	2.37
SLCO2B1	-1.00	-0.99	11.48	5.19	1.20
TSC22D3	-1.00	-1.00	8.46	3.05	1.35
GPR124	-1.00	-0.98	6.75	5.64	2.43
VEGFC	-1.00	-1.00	8.45	4.85	2.02
PLAT	-1.00	-1.00	7.55	6.14	3.22
SLIT2	-1.00	-0.98	9.46	3.44	2.84
SCRG1	-1.00	-1.00	11.93	5.44	1.03
KLKB1	-1.00	-0.99	9.58	3.62	3.76
CDH11	-0.85	-0.91	10.09	5.18	2.02
MT1X	-1.00	-1.00	14.69	3.97	5.36
MT1M	-1.00	-0.99	10.16	4.70	4.71
MT1E	-1.00	-0.99	11.41	3.52	4.58
MT1G	-1.00	-1.00	12.88	3.05	4.88
MT1F	-1.00	-1.00	11.62	3.42	4.52
MT1H	-1.00	-1.00	9.08	2.50	4.22
CRISPLD2	-1.00	-1.00	9.64	5.25	4.16
CDH13	-1.00	-0.99	5.86	5.25	0.45
PDGFD	-1.00	-1.00	5.64	2.57	1.78
THY1	-1.00	-1.00	5.90	2.94	1.97
CADM1	-1.00	-0.99	5.67	2.32	2.47

APOA4	-1.00	-0.99	19.40	7.49	2.46
MNS1	-1.00	-0.98	7.06	2.40	0.84
DNAJA4	-1.00	-1.00	6.34	0.94	2.38
TSPAN3	-1.00	-1.00	6.06	1.54	2.21
LOXL1	-1.00	-0.97	6.86	3.66	1.71
PLOD2	-1.00	-1.00	6.69	4.38	2.08
IGDCC3	-0.92	-0.87	8.97	3.94	3.50
SLCO2A1	-1.00	-1.00	6.53	4.09	2.78
SBSPON	-1.00	-0.92	6.91	5.48	1.19
RASGRP2	-1.00	-0.99	6.15	0.68	2.03
CST1	-1.00	-1.00	7.39	8.05	1.13
PRSS35	-0.99	-0.97	11.30	4.16	2.40
UCP2	-1.00	-1.00	8.22	1.93	2.59
DIO1	-1.00	-0.97	14.55	6.13	0.68
PARM1	-1.00	-1.00	12.20	4.26	2.81
KANK4	-1.00	-1.00	9.19	4.58	1.77
FOXA1	-1.00	-0.99	8.92	4.63	1.35
TGFBI	-1.00	-0.99	13.99	6.89	4.48
SLC51A	-1.00	-1.00	11.18	4.11	2.08
PRTG	-1.00	-0.99	7.03	2.90	3.57
SOSTDC1	-1.00	-0.97	7.06	2.51	2.80
ALX1	-1.00	-1.00	8.59	1.80	2.73
NODAL	-1.00	-1.00	5.88	4.00	1.31
HAND1	-1.00	-1.00	14.10	7.00	4.13
CDKN1C	-1.00	-0.96	11.07	3.93	1.35
CPE	-1.00	-1.00	6.76	2.92	1.53
SPON1	-1.00	-0.95	7.06	3.71	1.89
SLC19A3	-1.00	-1.00	5.85	2.92	0.62
RGS4	-0.82	-0.90	6.58	4.28	1.60
LEFTY2	-1.00	-1.00	10.57	7.26	3.45
LEFTY1	-1.00	-1.00	12.10	6.50	3.57
ECHDC3	-1.00	-0.99	6.32	1.65	2.59
TGFB2	-1.00	-1.00	5.72	2.72	2.15
PRSS23	-1.00	-1.00	7.36	5.92	1.98
CITED2	-1.00	-1.00	5.88	1.82	2.17
GPR176	-1.00	-0.97	7.38	1.46	3.04
NPNT	-1.00	-1.00	8.88	4.84	1.82
FBXO2	-1.00	-1.00	5.90	3.11	2.37
AMOT	-1.00	-0.98	9.64	4.03	2.36
KCNJ2	-1.00	-0.99	6.85	3.72	1.30

S100A10	-1.00	-1.00	8.96	3.16	1.52
SLC35F2	-1.00	-0.99	6.09	2.16	2.70
MFAP4	-1.00	-1.00	9.76	7.62	1.67
MUM1L1	-1.00	-1.00	11.21	3.73	2.43
CXXC4	-1.00	-1.00	7.16	2.70	0.99
SHISA2	-1.00	-0.96	5.76	4.01	3.48
TMEM88	-1.00	-1.00	8.28	4.20	1.69
EPCAM	-1.00	-1.00	8.01	1.42	3.70
TSHZ1	-1.00	-1.00	5.77	3.55	2.49
OSR1	-1.00	-1.00	5.91	4.90	3.03
NDNF	-1.00	-1.00	7.76	4.96	2.27
FAM110B	-1.00	-0.96	5.89	2.86	2.02
TMEM200A	-1.00	-0.99	6.61	4.55	2.47
FLRT3	-1.00	-1.00	11.71	2.86	3.32
RSPO2	-1.00	-0.98	5.72	4.60	2.39
REEP1	-1.00	-1.00	9.03	4.04	0.85
CTHRC1	-1.00	-1.00	8.53	4.27	2.74
TMEM119	-1.00	-1.00	7.57	5.02	3.16
HBZ	-1.00	-0.99	7.10	2.84	4.27
SLIT3	-1.00	-0.99	6.50	3.14	2.46
CD248	-1.00	-0.98	7.22	6.15	2.98
MAB21L2	-1.00	-1.00	8.17	2.87	2.09
GPC6	-1.00	-1.00	6.09	3.08	1.87
DSC3	-1.00	-0.98	7.70	2.95	1.05
IFITM1	-1.00	-0.99	6.03	4.57	1.74
CDH3	-1.00	-0.99	6.32	3.15	2.19
MYL4	-1.00	-0.98	7.87	2.46	3.26
AKR1B10	-1.00	-1.00	7.16	2.77	1.39
HS6ST2	-1.00	-0.99	6.01	3.11	2.34
CYTL1	-1.00	-1.00	6.76	3.67	0.92
CDC42EP5	-1.00	-1.00	6.72	6.34	1.10
IFITM1	-1.00	-0.99	6.51	4.56	1.64
S1PR3	-0.96	-0.94	6.88	2.46	2.21
TMEM100	-1.00	-0.99	5.96	3.95	2.50
CLDN3	-1.00	-0.99	8.73	2.71	1.90
THBD	-1.00	-1.00	11.04	3.04	0.80
SERPINA3	-0.98	-0.94	5.63	3.15	2.26
SERPINA3	-0.98	-0.94	5.63	3.15	2.26
TNNC1	-1.00	-1.00	5.77	3.58	3.10
KCTD12	-0.94	-0.94	8.17	3.83	2.36

TGFBR3	-0.92	-0.89	5.03	6.45	1.87

Table 10-5: Differentially expressed miRNAs during *in vitro* differentiation

miRNA	D6	D20	FC
hsa-miR-122-5p	1.116173	11695.98	10478.64
hsa-miR-122-3p	1.345	2331.818	1733.694
hsa-miR-194-5p	34.51123	11761.56	340.8037
hsa-miR-192-5p	65.5152	11588.42	176.8814
hsa-miR-215-5p	71.58354	6714.885	93.80488
hsa-miR-126-3p	61.17562	3826.199	62.54451
hsa-miR-146b-5p	32.51346	1632.775	50.21843
hsa-miR-660-5p	48.63475	2175.16	44.7244
hsa-miR-367-3p	2648.167	156.3541	37.16921
hsa-miR-22-3p	255.7126	8521.853	33.3259
hsa-miR-99a-5p	241.8919	7184.949	29.70314
hsa-miR-145-5p	89.83713	2113.623	23.52727
hsa-miR-4277	53.41646	1162.553	21.76396
hsa-miR-362-3p	53.32692	1126.184	21.11849
hsa-miR-100-5p	285.2226	5645.148	19.79208
hsa-miR-302c-5p	1513.711	322.8364	16.51097
hsa-miR-148a-3p	756.7924	11612.42	15.34426
hsa-miR-1271-5p	93.19828	1195.178	12.82403
hsa-miR-27b-3p	622.1667	7229.333	11.61961
hsa-miR-335-5p	316.9621	3646.825	11.50555
hsa-miR-101-3p	318.1882	1137.315	11.47451
hsa-miR-424-5p	281.3298	3198.852	11.37047
hsa-miR-148b-3p	249.9862	2716.484	10.86653
hsa-miR-532-3p	117.6664	1235.765	10.50228

Table 10-6: mRNA-miRNA intensities and interactions for human differentiation

Gene	D6	D20	miRNA	Source	D6_miRNA	D20_miRNA	miRNA-mRNA Interaction
COL1A2	7.56	12.89	hsa-miR-367-3p	Mirdb	2648.17	156.35	TRUE
EPB41L3	7.69	9.44	hsa-miR-1271-5p	Mirdb	93.20	1195.18	FALSE
FOXA1	5.96	7.60	hsa-miR-4277	Mirdb	53.42	1162.55	FALSE
PKIA	7.54	7.83	hsa-miR-27b-3p	Mirdb	622.17	7229.33	FALSE
ZFPM2	5.79	6.13	hsa-miR-148b-3p	Mirdb	249.99	2716.48	FALSE

ZFPM2	5.79	6.13	hsa-miR-148a-3p	Mirdb	756.79	11612.42	FALSE
ZFPM2	5.79	6.13	hsa-miR-302c-5p	Mirdb	1513.71	322.84	FALSE
NPNT	4.63	8.93	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
NPNT	4.63	8.93	hsa-miR-362-3p	Mirdb	53.33	1126.18	FALSE
RIN2	6.28	8.44	hsa-miR-145-5p	Mirdb	89.84	2113.62	FALSE
RIN2	6.28	8.44	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
SOSTDC1	4.78	5.81	hsa-miR-367-3p	Mirdb	2648.17	156.35	FALSE
CDH11	8.31	9.94	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
CDH11	8.31	9.94	hsa-miR-4277	Mirdb	53.42	1162.55	FALSE
CDH11	8.31	9.94	hsa-miR-27b-3p	Mirdb	622.17	7229.33	FALSE
SLC1A1	5.86	7.41	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
SLC1A1	5.86	7.41	hsa-miR-1271-5p	Mirdb	93.20	1195.18	FALSE
AMOT	6.42	7.74	hsa-miR-22-3p	Mirdb	255.71	8521.85	FALSE
SYT4	6.17	7.90	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
SYT4	6.17	7.90	hsa-miR-424-5p	Mirdb	281.33	3198.85	FALSE
CDH13	5.41	6.70	hsa-miR-660-5p	Mirdb	48.63	2175.16	FALSE
ALDH4A1	6.52	6.60	hsa-miR-27b-3p	Mirdb	622.17	7229.33	FALSE
PCDH7	5.69	6.32	hsa-miR-367-3p	Mirdb	2648.17	156.35	FALSE
LCP1	4.95	9.45	hsa-miR-1271-5p	Mirdb	93.20	1195.18	FALSE
PLOD2	8.44	10.26	hsa-miR-1271-5p	Mirdb	93.20	1195.18	FALSE
HAND1	4.66	10.62	hsa-miR-335-5p	Mirdb	316.96	3646.82	FALSE
TMEM100	4.78	6.51	hsa-miR-424-5p	Mirdb	281.33	3198.85	FALSE
KIFAP3	7.79	8.80	hsa-miR-302c-5p	Mirdb	1513.71	322.84	FALSE
FAM110B	6.33	7.00	hsa-miR-367-3p	Mirdb	2648.17	156.35	FALSE
SLIT2	8.16	9.22	hsa-miR-4277	Mirdb	53.42	1162.55	FALSE
SLIT2	8.16	9.22	hsa-miR-424-5p	Mirdb	281.33	3198.85	FALSE
GPC6	7.87	8.19	hsa-miR-367-3p	Mirdb	2648.17	156.35	FALSE
COL2A1	8.00	6.37	hsa-miR-148b-3p	Mirdb	249.99	2716.48	FALSE
COL2A1	8.00	6.37	hsa-miR-148a-3p	Mirdb	756.79	11612.42	FALSE
PRTG	6.00	5.85	hsa-miR-1271-5p	Mirdb	93.20	1195.18	FALSE
FLRT3	7.33	8.82	hsa-miR-101-3p	Mirdb	318.19	1137.31	FALSE
FLRT3	7.33	8.82	hsa-miR-27b-3p	Mirdb	622.17	7229.33	FALSE
CDH10	5.75	8.09	hsa-miR-367-3p	Mirdb	2648.17	156.35	TRUE
SPARC	7.56	12.67	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
ODC1	10.68	12.96	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
CADM1	7.65	8.14	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
SOCS2	9.28	8.94	hsa-miR-194-5p	MirTarbase	34.51	11761.56	FALSE

RGS2	7.56	9.72	hsa-miR-22-3p	MirTarbase	255.71	8521.85	FALSE
SOX9	6.46	6.49	hsa-miR-101-3p	MirTarbase	318.19	1137.31	FALSE
BIRC5	8.80	7.40	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SERPINA3	4.82	7.47	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CYTL1	5.43	6.91	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CRISPLD2	5.25	9.90	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SCRG1	7.62	6.81	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
THBD	5.72	6.91	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
TNNC1	7.80	7.19	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
TGFB2	5.05	6.10	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
LOXL1	7.10	8.35	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CDH3	7.08	9.64	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
BMP2	9.49	8.33	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CITED2	8.38	9.25	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
DIO1	3.22	7.68	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
PLAT	5.93	9.47	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SOX9	6.46	6.49	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CLDN7	6.32	8.66	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CLEC11A	6.81	8.04	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SMOC2	5.25	8.42	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
COL15A1	4.84	10.21	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CTHRC1	5.70	6.25	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SOSTDC1	4.78	5.81	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CDKN1C	7.60	11.08	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CST1	7.74	7.11	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
COL3A1	6.32	13.38	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
PPP1R1A	5.89	6.45	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
BNC1	4.82	6.35	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CDH11	8.31	9.94	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
GATA5	4.08	6.39	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
SLC1A1	5.86	7.41	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
FAM198B	4.76	7.83	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CDX2	4.64	6.57	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
BIN1	6.99	8.79	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
IFITM1	12.76	8.60	hsa-miR-335-5p	MirTarbase	316.96	3646.82	TRUE
PROS1	6.40	8.42	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
TPD52L1	6.04	8.08	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CPE	9.53	9.69	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
TSC22D3	6.26	8.55	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
RCAN1	5.83	8.07	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE

RSPO2	4.40	6.01	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
LAMB1	7.60	9.88	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
KCTD12	5.52	7.99	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
WISP1	4.68	6.89	hsa-miR-335-5p	MirTarbase	316.96	3646.82	FALSE
CTHRC1	5.70	6.25	hsa-miR-148b-3p	MirTarbase	249.99	2716.48	FALSE
LOXL1	7.10	8.35	hsa-miR-148b-3p	MirTarbase	249.99	2716.48	FALSE
SNAI2	6.02	10.05	hsa-miR-148b-3p	MirTarbase	249.99	2716.48	FALSE
AKR1B10	5.65	6.25	hsa-miR-145-5p	MirTarbase	89.84	2113.62	FALSE
CLEC11A	6.81	8.04	hsa-miR-122-5p	MirTarbase	1.12	11695.98	FALSE
NODAL	8.88	5.13	hsa-miR-122-5p	MirTarbase	1.12	11695.98	TRUE
THBD	5.72	6.91	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
MT1F	8.95	7.18	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
KCTD12	5.52	7.99	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
ASRGL1	6.74	8.33	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
REEP1	5.39	8.15	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
PRSS23	5.72	8.64	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
FOXA1	5.96	7.60	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
PARM1	5.53	8.20	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
MNS1	6.40	5.94	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
TMEM200A	6.17	7.65	hsa-miR-215-5p	MirTarbase	71.58	6714.89	FALSE
TMEM200A	6.17	7.65	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
THBD	5.72	6.91	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
REEP1	5.39	8.15	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
PCDH7	5.69	6.32	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
ASRGL1	6.74	8.33	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
NPNT	4.63	8.93	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
MT1X	10.65	8.68	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
MNS1	6.40	5.94	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
FOXA1	5.96	7.60	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
MT1F	8.95	7.18	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
PARM1	5.53	8.20	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
PRSS23	5.72	8.64	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
KCTD12	5.52	7.99	hsa-miR-192-5p	MirTarbase	65.52	11588.42	FALSE
BIRC5	8.80	7.40	hsa-miR-101-3p	MirTarbase	318.19	1137.31	FALSE
E2F2	8.09	6.32	hsa-miR-22-3p	MirTarbase	255.71	8521.85	FALSE
S100A10	10.79	12.24	hsa-miR-100-5p	MirTarbase	285.22	5645.15	FALSE
ODC1	10.68	12.96	hsa-miR-100-5p	MirTarbase	285.22	5645.15	FALSE
KCTD12	5.52	7.99	hsa-miR-22-3p	MirTarbase	255.71	8521.85	FALSE