Methylome and Transcriptome Profiling of Hepatocytes Derived from Human Pluripotent Stem Cells

Nidal Ghosheh

Department of Clinical Chemistry and Transfusion Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2018

Methylome and Transcriptome Profiling of Hepatocytes Derived from Human Pluripotent Stem Cells

nidal.ghosheh@his.se

ISBN 978-91-629-0418-0 (PRINT)
ISBN 978-91-629-0419-7 (PDF)

Printed in Gothenburg, Sweden 2018
Printed by BrandFactory Gothenburg

To the memory of my beloved grandmother

Methylome and Transcriptome Profiling of Hepatocytes Derived from Human Pluripotent Stem Cells

Nidal Ghosheh

Department of Clinical Chemistry and Transfusion Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Six hundred million people suffering from liver diseases worldwide of which the lethality is two million. Freshly isolated hepatocytes from the liver have been used for transplantation purposes and are extensively used to recapitulate drug metabolism. However, they lack stem cell ability and therefore cannot multiply, and will vary depending on each donator. Toward this, hepatocytes derived from human pluripotent stem cells (hPSC-HEP) recapitulate many features of their *in vivo* counterparts. However, the establishment of fully functional mature hepatocytes *in vitro* is still lacking. Abnormal DNA methylation emerging in *in vitro* cultured cells may underlie the immature functionality of hPSC-HEP and might explain the observed transcriptional differences between the *in vitro* generated hepatocytes and their *in vivo* counterparts. The aim of the thesis was to investigate the transcriptome and methylome of hPSC-HEP to identify their similarities and differences with human adult liver tissues.

Interestingly, on the transcriptome level, the results revealed stronger correlation and higher similarity of hPSC-HEP to adult liver than to fetal liver. Moreover, genes important for the functionality of hepatocytes with deviating expression and DNA methylation patterns, including a protein module consisting of seven drug-metabolizing enzymes that were downregulated in hPSC-HEP compared to adult liver, were identified.

In conclusion, the thesis shed light on significant deviations in the transcription and methylation of genes that are critical for the hepatic functionality. Further in-depth investigation and manipulation of these genes and their regulators in the differentiation protocol will pave the way for the generation of more functional hepatocytes *in vitro*.

Keywords: human pluripotent stem cells, gene transcription, gene regulation, DNA methylation, transcriptome, hepatocytes

ISBN 978-91-629-0418-0 (PRINT)
ISBN 978-91-629-0419-7 (PDF)

SAMMANFATTNING PÅ SVENSKA

Sexhundra miljoner människor över hela världen drabbas av kroniska leversjukdomar vilka leder till cirka 2 miljoner dödsfall per år. För närvarande är levertransplantation den enda effektiva behandlingen för slutstadiet av leversjukdom, men det råder stor brist på organ som uppfyller kraven för transplantation. Alternativa behandlingar behöver därför utvecklas för att svara upp mot det otillfredsställda behovet av levertransplantationer.

Pluripotenta stamceller utgör en population celler med unika egenskaper. De kan utvecklas till vilken celltyp som helst i kroppen och har en obegränsad förmåga att dela sig. Stamceller har därför stor potential för en rad olika tillämpningar inom bl.a. regenerativ medicin, läkemedelsutveckling och cellbehandling. Stamcellsderiverade leverceller (hPSC-HEP), har många gemensamma egenskaper med leverceller i kroppen men de saknar också viktig funktionalitet. Studier har visat att celler som odlas i laboratorium kan utveckla ett avvikande metyleringsmönster. Metyleringen kan påverka genuttrycket antingen genom att minska eller öka uttrycket av olika gener. Detta kan vara en av orsakerna till den bristande funktionaliteten hos stamcellsderiverade leverceller.

Denna avhandling omfattar både genuttrycks- och DNA metyleringsanalys av hPSC-HEP och dessa celler har jämförts med human adult levervävnad. Resultaten visar på substantiella likheter och skillnader. Syftet med avhandlingen var att studera genuttryck och DNA metyleringen av hela genomet i hPSC-HEP för att identifiera likheter och skillnader mellan human levervävnad från vuxna individer.

Resultaten visar att hPSC-HEP på genuttrycksnivå har en starkare korrelation och är mer lik adult lever än fosterlever. Dessutom identifierades avvikande genuttryck och DNA-metyleringsmönster i en grupp gener som ansvarar för viktiga metaboliska funktioner i levern. I denna grupp ingår gener som kodar för proteiner i en proteinmodul som består av sju läkemedelsmetaboliserande enzymer, och dessa uppvisade nedreglering i hPSC-HEP jämfört med i adult lever.

Sammanfattningsvis så påvisas betydande avvikelser på både genuttrycks- och metyleringsnivå hos gener som är kritiska för viktiga leverfunktioner. Fördjupade studier och manipulering av dessa gener och deras regulatorer i differentieringsprotokollen kan möjliggöra utveckling av stamcellsderiverade

leverceller med ökad funktionalitet som bättre lämpar sig för behandlingen av leversjukdomar.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Ghosheh N, Olsson B, Edsbagge J, Küppers-Munther B, Van Giezen M, Asplund A, Andersson TB, Björquist P, Carén H, Simonsson S, Sartipy P, and Synnergren J. Highly synchronized expression of lineage-specific genes during in vitro hepatic differentiation of human pluripotent stem cell lines. Stem Cells Int. 2016:8648356.

II. Ghosheh N, Küppers-Munther B, Asplund A, Edsbagge J, Ulfenborg B, Andersson TB, Björquist P, Andersson C. X, Carén H, Simonsson S, Sartipy P, and Synnergren, J. Comparative transcriptomics of hepatic differentiation of human pluripotent stem cells and adult human liver tissue. Physiol Genomics 2017. doi:10.1152/physiolgenomics.00007.2017

III. Ghosheh N, Küppers-Munther B, Asplund A, Andersson C. X, Björquist P, Andersson TB, Carén H, Simonsson S, Sartipy P, and Synnergren J. Novel transcriptomics targets for functional improvement of hepatic differentiation of human pluripotent stem cells. (manuscript).

IV. Ghosheh N, Ulfenborg B, Küppers-Munther B, Asplund A, Andersson C. X, Andersson TB, Björquist P, Carén H, Simonsson S, Sartipy P, and Synnergren, J. Identification of hypermethylated genes involved in hepatic functionality in human pluripotent stem cell-derived hepatocytes. (manuscript).

CONTENTS

ABBREVIATIONS

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ADME	Absorption, Distribution, Metabolism, Excretion
ADR	Adverse drug reaction
AL	Adult liver tissues
BAL	Bioartificial liver device
BER	Base excision repair
BMIQ	Beta Mixture Quantile dilation
CpG	Cytosine phosphate guanine
CGI	CpG island
DE	Definitive endoderm
DILI	Drug-induced liver injury
DMP	Differentially methylated probes
DMR	Differentially methylated regions
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT3l	DNA methyltransferase 3 like

ECM	Extracellular matrix
ESC	Embryonic stem cells
ESLD	End-stage liver disease
FDR	False discovery rate
FL	Fetal liver tissues
GO	Gene Ontology
ICM	Inner cell mass
H3K4	Histone 3 lysine 4
HESC	Human embryonic stem cells
HiHEP	Human-induced hepatocytes
HiPSC	Human-induced pluripotent stem cells
HPSC	Human pluripotent stem cells
HPSC-HEP	HPSC-derived hepatocytes
HSC	Hepatic stellate cells
iPSC	Induced pluripotent stem cells
lncRNA	Long non-coding RNA
LSEC	Liver sinusoidal endothelial cells
MDR	Methylation-determining regions
miRNA	Micro RNA
ncRNA	Non-coding RNA
NPC	Neural progenitor/stem cells
QN	Quantile normalization

SAM	Significance analysis of microarrays
TET	Ten-eleven translocation
TSS	Transcription start sites
XME	Xenobiotics-metabolizing enzymes

1 INTRODUCTION

The liver is the central organ of homeostasis; therefore, liver diseases are a major source of global health burden (17, 89, 90). Over 600 million people worldwide suffer from chronic liver diseases (114, 149), causing the death of about 2 million patients each year (17). Currently, liver transplantation is the only proven treatment for end-stage liver disease (ESLD). However, shortage of transplantable donor livers requires a search for alternative treatments (114, 148, 153).

1.1 THE LIVER

The liver is the major organ of homeostasis and conducts a wide range of metabolic and detoxification functions (89, 114). In addition, it has an unique capacity to regenerate itself after injury (89). Moreover, it is the largest gland in the body, since it exhibits both endocrine secretion of hormones and exocrine function, including the production of bile (121). The liver has more than 500 functions (114), and liver diseases result often in serious morbidity and mortality (89).

1.1.1 FUNCTIONS OF THE LIVER

The liver performs multiple critical functions. It regulates the glucose concentration in the blood by removing excess glucose and transforming it into glycogen and further to triglycerides. The liver synthesizes urea for the detoxification of nitrogen formed by different processes in the liver and by the production of ammonia. The liver is also the major organ of biotransformation, in which xenobiotics are detoxified by oxidation, reduction or hydrolysis reactions combined with conjugations performed by drug-metabolizing enzymes (63). Furthermore, the liver synthesizes and extracts bile acids and bilirubin from the blood, excreting them into the bile (63, 102). The liver synthesizes most of the plasma proteins, such as albumin and apolipoproteins (63, 121), and it is the major site for degradation of plasma proteins (33). In addition, the liver regulates the synthesis and the transport of cholesterol (121). It plays a major role in immunotolerance, leading to the suppression of immune responses against vital exogenous molecules (59, 75). Normally, the immune response in the liver is skewed toward immune tolerance rather than immune activation (75). However, this property of the liver could be exploited by pathogens causing liver diseases (23).

1.1.2 LIVER FAILURE

Despite the ability of the liver to regenerate, under certain conditions it loses this ability and fails to function, resulting in an accumulation of toxins which results in coma and death if not treated (130). The tolerogenic property of the liver causes harm in liver diseases such as cancer and hepatitis because antigens are continuously expressed during the pathology of the disease, resulting in systematic immune tolerance and the body's difficulty controlling these diseases (75).

Liver failure can develop by versatile diseases and injuries including liver chronic diseases which can be triggered by viral infections, drugs, alcohol abuse, and non-alcoholic steatohepatitis, as well as from autoimmune, cholestasis and inherited metabolic diseases (8, 30). However, hepatitis B and C constitute most of the chronic liver diseases. Infections by hepatitis viruses result in immune-mediated hepatocyte death (135). Most chronic liver diseases lead to serious fibrosis (8), which may progress into liver cirrhosis and cancer (135).

Liver fibrosis is the accumulation of extracellular matrix (ECM) components in the liver, forming fibrous scars as a response to chronic activation of inflammatory signals during liver injury (8). This process is considered a wound-healing model of chronic liver disease mediated by hepatic stellate cells (HSC) (8, 135). HSC are activated by stress signals, such as reactive oxygen species released by dying hepatocytes (135). The accumulation of excess fibrous scars in the liver parenchyma leads to liver cirrhosis which is associated with the formation of nodules of regenerated hepatocytes (135). Liver cirrhosis leads to liver failure and hampering of regenerative pathways (135). In addition, patients with cirrhosis are prone to develop cancer if not treated (6). The most common type of liver cancer is hepatocellular carcinoma and it may be initiated by chronic inflammatory and cirrhotic microenvironment (135, 153).

Acute liver injury is also a type of liver failure which may lead to hepatic encephalopathy and death if not treated. Hyperacute liver injury occurs mostly in the context of acetaminophen (paracetamol) toxicity or viral infections, while subacute cases are caused by unpredictable drug-induced liver injury (DILI). DILI is a type of adverse drug reaction (ADR) caused by the idiosyncratic induction of inappropriate immune response by drugs or metabolites. Such idiosyncratic reaction is attributed to the polymodality of Phase I and II drug metabolizing enzymes that results in either extensive or poor metabolic activity (27). Patients with rapid onset of acute liver injury

have a better prognosis when treated by medications alone than those with slower onset of the disease (9).

1.1.3 AVAILABLE TREATMENT FOR LIVER DISEASE

LIVER TRANSPLANTATION

Liver transplantation is the optimal treatment and the most successful for ESLD, as it is an irreversible condition (153). However, due to shortage of donor livers, 40% of the patients on the waiting lists do not receive liver transplantation, resulting in further progression of the disease and death (148). In addition to the shortage of donor livers, the invasive property of the transplantation procedure and the immunological incompatibilities between donor and recipient demand the consideration of alternative treatments (9, 25, 130, 153).

HEPATOCYTE TRANSPLANTATION

Hepatocyte transplantation is less invasive than organ transplantation and has been applied to restore liver function or to bridge patients to liver transplantation (9, 25, 130). Nevertheless, this treatment is also limited due to the availability of untransplantable donor livers and the quality of the isolated hepatocytes. In addition, loss of grafts has been observed a few months after transplantation (9, 25, 130). Moreover, hepatocytes typically lose their function and ability to proliferate upon isolation, which is another limitation of this treatment (6). However, patients with acute liver injury treated with hepatocyte transplantation showed improvement in the reduction of ammonia, bilirubin and hepatic encephalopathy, but with no survival benefit (25). Nevertheless, successful recovery of patients, especially neonates and children, with monogenic inherited metabolic disorders, has been reported (9, 25, 130). In some cases, liver functionality was improved by the replacement of 2 - 5 % of liver parenchyma with normal hepatocytes (114).

To avoid the immune rejection of transplanted hepatocytes, encapsulation in immunoisolated microbeads has shown to elongate the lifespan of transplanted hepatocytes in the host. However, transplantation of co-cultured hepatocytes with mesenchymal cells, which not only inhibit the death of hepatocytes in damaged liver but also stimulate their proliferation, is being considered (25).

EXTRACORPOREAL LIVER-ASSIST DEVICES

Extracorporeal liver-assist devices include non-biological dialysis, which removes toxins from the circulating system, and the bioartificial liver device (BAL), which removes xenobiotics by filtration or adsorption while the

included hepatocytes in the device compensate for biotransformation and synthetic functions of the liver. Primary human hepatocytes, porcine hepatocytes, immortalized hepatic cell lines, and hepatocyte-like cells (HLC) are used in BAL. Extracorporeal liver-assist devices are regarded as supportive therapy that bridges patients to transplantation or facilitates liver regeneration (9).

1.2 THE STRUCTURE OF THE LIVER

Structurally, the liver is composed of parenchymal (hepatocytes) and non-parenchymal cells (including bile ductile, and connective tissue cells) (63, 102). The liver obtains blood from both the hepatic artery and the portal vein that flows through the sinusoidal capillaries to the central vein (59, 121). The portal vein contains nutrients, metabolites, toxins, and antigens derived from the gut and transported through the blood to the liver (59). The lobule is the structural unit of the liver, and it includes the periportal zone where the portal vein, the hepatic artery, and the bile duct are located (89, 130). The blood in this zone is enriched with oxygen, hormones, and substrates (63). In contrast, the pericentral zone, where the central vein is located (130), has a low concentration of oxygen, hormone, and substrates. However, it is rich with CO_2 and other products (63). The midlobular zone between the periportal and the pericentral vein contains gradients of oxygen and different products (130). Both the hepatocytes and other non-parenchymal cells residing in these different zones have distinct gene expression patterns and functional responsibilities (64, 130). The space of Disse separates the hepatocyte from the sinusoids. The sinusoids consist of Kupffer and fenestrated endothelial cells (151). Many kinds of liver cells contribute to antigen presentation, a process that allows the immune system to recognize antigens as non-self (141). Cells that constitute the key component of the innate immune system are enriched in the liver, including the Kupffer cells that reside in the liver, natural killer cells, and natural killer T cells, which are recruited from the blood stream (107). Figure 1 shows an overview of the liver lobule and illustrates the zonation and the localization of the different hepatic cell types.

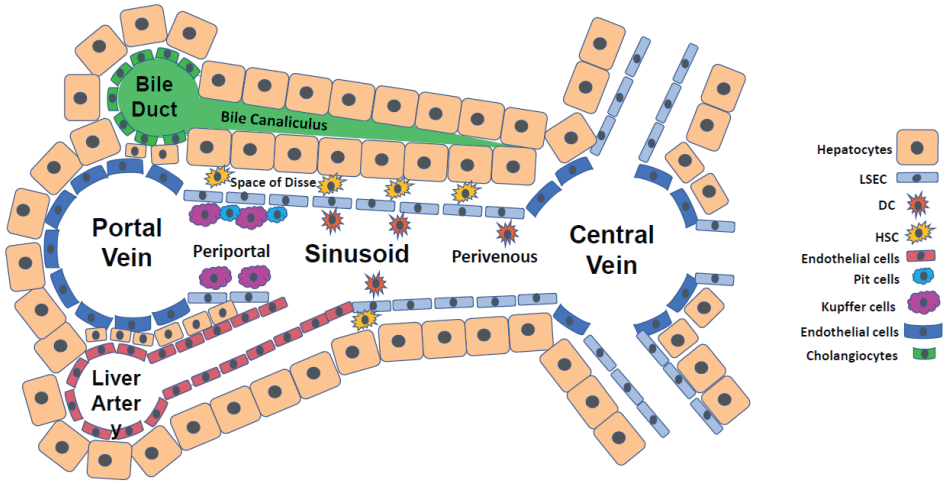

Figure 1. The structure of the liver lobule, and the localization of hepatic cells. The perivenous zone is located closer to the central vein, while the periportal zone is located closer to the portal vein. Abbreviations: Dendritic cells (DC), hepatic stellate cells (HSC), liver sinusoid endothelial cells (LSEC).

1.3 NON-PARENCHYMAL CELLS

The non-parenchymal cells are responsible for specific physiological liver functions. They play a role in liver damage caused by DILI, acute inflammation, hepatitis, and chronic liver diseases such as fibrosis and cirrhosis (102). The majority of non-parenchymal cells is represented by biliary epithelial cells (cholangiocytes), sinusoidal endothelial cells, hepatic stellate cells (Ito cells), Kupffer cells, and pit cells (151).

1.3.1 KUPFFER CELLS

Kupffer cells are residential microphages in the liver, and they constitute 20% of the non-parenchymal cells there. They are located in the sinusoidal space, predominantly in the periportal area. Their main function is to clear toxins, debris, and microorganisms through phagocytosis from the blood stream (75, 107, 151). Kupffer cells produce pro- and anti-inflammatory cytokines to control the immune response, and they have a central role in inducing tolerance and liver regeneration (102). In addition, they act as antigen-presenting cells and serve as educators of circulating cells such as T cells, natural killer cells, natural killer T cells, and myeloid-derived suppressor cells (75). However, they can also produce reactive oxygen intermediates, damaging the parenchymal

and non-parenchymal cells during immune response, and thus contributing to the pathogenesis of liver diseases (69, 102).

1.3.2 LIVER SINUSOIDAL ENDOTHELIAL CELLS

Liver sinusoidal endothelial cells (LSEC) are of mesenchymal origin (102), and they constitute 50% of the non-parenchymal cells in the liver. LSEC form a layer of fenestrated thin vessels called the sinusoid, which allow blood to stream through the liver lobule (107). They separate the hepatocytes from the bloodstream while still allowing exchange of substances between hepatocytes and the bloodstream through the fenestrations (102, 135). Moreover, they express molecules that promote the uptake of antigens (75, 107) and function as antigen-presenting cells and educators of circulating cells (75).

1.3.3 HEPATIC STELLATE CELLS

Hepatic stellate cells (HSC) are non-parenchymal cells originating from the mesenchyme (75, 102) and constitute 5-8 % of total liver cells (141). They reside in the space of Disse between the sinusoid and the hepatocytes (151). Their main function is the storage and metabolism of retinol (75, 102), although they also synthesize ECM components and regulate the homoeostasis in the microenvironment in the liver (141). They also act as antigen-presenting cells, which capture and process antigens in the liver (75, 107, 141). In addition, they educate circulating cells (75) and regulate the inflammatory and immunological processes (141). HSC are normally quiescent and only activated upon liver injury. When activated, they transdifferentiate into myofibroblasts and start to produce ECM components, which have a crucial role in the development of fibrosis and cirrhosis (102, 135, 141). The deactivation of hepatic stellate cells is important for the resolution of fibrosis (141).

1.3.4 LIVER DENDRITIC CELLS

Liver dendritic cells are professional antigen-presenting cells, and they are involved in educating circulating cells and liver tolerance. Immature hepatic dendritic cells are differentiated into tolerance- or effective immunity-inducing cells upon uptake of antigens (23).

1.3.5 BILIARY EPITHELIAL CELLS

Biliary epithelial cells, also known as cholangiocytes, form the bile duct (89). They are also considered to be facultative stem cells that acquire stemness only in certain circumstances to maintain organ homeostasis. In case of severe injury, they contribute to the generation of hepatocytes (108, 130).

1.3.6 OVAL CELLS

The atypical ductal cells, sometimes known as oval cells, are also considered to be facultative stem cells and hepatic progenitor cells (130, 145). These cells express markers for both cholangiocytes and hepatocytes (6). Oval cells are thought to emerge from cholangiocytes upon toxin-mediated liver injury and differentiate into cholangiocytes and hepatocytes (130, 145). Importantly, the total inactivation of the regenerative ability of hepatocytes is crucial for the oval cells to become activated and differentiated to regenerate the damaged liver (1).

1.3.7 PIT CELLS

Pit cells are liver-associated natural killer cells. They are hepatic large granular lymphocytes located in the sinusoid lumen and weakly attached to the sinusoid wall, adhering to Kupffer cells and LSEC. These cells are originated from natural killer cells from the peripheral blood that differentiate into pit cells after residing in the liver and adhering to the liver sinusoid. Pit cells have been reported to target and kill tumor cells and suppress metastasis (95).

1.4 PARENCHYMAL CELLS

1.4.1 HEPATOCYTES

The hepatocytes constitute the parenchyma of the liver and responsible for most of its functions (33, 75). They compose 60% of the liver cells and 80% of its volume (33). In addition, they are highly polarized and heterogeneous regarding the uptake, release and metabolism of compounds (33, 89).

THE FUNCTIONS OF HEPATOCYTES

The hepatocytes are responsible for most of the liver functions including biotransformation and detoxification of xenobiotics, the synthesis and secretion of bile, energy metabolism, and glycogen storage (89, 102). In addition, the hepatocytes serve as antigen-presenting cells and induce immune tolerance due to constant contact with gut antigens and neoantigens because of their metabolic function. Therefore, they contribute to preventing autoimmune reactions (75). Morphologically, the hepatocytes are polygonal with large nuclei, enriched with mitochondria, lipid bodies, peroxisomes and microvilli vesicles. In addition, they exhibit intact Golgi apparatus, rough endoplasmic reticulum, and junctional complexes (114). The hepatocytes are highly polarized. Their basolateral membrane borders the fenestrated sinusoidal endothelial cells where the exchange of the different substances occurs, while

the apical membrane of hepatocytes borders the bile canaliculi, which is formed by tight junctions between adjacent hepatocytes. The hepatocytes secrete bile acids and salts to the bile canaliculi to be transported to the bile duct (121).

The zonation of the liver causes heterogeneity of hepatocytes. Herewith, periportal and perivenous hepatocytes have distinct areas of responsibility. For instance, the storage of glycogen in hepatocytes is achieved by uptake of glucose and glycolysis in perivenous hepatocytes, but by gluconeogenesis and the release of glucose in periportal hepatocytes (33, 63). The catabolism of amino acids and fatty acids requires a high concentration of oxygen. Therefore, this process is performed by periportal hepatocytes. Ureagenesis is performed by both periportal and perivenous hepatocytes. However, periportal hepatocytes synthesize urea from amino acids, while perivenous hepatocytes use ammonia. Biotransformation is the process of detoxification of xenobiotics by oxidation, reduction or hydrolysis followed by conjugation. Biotransformation, such as drug metabolism and detoxification, occurs in perivenous hepatocytes, mostly by monooxygenases such as Cytochrome P450 enzymes of the smooth endoplasmic reticulum, which is predominant in these cells. The formation of toxic metabolites is high in perivenous hepatocytes; however, oxidation protection against these metabolites by conjugation to substrates resulting in harmless products is higher in periportal hepatocytes. In addition, the UDP-glucuronosyltransferase activity is predominant in perivenous, while sulfotransferase activity is dominant in periportal hepatocytes (33).

Secretion of bile acids and bilirubin occurs normally in periportal hepatocytes. This process could be performed in perivenous hepatocytes upon injury of the periportal zone. However, the synthesis of bile acids occurs mainly in perivenous hepatocytes (33, 63). Moreover, hepatocytes synthesize most of the plasma proteins. Due to hepatocyte heterogeneity, different acute phase proteins are preferentially synthesized in different zones of the liver, but upon injury this preference is abolished (63).

1.5 REGENERATION OF THE LIVER

The liver is an extraordinary regenerative organ in the body, as it has the capacity to regenerate after a two-thirds physical resection (130). In healthy liver, hepatocytes and cholangiocytes have unlimited proliferation potential, which accounts for liver turnover and regeneration (67, 145). Regeneration is initiated by Wnt signals that promote the proliferation of perivenous hepatocytes. In mildly diseased liver and hepatectomy, perivenous hepatocytes

are not capable of replacing damaged hepatocytes, and in this case the periportal hepatocytes proliferate and regenerate the damaged liver (1, 67, 130). Wnt signaling, in addition to cytokines and other signaling pathways, regulates the proliferation process of hepatocytes, and replication of hepatocytes is the dominant regeneration mechanism in healthy and mild diseased liver (1, 145).

In severely damaged liver, the proliferative potential of the hepatocytes and the cholangiocytes is hampered as a response to liver inflammation. Therefore, these cells lose their ability to regenerate the liver (1, 67). Upon toxin-mediated liver injury, hepatic progenitor cells are activated by HSC. They then proliferate and give rise to oval cells that have the potential to differentiate into hepatocytes and cholangiocytes to regenerate the liver (1, 67, 145), while HSC contribute to ECM remodeling during liver regeneration (135).

1.6 HEPATOGENESIS

Liver development is currently mostly understood in mouse embryos due to the availability of the mouse model system. However, studies in other animal models such as the chicken, zebrafish and Xenopus, in addition to studies in primary cell cultures, have revealed that most of the hepatogenesis is evolutionarily conserved. The liver is comprised of different cell types of which the hepatocytes are the principal cell type. Hepatocytes and cholangiocytes are the only cell types in the liver that are derived from the endoderm germ layer, while the remaining cell types are derived from the mesoderm germ layer (154).

During embryogenesis, the definitive endoderm (DE) emerges from the primitive streak at the gastrulation stage (151). Subsequently, the primitive gut, which is subdivided into foregut, midgut and hindgut regions, is formed from the DE. Hepatocytes are generated from the foregut endoderm, where the hepatic endoderm (hepatoblasts) originates from the ventral foregut (154). In humans, gastrulation and the differentiation into three germ layers occur at week three after fertilization. The endoderm is formed by high levels of Nodal signaling, which is stimulated by the canonical Wnt pathway (38), then it is patterned further into foregut, midgut and hindgut, where high Nodal levels promote the anterior endoderm generation and the expression of hHEX, SOX2, and FOXA2 (155). The foregut endoderm is subsequently generated by repression of Wnt/β-Catenin and FGF4 signals. The hepatic fate in the ventral foregut is promoted by FGF and BMP signaling from the cardiogenic mesoderm and the septum transversum mesenchyme respectively (151, 154). Hepatoblasts are generated during weeks 3 and 4 after fertilization (153), and

they express hHEX, AFP, ALB, and HNF4A. Afterwards, they migrate and invade the adjacent septum transversum mesenchyme of a mesodermal origin (72, 151, 154) to form the liver bud at week five after fertilization, which is also when hematopoiesis is initiated (39). Hence, the fetal liver becomes the major hematopoietic organ, resulting in blood cells constituting the majority of the liver cells (89). The liver bud then becomes surrounded by angioblasts and endothelial cells to form hepatic vasculature (151). Of note, the presence of FLK1 positive endothelial cells is required for the hepatocyte to be established from the liver bud (89, 151). Furthermore, mesothelial cells are required for the proliferation of hepatoblasts as they provide growth factors such as HGF, midkine, and pleiotrophin (89). However, hematopoiesis in human fetal liver ends towards week 26 after fertilization (39). The formation of the liver bud requires the expression of hHEX, GATA4/6, HNF6, OC2, TBX3 and PROX1 (72, 154). Notably, the proliferation of hepatoblasts is controlled by the surrounding endothelial cells. The hepatocyte differentiation starts between week six and eight after fertilization (39, 153), requiring low levels of TGFβ and Wnt/ β-catenin signaling (154). Hematopoietic cells contribute to the maturation of hepatocytes by the secretion of cytokines such as oncostatin-M (OSM) and interleukin 6 (89). On the other hand, high levels of TGFβ and Wnt/β-catenin direct differentiation of the hepatoblasts towards cholangiocytes (151, 154). Here, HNF6 controls the timing and the positioning of cholangiocytes, as the lack of HNF6 in the embryo allows the early differentiation of hepatoblasts to cholangiocytes and their extension from portal mesenchyme to the liver parenchyma (151). Finally, the hepatocytes undergo a long process of maturation that continues until after birth. A network of transcription factors, HNF1A, HNF1B, FOXA2, HNF4A, HNF6, and LRH1, in addition to OSM, WNT, HGF, and glucocorticoids, controls the maturation process (38, 74).

1.7 HEPATOCYTES IN PHARMACOLOGY

The drug metabolism function of the liver constitutes a risk factor for the development of DILI. Therefore, different *in vitro* hepatocyte models have been developed for safety pharmacology and toxicology research in order to understand the mechanism of DILI and to screen for new chemical entities. Freshly isolated primary hepatocytes are the gold standard *in vitro* model. However, due to their limited availability, short life span, inter-donor differences, and variable viability following isolation, in addition to their rapid dedifferentiation and loss of function (65), immortalized hepatocyte cell lines such as HepG2 and HepaRG have been developed to overcome these limitations (41). However, these cell lines only partly recapitulate the

hepatocyte functions and do not represent the genetic polymorphism of the hepatocyte population (65).

ADR, including DILI, is a major problem for pharmaceutical industries and clinicians due to the polymodal property of Phase I and Phase II drug-metabolizing enzymes, Phase III transporters, and the receptors that regulate the different phases in hepatocytes. ADR is responsible for the withdrawal of 4% of the drugs that enter the market and for 50% of the drug candidates during drug development (57). This is often the cause of either extensive or poor drug metabolism (65, 128). Hence, in order to reduce late attrition of drug candidates and investigate the mechanism of DILI, improved hepatocyte models are urgently needed.

1.7.1 HEPATOCYTE MODELS

Hitherto, hepatocyte models applied in pharmaceutics have failed to recapitulate accurately the morphology, functionality and phenotype of *in vivo* hepatocytes (114). Hepatocyte models include human primary hepatocytes, liver cell lines, animal models, and hepatocyte-like cells (HLC).

HUMAN PRIMARY HEPATOCYTES

Human primary hepatocytes are considered the gold standard for *in vitro* investigation of drug discovery and development, in addition to metabolism and toxicity assessment of drugs (102, 114). However, they do not recapitulate accurate and robust hepatocyte functionality as they lack the important cell-cell interactions between the different liver cell types (102). Moreover, the availability of these cells from untransplantable donor liver and liver tissue is limited (148). In addition, they exhibit inter-individual and batch-to-batch variability with different viability levels (148). They also dedifferentiate and fail to recapitulate most normal levels of hepatocyte functions when kept in culture (114).

LIVER CELL LINES

Other hepatocyte models were developed by establishing liver cell lines from hepatocellular carcinoma and through SV40 transformation (114). These cells are easier to maintain and have a longer lifespan (153). However, they do not exhibit normal hepatic functionality. Furthermore, they dedifferentiate *in vitro* and accumulate genetic abnormalities (114). Although HepG2 is a liver cell line derived from FL, it exhibits low metabolic functionality.

HepaRG was derived from hepatocellular carcinoma and expresses CYP1A2, CYP2B6, CYP2C9, CYP2E1, and CYP3A4. Nevertheless, it also has lower

metabolic activity and lacks the ability to accurately predict drug toxicity (153).

ANIMAL MODELS

Animal models have distinct physiological and metabolic properties. Therefore, they do not accurately recapitulate the functions of human hepatocytes (114). Several studies have shown that animal models fail to predict human response to drugs, and thus they contribute to the high attrition rate in drug development, since most of the drugs at the later stages in drug development are optimized in animal models (117).

HEPATOCYTE-LIKE CELLS

HLC are cells that exhibit some properties of true hepatocytes. These cells are generated from extra-hepatic cells, including human pluripotent stem cells (hPSC), most efficiently by direct differentiation mimicking embryonic hepatocyte development to produce hPSC-derived hepatocytes (hPSC-HEP). These cells do exhibit some immature aspects, as they fail to turn off some genes of earlier stages during the differentiation (32).

HLC could also be generated from human fibroblasts or other cell types by transdifferentiation, producing human-induced hepatocytes (hiHEP) through the introduction of a combination of hepatic transcription factors such as FOXA2, HNF4A, and CEBPB, or HNF1A, HNF4A, and FOXA3 (32, 68). Of note, functional differences between HLC derived by differentiation of hPSC and those derived by transdifferentiation were observed at the transcriptional level. Importantly, hPSC-derived HLC expressed endoderm progenitor and hepatoblast markers as well as *CDX2,* a colon-specific transcription factor, while hiHEP did not. Nevertheless, both were observed to abolish tissue-specific genes of the somatic cells from which they were derived. Interestingly, some HLC were reported to gain improved cell phenotype and exhibited regeneration ability of damaged liver upon transplantation (3, 32).

1.8 ADME GENES

ADME genes are crucial for the absorption, distribution, metabolism and excretion (ADME) of drugs and xenobiotics. The ADME proteins include Phase I and II xenobiotic metabolizing enzymes (XME) and Phase III transporters, in addition to receptors that regulate the members of Phases I, II and III (143). These genes show high inter-individual variability, accounting for ADR as a result of different responses to drugs (57).

1.8.1 PHASE I XME

Phase I XME convert hydrophobic xenobiotics into hydrophilic molecules. These enzymes include cytochrome P450 (CYP), flavin-containing monooxygenase (FMO), carboxylesterase (CES), alcohol and aldehyde dehydrogenases (ADH, ALDH), aldo keto reductase (AKR), and amine oxidases (73, 92). In FL, Phase I drug metabolism is immature and includes mainly CYP enzymes, while other XME are either expressed at very low levels or absent. However, FMO1, CYP1B1, and CYP3A7 are upregulated in FL. In AL, CYP3A7 is replaced by CYP3A4 (92).

1.8.2 PHASE II XME

Phase II XME convert the xenobiotics into water-soluble molecules by the addition of endogenous compounds. Phase II XME include glutathione S-transferase (GST), sulfotransferase (SULT), UDP-glucuronosyltransferase (UGT), and N-acetyltransferase (NAT) (73, 92). Most Phase II XME are not expressed in FL, although a few, such as SULT1A3, are only expressed in FL (92).

1.8.3 PHASE III TRANSPORTERS

These transporters regulate the uptake and efflux of substances to and from the liver. Uptake transporters include organic anion transporters (OATS) and organic anion transporting polypeptides (OATPS). The efflux transporters consist of multidrug resistance (MDR) transporters and resistance-associated proteins (MRPS, ABCC). Bile acid transporters such as BSEP (uptake) and NTCP (efflux) are downregulated in FL (92).

1.8.4 XENOBIOTIC RECEPTORS

The regulation of Phase I, II, and III drug metabolism enzymes and transporters is mediated through receptors that sense the microenvironment of hepatocytes and activate the detoxification machinery accordingly. This group of receptors contains AHR, orphan nuclear receptors, pregnane X receptor (PXR), constitutive androstane receptor (CAR/NR1I3), peroxisome proliferator-activated receptors (PPAR), liver X receptor (LXR), farnesoid X receptor (FXR), and retinoid X receptor (RXR) (143).

1.9 HUMAN STEM CELLS

Human stem cells are primitive cells, characterized by their unique capacities of self-renewal and differentiation into one or several specialized cell types

(22). The different types of stem cells can be classified into two categories: pluripotent stem cells and adult stem cells (26).

1.9.1 HUMAN ADULT STEM CELLS

Human adult stem cells are quiescent tissue resident stem cells that mediate tissue homeostasis and regeneration upon receiving appropriate activation signals (91, 111). There are two types of human adult stem cells: multipotent and unipotent.

Multipotent stem cells have the potential to differentiate into multiple cell types, normally within a single lineage or germ layer. For instance, hematopoietic stem cells can differentiate into all types of blood cells (115); neuronal stem cells can differentiate into neurons, astrocytes, and oligodendrocytes (132); and mesenchymal stem cells can differentiate into specialized cells of the skeletal tissues (111).

On the other hand, unipotent stem cells can only differentiate into one cell type: satellite stem cells that differentiate into skeletal muscle cells (26).

Remarkably, some types of adult stem cells have the capacity to transdifferentiate into cell types from different germ layers upon transplantation. This can be seen with hematopoietic stem cells and mesenchymal stem cells, which have been suggested to have the potential to differentiate into hepatocytes and repair metabolic function and liver regeneration (149). Another possible explanation for this phenomenon is that the adult stem cells were fused with hepatocytes and reprogrammed by them (49). Interestingly, mesenchymal stem cells can inactivate human stellate cells, resulting in the inhibition of the fibrogenic process. Transplantation of mesenchymal stem cells in ESLD patients showed promising results in Phase I and II clinical trials. However, for patients with ESLD caused by hepatitis B, there was no long-term improvement of hepatic function (153).

1.9.2 HUMAN PLURIPOTENT STEM CELLS

Human pluripotent stem cells (hPSC), including both human embryonic stem cells (hESC) and human-induced pluripotent stem cells (hiPSC), are characterized by their unique capacities of self-renewal and differentiation into all mature cell types of the different germ layers (60). Considering the aforementioned characteristics of hPSC, they provide an excellent human cell source in basic research, regenerative medicine, and cell therapy. In addition, they could compensate for many of the drawbacks of the current methods and models used in liver disease treatments, drug discovery, and toxicology (65).

HUMAN EMBRYONIC STEM CELLS

Embryonic stem cells (ESC) were first isolated from mouse blastocysts in 1981, but it was not until 1998 that the isolation of the first human embryonic stem cells was accomplished (149). Human ESC are isolated from the inner cell mass (ICM) of preimplantation blastocysts (48, 60). The zygote, which is formed after fertilization of the egg, undergoes multiple mitotic cell divisions leading to the formation of the blastocyst, which consists of ICM and the trophoblasts. The ICM includes the epiblast, which gives rise to the embryo. The outer cell layer consists of trophoblasts, which give rise to the placenta, the chorion and the umbilical cord. The preimplantation blastocyst is encapsulated by zona pellucida, a glycoprotein protective layer (42, 140). Isolation of embryonic stem cells could be performed by different techniques including immunosurgery, spontaneous hatching of the blastocyst, and enzymatic removal of zona pellucida. Culturing of the isolated stem cells could be achieved either in the presence of feeder cells and basic fibroblast growth factor (bFGF) (48) or in a feeder-free culturing system such as in DEF-CS (www.clontech.com) (35). The quality and the pluripotency properties of the established stem cell lines are assessed by different *in vivo* and *in vitro* characterization, including the morphology of the cells or the colonies (48); the expression of ESC markers such as OCT4, NANOG, SOX2, SSEA-3, SSEA-4, and TRA-1-81(48, 152); the presence of telomerase activity and high levels of alkaline phosphatase activity; karyotype analysis; the formation of embryoid body containing cell types from all three different germ layers to confirm pluripotency *in vitro;* and the formation of teratoma from stem cells transplanted into SCID mice (48). The application of hESC in a clinical context may trigger ethical issues due to the isolation of these cells from fertilized human eggs (22).

HUMAN-INDUCED PLURIPOTENT STEM CELLS

Human iPSC are generated through reprogramming of adult somatic cells by the transduction of genes encoding transcription regulators of stem cells, such as OCT4, SOX2, LIN28, KLF4, NANOG and C-MYC, using viruses, plasmids, synthesized RNA, and proteins. Although both hESC and hiPSC have the same morphology and function, there is evidence of differences in the gene expression levels between these two cell types (51). Moreover, hiPSC can retain an epigenetic memory of the somatic cell from which they originate (66). However, hiPSCs are still preferred over hESC as these offer the potential to generate patient-specific cell types and *in vitro* models of rare diseases (114), and provide a model of higher relevance than animal disease models whose distinct physiology limits the translatability of the results (148). Furthermore, they offer the possibility of generating versatile phenotypes, which improves

drug discovery and toxicology studies (65). In addition, hiPSC have potential therapeutic applications including tissue replacement and gene therapy, thus promoting patient-specific treatment (148). Importantly, hiPSCs bypass the ethical issue of using hESC in a clinical setting. Indeed, clinical application of hiPSC-derived retinal tissues was already implemented in patients in 2014 (22).

1.9.3 DIFFERENTIATION OF HPSC INTO HEPATOCYTES

Different strategies have been established to differentiate hPSC into hepatocytes *in vitro.* One strategy is directed differentiation, done by applying Activin A and WNT3A to stimulate the hepatic differentiation of stem cells (6). There are also differentiation protocols of hPSC that include a step of formation of embryoid bodies, but these are unreproducible due to a spontaneous regional differentiation resulting in a variety of alternate cell lineages (114).

Efficient differentiation of hPSC into hepatocytes that recapitulate many features of their *in vivo* counterparts, including the expression of hepatic markers and genes involved in drug metabolism and transport, has been achieved by mimicking the embryonic hepatogenesis by encompassing the DE, the hepatoblast, and hepatocyte maturation stages (14, 46, 65, 150). Notably, the generation of pure culture of DE is essential for effective hepatic differentiation (6). Although the results from hPSC-differentiation are encouraging, establishment of fully functional hepatocytes *in vitro* is still lacking (113). In comparison to their *in vivo* counterpart, hPSC-HEP produce lower levels of albumin and exhibit lower cytochrome P450 activity. In addition, they have lower consumption of oxygen, immature mitochondria and incomplete urea activity. Furthermore, they fail to turn off early hepatic markers as, unlike *in vivo* mature hepatocytes, they express the fetal hepatic marker AFP (148).

One standardized protocol to generate homogenous hPSC-HEP cultures from a panel hPSC lines displaying metabolic diversity reminiscent of intra-individual variation present in human population was reported by Asplund and colleagues (5). Their study showed notable similarities between the large numbers of cell lines analyzed in addition to variability of hepatic enzyme activity, including CYP1A, CYP2C9, CYP2D6, and CYP3A, where CYP1A and CYP3A activity were in a range similar to that of human primary hepatocytes, while CYP2C9 and CYP2D6 showed lower activity in hPSC-HEP. In addition, these cells were proven to be useful in chronic toxicity

studies (5, 53). These findings promote the application of hPSCs-HEP in patient-specific therapeutics, drug discovery and DILI investigations. However, in order to fulfill these tasks and reach the functionality levels of freshly isolated primary hepatocytes, further improvement of the differentiation protocols is required.

Importantly, cell-cell interaction that occurs during hepatogenesis contributes to normal development of the organ. Camp *et al.* adapted 3D approaches in hepatocyte differentiation from stem cells, applying co-culturing with endothelial and mesenchymal cells. The results revealed high transcriptional correspondence between *in vitro* developed liver bud and FL (18).

Interestingly, hPSC-HEP were also found to support the life cycle of viral hepatitis C as well as to exhibit proper immune response. Therefore, these cells could be used to study the pathogenesis of the virus (114), considering the host-limited tropism of these viruses to humans and chimpanzees (148).

Human PSC-HEP are anticipated to replace the current hepatic models in pharmaceutics and to compensate for the shortage of donor livers. Herewith, hPSC-HEP transplantation may serve as an alternative treatment for chronic and acute liver failure, viral infections, and inherited metabolic disorders. Notably, transplantation of hPSC-HEP in rodents showed reduced fibrosis and enhanced liver regeneration. Moreover, they stabilized chronic liver diseases (148). The application of hiPSC to produce hPSC-HEP facilitates the establishment of cell libraries with known genotypes to match patients with HLA/MHC to avoid graft rejection by the immune system (114). In addition, they could be applied to the establishment of disease models of inherited liver disorders to investigate the biology of the disease pathology (153). However, to achieve these goals, efficient and reproducible differentiation protocols must be developed (114). For advanced application of hPSC-HEP in pharmaceutics, they should approach the activity ranges of *in vivo* hepatocytes. The hepatic functions that hPSC-HEP need to perform for this aim are: metabolism of xenobiotics and endogenous substances; synthesis and secretion of albumin, clotting factors, complement, transporter proteins, bile, lipids and lipoproteins; and storage of glycogen, fat-soluble vitamins A, D, E and K, folate, vitamin B12, copper and iron (49). Furthermore, hPSC-HEP must be completely free from all non-liver cells. In addition, the functional maturity of hPSC-HEP must be improved (130). This could be achieved by mimicking the transcription of liver development *in vitro* to provide all necessary signals for hepatocyte generation (114). However, the resolution of safety challenges of stem cell-based therapies, such as the possibility of monitoring the engrafted cells and distinguishing them from host cells, in addition to the production of pure

populations of hepatocytes to prevent the formation of tumors and avoiding immune system response leading to the rejection of the graft, must be achieved in order to permit the application of hPSC-HEP in regenerative medicine and cell-based therapies. To achieve this goal, new techniques need to be developed to allow the distinguishing of grafts from host cells after transplantation in addition to techniques that eliminate the transplanted cells upon the emerging of abnormalities. Importantly, establishment of stem cell lines and the differentiation procedure must be conducted according to good manufacturing practices. Herewith, the origin of the cells and the different steps during the differentiation procedure must be rigorously controlled and characterized to confirm the absence of genetic abnormalities (37).

1.10 EPIGENETICS

Epigenetics is defined as ‘‘the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence’’(10). Epigenetic events regulate gene expression at both transcriptional and translational levels (58) and allow a single genome to produce different types of cells (2). During the differentiation process, there is a gradual silencing of developmental genes and genes in control of pluripotency and cell proliferation, while tissue-specific genes are activated (22, 77). When new cell type is generated, it is epigenetic mechanisms that provide it with stability by maintaining the expression of key genes specific for the generated cell type while preventing the expression of genes specific for other cell types (52).

Epigenetic mechanisms include histone modification, DNA methylation, chromatin remodeling, and non-coding RNA (58). In this thesis, the focus is on DNA methylation.

1.10.1 DNA METHYLATION

DNA methylation is an epigenetic mechanism that plays a crucial role in versatile biological processes including development, aging, X chromosome inactivation, repression of retrotransposons, and genomic imprinting (22, 77). DNA methylation of promoters, the binding motifs of transcription factors, enhancers and super enhancers (large cluster of transcriptional enhancers), is associated with transcription inactivation (77). Moreover, DNA methylation is also involved in the pathogenesis of some diseases, such as cancer, when abnormally regulated. Both aberrant gain and loss of DNA methylation can be associated with the initiation and progression of diseases (2, 77). In contrast to

previous conceptions of being a stable modification, DNA methylation has been found to be dynamically regulated and reversible (83, 134).

DNA methylation occurs predominantly on the cytosine carbon 5 in CpG dinucleotide by a covalent binding of a methyl group. The methylation of cytosine is catalyzed by DNA methyltransferases, producing 5-methycytosines (5mC) (24). Methylation of CpG is typically symmetrical, that is, it occurs on both strands of the DNA (99). About 60-90% of all cytosines in CpG are methylated throughout the genome. However, sequences that are enriched with CpG dinucleotides, referred to as CpG islands (CGI), are frequently unmethylated (24, 77). Due to the mutagenic property of 5mC (24), CpG dinucleotides are depleted in the vertebrate genome, including that of mammals, where the ratio of observed to expected CpG is less than 35% (138). The amount of 5mC in all cell types is about 3-4% of all cytosines (142), where the same percentages of CpG are being methylated across all cell types (40). Moreover, 5mC correlates with heterochromatin, a condensed packaging of DNA and transcriptionally inactive regions (112), and provides a long-term repression of genes, transposons and repetitive sequences (19).

CpG ISLANDS

The vertebrate genome has a CpG depletion due to the mutagenic property of 5mC. However, some regions on the vertebrate genome, known as CGI, show enrichment in CpG content that are normally unmethylated and thus have a low rate of CpG depletion (24). CGI are regions of DNA that are greater than 500 base pairs long on average, with a CpG content of 55%, and the ratio of observed to expected CpG is higher than 65% (43). CGI are typically associated with transcriptional regulatory motifs. About 50% of CGI include transcription start sites (TSS), while the remaining are of uncertain function and referred to as orphan CGI. Orphan CGI are located either in the gene body (intragenic) or between annotated genes (intergenic). Hitherto, 40% of orphan CGI were also associated with TSS of genes, including non-coding RNA. CGI are rarely methylated at TSS even if the gene is transcriptionally inactive, while methylation of orphan CGI occurs more frequently, particularly in intergenic CGI. About 70% of annotated promoters contain CGI, including housekeeping, tissue-specific, and genes associated with development (24).

NON-CpG METHYLATION

Methylation of cytosine in non-CpG was also observed in mammals and plants, where cytosine is methylated in CpA, CpT or CpC. However, unlike methylations of CpG, methylation of non-CpG is asymmetrical and occurs only on one strand of the DNA. Non-CpG methylations of promoters have been

shown to be associated with transcription repression (99, 105). Nevertheless, non-CpG methylation in the gene body has been reported to be associated with gene activation (54, 105). Non-CpG methylation is frequent in gametes and stem cells, but absent in most somatic cells with the exception of infrequently dividing cells such as neurons and skeletal muscle. In addition, it reappears in reprogrammed cells, suggesting a role in pluripotency. However, how non-CpG methylation regulates gene expression is still to be elucidated (99, 105).

DNA METHYLTRANSFERASES

The methylation of cytosines is catalyzed by three DNA methyltransferases There are three enzymes that catalyze the DNA methylation. The first is DNA methyltransferase 1 (DNMT1), which is recruited to replication foci on the DNA and copies the methylation profile to the newly synthesized strand, hence serving as the DNA methylation maintenance enzyme. In addition, DNA methyltransferase 3A and 3B (DNMT3A and DNMT3B) act as the DNA *de novo* methylation enzymes. While DNMT1 has affinity to hemimethylated sequences, DNMT3A and DNMT3B are recruited to unmethylated sequences (86). In addition, DNA methyltransferase 3-like (DNMT3l), a family member of the *de novo* DNA methyltransferases lacking the transferase activity, was observed to identify unmethylated histone 3 lysine 4 (H3K4) and recruit DNMT3A and DNMT3B to methylate the DNA in that region (19, 77). Notably, the methylation of H3K4 is associated with gene activation and protects the DNA from *de novo* methylation. Nevertheless, DNMT3A and DNMT3B have also displayed demethylation activity (20). Another difference between DNMT3A, DNMT3B and DNMT1 is the fact that DNMT1 methylates only CpG sequences, while DNMT3A and DNMT3B methylates non-CpG as well (58).

However, DNA methylation is also regulated by long noncoding RNA (lncRNA), which have been observed to both guide and maintain DNA methylation as well as prevent the methylation of DNA in some contexts. They were also implicated in the formation of aberrant DNA methylation in cancers (77).

METHYLATION-DETERMINING REGIONS

Determination of DNA methylation profile was demonstrated to be regulated by cis elements found in methylation-determining regions (MDR), which are located on the promoters near the transcription start sites (TSS). The MDR includes DNA-binding motifs for transcription factors such as SP1, CTCF, and the RFX family, which regulate the methylation of DNA sequences. Mutations in these motifs result in hypermethylation of DNA sequences when

hypomethylation is expected. MDR have also exhibited protection of DNA methylation that extended to adjacent heterologous DNA sequences (76).

1.10.2 DNA DEMETHYLATION

DNA demethylation occurs in different biological contexts by two different mechanisms: active DNA demethylation and passive DNA demethylation. In the preimplantation embryo, rapid active DNA demethylation occurs on the parental DNA, where all DNA methylations are erased except for imprinted genes. On the other hand, the maternal DNA goes through slow passive demethylation. DNA demethylation of paternal and maternal genome in the early embryo is believed to promote the activation of the pluripotent genes to initiate the developmental process (88).

ACTIVE DNA DEMETHYLATION

Active DNA demethylation is catalyzed by different DNA demethylases, where different potential demethylation pathways were suggested (77). One example of this is DNA demethylation by the ten-eleven translocation (TET) family enzymes (TET1, TET2 and TET3) (22). TET enzymes demethylate 5mC by oxidation to 5-hydroxymethylcytosine (5hmC), which is oxidized further to 5-formylcytosine (5fC), and finally 5fC is oxidized to obtain 5-carboxylcytosine (5caC). There are versatile mechanisms to acquire the unmodified cytosine, such as the base excision repair (BER)-mediated excision of 5fC and 5caC by thymine DNA glycosylase that produces a gap in the DNA strand. Subsequently, the gap is repaired by ligation of the unmodified cytosine (77).

PASSIVE DNA DEMETHYLATION

Passive DNA demethylation occurs in a DNA replication-dependent manner through the inactivation of DNMT1 during the cell cycle, preventing the reproduction of DNA methylation pattern in the newly synthesized strand (77).

1.10.3 DNA METHYLATION AND DEMETHYLATION IN HPSC

DNA methylation in hESC is characterized by depletion at protein motifs and enrichment in the methylation of non-CpG dinucleotides (about 25%). Methylation of non-CpG occurs predominantly in the gene body, decreasing upon differentiation. Interestingly, non-CpG methylation is restored in reprogrammed hiPSC (22).

DNA methylation plays a key role in the reprogramming of somatic cells, as the activation of pluripotent genes is crucial for successful reprogramming of hiPSC (22). It is noteworthy that changes in DNA methylation occur towards the end of hiPSC reprogramming and take place after histone modifications (22). Demethylation is implied to have an important role in hiPSC reprogramming because using demethylation agents in the reprogramming procedure resulted in higher programming efficiency (22).

The methylation profile has also been suggested to affect the differentiation of hiPSC, indicating that an epigenetic memory may be extended to the hiPSC derivatives. In this regard, Xu *et al.* demonstrated that the differentiation of murine iPSCs to ventricular cardiomyocytes (chamber-specific cardiomyocytes) is far more efficient and homogenous than when murine ESC or murine iPSC, derived from distinct somatic cells, were differentiated to ventricular cardiomyocytes (144). On the other hand, Hewitt *et al.* observed that there was no significance difference between hESC-derived fibroblasts and fibroblasts derived from hiPSC derived from fibroblasts in regard to morphology and functionality (50).

1.10.4 DNA HYDROXYMETHYLATION

DNA hydroxymethylation occurs on cytosine, where a hydroxymethyl group replaces the hydrogen on carbon 5 (5hmC). 5mhC is detected in all somatic cells, but its content varies in different tissues, with the highest level being found in the brain. In the central nervous system, 5mhC level is 0.4-0.9% of all cytosines, corresponding to 10-20% of total 5mC, while in other tissues the 5mhC level is 0.03-0.2% of all cytosines (142). In addition, 5hmC is also enriched in ESC, where it consists of 5-10% of total 5mC (22). The 5hmC decreases upon differentiation, rising again in terminally differentiated cells (131).

5mhC is considered either an intermediate for oxidative demethylation or a stable modification that changes the affinity of 5mC binding proteins or promotes the recruitment of 5hmC-binding proteins (22, 119). In contrast to 5mC, 5mhC is enriched in the euchromatin region of the genome, where most of the genes are transcriptionally active. 5hmC is also enriched in the gene body of expressed genes, in addition to active enhancers and pluripotent transcription factors-binding motifs. (119). Nonetheless, 5hmC is correlated with both activated and repressed genes (56). The discovery that DNMT enzymes are involved in the production of 5hmC advocates for 5hmC being a stable epigenetic mark that influences genomic structure and function (131).

DNA HYDROXYMETHYLATION IN HPSC

The genomic DNA of HPSC contains a high level of DNA hydroxymethylation, ranging between 5% and 10% of all methylated cytosines. Considering the oxidation mechanisms that lead to the formation of DNA hydroxymethylation, a regulation role of 5hmC is implied in pluripotency establishment and differentiation (22).

Evidence shows that DNA hydroxymethylation is increased during the reprogramming of hiPSC. In fact, TET1 and NANOG enhance the efficiency of reprogramming and increase the 5mhC levels in the pluripotency regulators ESRRB and OCT4. Moreover, overexpression of TET1 was demonstrated to reprogram fibroblasts to hiPSC. TET is regulated by two miRNAs, miR-29a/b and miR-22, which are proposed to influence the efficiency of hiPSC reprogramming by inhibiting the translation of TET proteins (22). Moreover, ascorbic acid was also demonstrated to improve the efficiency of reprogramming of hiPSC through the regulation of TET activity, as it was observed to increase the level of 5hmC in the DNA (22).

1.10.5 ABNORMAL METHYLATION OF CULTURED CELLS

Abnormal hypo- and hyper-methylation of CpG islands was observed in cultured somatic cells. Furthermore, several studies reported the emergence of abnormal methylation patterns in hESC lines, where these aberrances were demonstrated to worsen upon differentiation (81). For instance, Shen *et al.* revealed abnormal DNA methylation in hESC-derived neural progenitor/stem cells (NPC), compared to primary NPC, that affected the hESC-derived NPC functionality (120). Similar abnormalities have also been reported by Brunner and colleagues (15). In their work, they demonstrated that the methylation profiles of hESC and hESC-HEP have larger fractions of DNA-methylated regions compared to FL and AL. Moreover, in a previous study, they showed that DNA demethylation events are characteristic for *in vivo* liver development, while *in vitro* differentiation is characterized by both demethylation and *de novo* methylation (15). Considering these findings, there is a probability that abnormal DNA methylation underlies the immature functionality of hPSC-HEP, potentially explaining the differences in the expression profiles between the *in vitro* products and the *in vivo* tissues.

1.10.6 DNA METHYLATION AND HYDROXYMETHYLATION IN THE LIVER

Studies have shown that in the liver, DNA methylation may contribute to the inter-individual variation in drug response. Interestingly, the genetic factor was observed to account for about 20-30% of the inter-individual variation in the response of ADME genes to drugs, proposing the involvement of epigenetic regulation of these genes in higher proportion of the inter-individual variation (57). For example, the variability in the expression of *GSTM1* was only 11% explained by the strongest SNP in the DNA sequence of the gene, while the methylation level of a specific CpG site in *GSTM1*, was reported to account for up to 55% of the expression variability (12). A study by Ivanov *et al.* revealed that AL is abundant with 5hmC, which correlated with actively transcribed genes. A subsequent study revealed that 5hmC abundant genes are actively transcribed, while 5hmC-poor genes were downregulated or silent. Furthermore, the 5hmC content in some ADME genes was demonstrated to exceed the level of 5mC in some CpG sites. In addition, the content of 5hmC was reported to be highly variable across AL from different individuals, while much less variation was seen for the 5mC content. One possible explanation was that 5hmC are more sensitive to environmental changes than 5mC (57).

2 AIMS

The aim of the thesis was to investigate the transcriptome and methylome of *in vitro* hepatic differentiation of human pluripotent stem cells in order to identify similarities and differences between human pluripotent stem cell-derived hepatocytes and human adult liver tissues.

2.1 SPECIFIC AIMS

- The evaluation of the efficiency of the applied hepatic differentiation protocol considering the generation of functional biological replicates with minimal variation and synchronicity of the expression of genes throughout the differentiation procedure (addressed in Paper I).

- The verification of whether the *in vitro* hepatic differentiation procedure recapitulates the *in vivo* development of the liver via thorough transcriptomic analysis of the following developmental stages: the definitive endoderm, the hepatoblasts, the early hepatocytes and the mature hepatocytes (addressed in Papers I and II).

- The identification of similarities and differences in the transcriptome between hPSC-HEP and human adult liver tissue (addressed in Paper III).

- The determination of whether hPSC-HEP is mostly correlated to human adult or fetal liver tissue (addressed in Paper III).

- The identification of similarities and differences in DNA methylation between hPSC-HEP and human adult liver tissue (addressed in Paper IV).

- The integration of methylome and transcriptome data to identify associations between the pattern of DNA methylation and gene transcription.

3 METHODS

3.1 DIFFERENTIATION OF HPSC-HEP

To minimize the variability in the differentiation experiments, hPSC lines that were used in this study were all of low passage (p.10-21) and from one gender only (XY genotype). In addition, to reveal any potential differences in the differentiation procedure related to the origin of the hPSC, three hESC and three hiPSC were included. All the hPSC were thawed, maintained, and passaged in the feeder-free DEF-CS™ Culture System prior to differentiation. The differentiation procedure started with the application of the Cellartis® Definitive Endoderm Differentiation Kit, developed by Takara Bio Europe AB and at day 7 of the differentiation procedure >90% pure definitive endoderm cells were obtained. At day 7, the cells were dissociated and harvested according to the manufacturer's recommendations, and the differentiation of hPSC-HEP was initiated applying a prototype of the Cellartis® Hepatocyte Differentiation Kit, provided by Takara Bio Europe AB. The differentiation procedure resulted in >90% pure hPSC-HEP.

The cells were harvested every day during the differentiation period (day 0 to day 35). RNA was extracted from all the samples, and DNA was extracted on days 0, 5, 14, 25 and 30, using commercial kits, for further examination in downstream analysis.

3.2 OLIGONUCLEOTIDE MICROARRAYS

Oligonucleotide microarray (DNA chip) is a large-scale quantification technique using DNA chips, which allows for parallel quantification of mRNA transcripts using hybridization assay (45). This technique takes advantage of nucleic acids property to hybridize to each other forming hydrogen bonds.

3.2.1 TRANSCRIPTOME MICROARRAY ANALYSIS

To investigate the dynamics of the transcriptome during the differentiation process of hPSC-HEP, the GeneChip® Human Transcriptome Array 2.0 from Affymetrix was applied and transcriptome differences between samples from the differentiation of hPSC to hPSC-HEP and human liver tissues were investigated. This array has high coverage including at least 285,000 full-length transcripts, 245,000 coding transcripts, 40,000 non-coding transcripts, and 339,000 probe sets covering exon-exon junctions.

The normalized expression measures from the microarray runs, were extracted by applying the robust multi-array average (RMA) algorithm, which corrects for background noise, performs $\log_2$ transformation on the corrected intensities, then performs quantile normalization and finally fits a linear model for each probe set using median polish as estimator for the parameters of the models (55).

3.2.2 METHYLATION MICROARRAY ANALYSIS

To study the genome-wide DNA methylation dynamics during the differentiation process of hPSC-HEP, the Illumina Infinium MethylationEPIC BeadChip was applied. This array includes over 850,000 CpG sites, that cover 99% of the RefSeq genes and 95% of the CpG islands with an average number of six probes per island (80). Illumina applies a bead technology to produce the Infinium microarrays. Each bead holds thousands of copies for a specific probe of 73 bases long, of which 23 bases represent the base address and 50 bases target specific sequence of the genome. The beads are distributed randomly on the chip, while the base addresses identify the location of each probe (104). The technology is based on a two-color array technique (red and green) to determine the methylation status of each probe. For each methylation locus two signals are recorded methylated (M), and unmethylated (U) (2).

In addition to standard normalization procedures required to reduce the technical variation introduced during the experimental processing of the microarrays, normalization to reduce effects of bead type is also performed using the BMIQ method (85).

BISULFITE CONVERSION

Bisulfite conversion is the gold standard of measuring DNA methylation at a single-base resolution (77). Bisulfite treatment of DNA deaminates unmethylated cytosines into uracil while maintaining methylated cytosines, allowing the identification of methylated loci (47). However, bisulfite conversion does not distinguish between 5mC and 5hmc. Two different methods were developed to distinguish between 5mC and 5hmC: oxidative bisulfite (oxBS) treatment, and TET-assisted bisulfite (TAB) treatment. Comparative studies between these two methods were conducted producing contradictory outcomes regarding the efficiency of both methods. Ivanov *et al.* claim that TAB treatment is more sensitive than oxBS. However, Li *et al*. presumed that the observed sensitivity of TAB over oxBS could be a result of the higher kinetic efficiency of the TET enzymes in producing 5hmC compared to subsequent products, which would result in incomplete treatment of the TAB method, thus introducing *de novo* 5hmC (31).

OXIDATIVE BISULFITE TREATMENT

Oxidative bisulfite (oxBS) treatment applies potassium perruthenate (KRuO4) to the DNA to oxidize all 5hmc to 5fC. Similarly to unmethylated cytosines, 5fC is then converted into uracil by bisulfite treatment (13).

TET-ASSISTED BISULFITE TREATMENT

TET-assisted bisulfite (TAB) treatment is initiated with the application of glucosylation with β-glucosyltransferase to generate β-glucosyl-5-hydroxymethylcytosine (5gmC) from all the 5hmC that are found in the investigated genome. This process prevents further oxidation of 5hmC by TET proteins. Subsequently, the DNA is treated with TET proteins that convert all 5mC to 5hmC and further to 5fC and 5caC. 5fC and 5caC are transformed to uracil with bisulfite conversion (147).

In this work, the oxBS has been used to distinguish between 5mC and 5hmC to avoid the possible introduction of *de novo* 5hmC.

3.3 CYTOCHROME P450 ENZYME ACTIVITY ASSAY

The activity of drug-metabolizing enzymes is determined by measuring the resulting metabolites from metabolized substrates that are unique for each metabolizing enzyme. The metabolites are measured applying liquid chromatography mass spectrometry (LC/MS), which combines the separation of molecules or ions by liquid chromatography and mass spectrometry. The technology separates molecules or ions, then the separated masses are charged and analyzed on the basis of mass-to-charge ratio by mass spectrometry to identify the different separated compounds (106). Ultimately, the metabolites were quantified applying Pierce BCA Protein Assay Kit (www.thermofisher.com) with standard spectrophotometry.

The substrates that were applied are phenacetin, a CYP1A substrate; bufuralol, a CYP2D6 substrate; diclofenac, a substrate of CYP2C9; mephenytoin, a substrate of CYP2C19; and midazolam, a substrate of CYP3A. The resulting metabolites were measured: paracetamol (CYP1A), 1-OH-bufuralol (CYP2D6), 4-OH-diclofenac (CYP2C9), 4-OH-mephenytoin (CYP2C19), and 3-OH-midazolam (CYP3A).

3.4 REAL-TIME POLYMERASE CHAIN REACTION

Real-time polymerase chain reaction (RT-PCR) is a technology that enables the detection and measurement of PCR amplicons during each cycle of the PCR. The increase of amplicons is exponential and proportional to the abundancy of the target sequence in the reaction mixture. In this thesis, TaqMan chemistry was applied to detect amplification of lineage markers during hepatic differentiation. This technology includes a probe complementary to the target sequence. This probe has fluorescence dye on its 5′ end and a quencher on its 3′ end, applying fluorescence resonance energy transfer (FRET) as quenching mechanisms. As long as the quencher and the fluorescence dye are attached to the probe and close to one another, the quencher silences the signal of the fluorescence dye. The probe is degraded gradually, releasing the fluorescence dye with the progression of the DNA polymerase on the template. The fluorescence signal will increase with each PCR cycle and will be proportional to the amount of the DNA template (100).

3.5 IMMUNOCYTOCHEMISTRY

Immunocytochemistry (ICC) is a method that makes use of the properties of antibodies to detect molecules in the cells. The antibody is a protein called immunoglobin that is generated as an immune response. The antibody targets species-specific epitopes on antigens. These antigens can be proteins, peptides or molecules that cause an immune response, while the epitope is a small part of the antigen that can be targeted by a specific antibody.

ICC starts with the fixation of the cells to stabilize and preserve the cells' contents and structure. Fixation usually is performed by crosslinking using chemicals with aldehyde groups that binds the different molecules in the cell and holds them in a consistent position as in living cells. Afterwards, a permeabilizing agent is applied, if the target is not located on the surface of the cell, to allow the penetration of antibodies to the interior of the cells. Then a blocker is used to prevent unspecific binding of the antibodies to charged molecules. Subsequently, in direct ICC, a fluorescence-tagged primary antibody specific for the investigated molecule is applied to the cells, allowing the presence of the target to be monitored in fluorescence microscopy. However, in indirect ICC, after incubation of the cells with untagged primary antibody, the cells are washed to remove excess antibodies, and a fluorescence-tagged secondary antibody that recognizes the primary antibody is applied to cells. Subsequently, the results can be viewed in fluorescence microscopy (16).

3.6 BIOINFORMATICS AND STATISTICAL ANALYSIS

3.6.1 NORMALIZATION

QUANTILE NORMALIZATION

Quantile normalization (QN) is customarily applied on signal intensities to reduce sample variation by making the distribution of the probe intensity of each array in a set of arrays equal (11). This method centers the signal between arrays and corrects for position bias (46).

BETA MIXTURE QUANTILE DILATION

The beta mixture quantile dilation (BMIQ) normalization corrects for probe-type bias with some reduction of the technical variation. However, applying only BMIQ normalization was shown to be a less effective reduction of technical variation compared to QN (46). BMIQ normalization fits mixture models for the beta-values of nonmethylated, hemimethylated and completely methylated probes (67).

3.6.2 SIGNIFICANCE ANALYSIS OF MICROARRAYS (SAM)

To identify differentially expressed genes, a significance analysis of microarrays (SAM) was applied. SAM applies a set of gene-specific t-tests in which each gene is assigned a score (relative expression) based on the ratio of the difference of gene expression to the standard deviation of the expression of the gene. Controls for random fluctuation of the data were generated by permutation in order to assign statistical significance for the observed difference of gene expression. Expected relative expression for each gene is calculated as the average of all permutation results for that specific gene. A scatter plot is generated for the observed relative expression based on the original data set results versus the expected relative expression calculated from the permutation results. If the observed relative expression is similar to the random fluctuation simulated by the permutations, the scatter plot will depict a straight diagonal line. To account for false positive results, two horizontal cutoffs were set to determine the least relative difference of induced and repressed genes. The horizontal cutoffs are determined by controlling the false discovery rate (FDR) (Figure 2). SAM has been demonstrated to perform well,

and the FDR estimate yielded by SAM is relatively low compared to other methods (137).

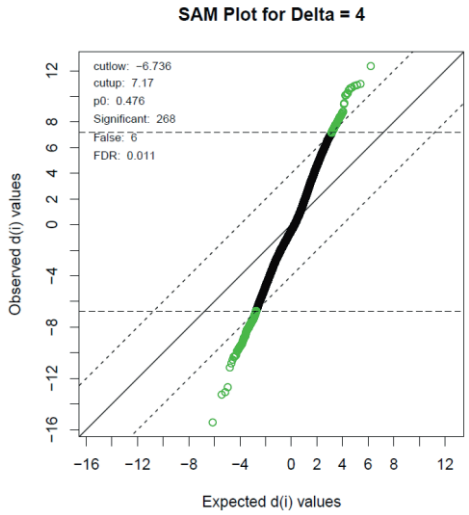

Figure 2. An example SAM plot showing the plot of expected to observed relative expression. The horizontal cutoffs are marked by diagonal dashed lines. The green circles represent differentially expressed genes.

3.6.3 METHYLATION ARRAY ANALYSIS

The unmethylated (U) and methylated (M) intensities are used to determine the proportion of methylation for each probe. The methylation levels are reported as beta-values (β = M/(M+ U)) or the logit of beta-values, the M-values (M value = log2(M/U)) (84).

DIFFERENTIALLY METHYLATED PROBES

Differentially methylated probes (DMPs) were identified applying *champ.DMP* from the *ChAMP* package in R (133), which makes use of the calculation algorithm of differentially expressed genes in the *Limma* package. This method generates a general linear model for the microarray experiment to analyze the entire experiment as a complete unit (109). To correct for multiple testing effects, the t-statistic is modified with the empirical Bayes approach. The probability distribution of the test is estimated from the marginal

distribution of the observed values. This method improves the statistical power and FDR compared to the conventional t-test (82, 103, 122, 123)

DIFFERENTIALLY METHYLATED REGIONS

Differentially methylated regions (DMR) are adjacent regions on the genome that are differentially methylated between phenotypes. To identify DMR, we used the *DMRcate* method. This method relies on the results of differential methylation for each probe performed by the *Limma* package, applying linear model fit for all probes, or by computing the variance, using Gaussian smoothing of the test statistics for each probe compared to the results of neighboring probes to reduce noise. Afterwards, a model is constructed to the smoothed test statistics and the *p*-values are calculated for these models and corrected for multiple testing. Finally, nearby significant sites are aggregated into significant DMR (101).

3.6.4 CORRELATION ANALYSIS

In this thesis correlation analysis were applied to investigate associations between samples or genes. In addition, it was used as proximity measure in the clustering analyses. Different correlation methods were applied dependent on the distribution of the data.

SPEARMAN'S RANK CORRELATION

Spearman's rank correlation is a non-parametric statistic test that is applied to assess reciprocal relation between variables. This method does not require the assumption of normal distribution of the input data. Furthermore, as the correlation is performed on ranks, Spearman's correlation is insensitive to outliers (70, 93). The correlation is measured by a correlation coefficient that has values ranging between -1 to +1, where -1 represent a perfect negative correlation, +1 represent a perfect positive correlation, and 0 represent no correlation (93).

PEARSON CORRELATION

The Pearson correlation is a parametric test to assess linear association between two normally distributed continuous variables. In contrast to Spearman's rank correlation, this method is sensitive to outliers. The correlation is measured using a correlation coefficient, which denotes the strength of the linear relationship between the variables. The correlation coefficient ranges between perfect negative (-1) and perfect positive (+1) correlation, and 0 indicates no correlation (93).

3.6.5 CLUSTERING ANALYSIS

Clustering analysis is used to reduce dimensionality in large-scale datasets and facilitates the downstream analysis in revealing possible similarities among the members in the resulting clusters. The clustering is based on the characteristics of the objects to be clustered. Robust clustering techniques generate clusters of objects with high similarity within a cluster as well as large difference between clusters. Commonly, the Euclidean distance (the straight line distance between two points) is used for the proximity measure (97). In this study, the Pearson correlation was applied as similarity measure in both the Kmeans and the hierarchical clustering analyses. Pearson correlation is widely used as proximity measure for the clustering of gene expression data (126).

AGGLOMERATIVE HIERARCHICAL CLUSTERING

Hierarchical clustering is a graph-based type of nested clustering that allows the formation of subclusters. The results are displayed graphically as a dendrogram. There are two types of hierarchical clustering, agglomerative and divisive. Agglomerative hierarchical clustering starts with each object as a cluster and performs merging of clusters until only one cluster remains, while divisive hierarchical clustering begins with all objects as one cluster and iteratively divides the clusters.

Agglomerative hierarchical clustering algorithm starts with computing the proximity matrix. Then it merges the closest objects in the matrix together and updates the proximity matrix according to the resulting clusters. This procedure is repeated until one cluster is produced. There are three approaches for merging of objects to clusters; the average linkage, the complete linkage, and the single linkage. The average linkage defines the distance between clusters as the average of all the pairwise distances between the objects from the different clusters. The complete linkage, considers the distance between the farthest two objects in the different clusters. The single linkage, considers the distance between the two closest objects (97).

KMEANS CLUSTERING

Kmeans clustering is a prototype-based partitioning clustering technique that divides the data into a specified number of non-overlapping clusters, where each object is assigned to only one cluster and is closer to the prototype of this cluster than any other cluster. In Kmeans, the prototype is the centroid of the cluster. The algorithm of the Kmeans starts with randomly selecting one centroid for each cluster. The number of clusters is determined by the user. Each object in the dataset is then assigned to the cluster with the closest centroids. Afterwards, the centroid of each generated cluster is recomputed,

and the objects are again assigned to the closest centroid. An iteration of calculation of new centroid and the assignment of the objects to the cluster with the closest centroid continues until convergence. Kmeans clustering is an efficient clustering algorithm; however, it is sensitive to outliers and can produce slightly different clusters each time it is run because of the arbitrary assignment of the initial centroids (97).

3.6.6 ENRICHMENT ANALYSIS

Enrichment analysis can be performed on lists of interesting objects compiled from large-scale profiling of, among others, DNA, RNA, or proteins, when compared with libraries of sets (e.g., lists of genes or proteins) generated based on prior knowledge (21, 71). These findings may reveal new knowledge and clues about the objects in the list of interest. Genes are organized in libraries associating these genes with functional terms including pathways, regulatory transcription factors, gene ontology, disease, drugs, cell types, and miscellaneous. The enrichment can be computed applying different methods, such as the Fisher exact test or the overrepresentation analysis (ORA) in which the common genes between the queried list and the sets in the knowledge database are counted and the statistical significance of the overlap is calculated (79). The EnrichR software implements another method that is a correction of the Fisher exact test. It computes mean rank and standard deviation of the queried list and compares them to the expected ranks of the sets in the libraries, returning a z score that shows the deviation from the estimated rank. A third method is computing the *p*-value of the Fisher exact test and multiplying it with the z scores to obtain a combined score (21).

PATHWAY ENRICHMENT ANALYSIS

Pathway enrichment analysis applies libraries of gene sets from pathway databases such as Wikipathways, KEGG, and Reactome, as reference for querying gene list to investigate overrepresentation of the members of the specific gene list in some of the predefined gene sets (79).

GENE ONTOLOGY ENRICHMENT ANALYSIS

Gene ontology (GO) is an annotation system in which genes and gene products are attributed, structured, and precisely defined, using common and controlled vocabulary for describing their role. This process of terminology is dynamic as each term of GO is linked to the different gene and protein databases to be updated with the latest changes and discoveries. GO is divided into three categories: biological process, molecular function, and cellular component. The GO terms have parent and children terms hierarchically connected to each other as nodes in a network (4). Gene lists are compared to lists of genes in the

different GO categories, and statistical tests are performed to detect significant results.

TRANSCRIPTION FACTOR ENRICHMENT ANALYSIS

Enrichment analyses can also reveal overrepresentation of binding sites for specific transcription factors in sets of genes. There are libraries that associate gene sets with transcription factors, including the ChEA database. The ChEA database includes targets for transcription factors extracted from available publications. In addition, scanning the promoters of all genes in the genome with position weight matrices generated from JASPAR and TRANSFAC transcription factor databases is another method that can be applied to predict motifs for different transcription factors on the promoters (21).

3.6.7 PROTEIN INTERACTION NETWORK ANALYSIS

Protein interaction networks are used to investigate relationships between genes in gene sets. The logic behind this method is that genes that are located close to each other in the network are likely to be involved in the same biological process (79). The protein interaction networks are illustrated as an assembly of nodes representing the components and edges signifying the interaction between the components (116).

In protein interaction networks, proteins that have many more interactions among themselves than the rest of the network cluster together in a protein module (125). The members of the module are commonly involved in the same regulatory pathways and have similar functions. Therefore, the function of unknown members in a module could be inferred from the function of other members (116). In this thesis, the prediction of modules was performed using the MCODE application in Cytoscape (118). The application predicts the modules based on the density of the regions in the protein interaction network, where the density of a graph is defined as the existing edges between the nodes divided by the theoretical maximum number of edges possible for the graph (7).

Hub proteins are proteins with a high level of interactivity and are usually essential, playing central roles in protein interaction networks. In addition, hub proteins tend to be evolutionarily conserved. As there is no defined threshold to determine the properties of hub proteins, research groups have applied different strategies to identify hub proteins (139). These strategies have been based on various combinations of different indices of topological properties of the genes both on the local and global levels (28, 78, 139). The local level includes the node degree, which is the number of connections of a node in the

protein network, and the clustering coefficient of the node, which is the proportion of interacting nodes that also interact with each other. The global levels include the network centralities: stress centrality, betweenness centrality, and closeness centrality. The closeness centrality is the measure of the shortest paths for a node to reach to all other nodes in the network. The betweenness centrality is the frequency of a node being on the shortest path between two other nodes, and the stress centrality is the number of shortest paths passing through the specific node (78).

3.6.8 ComBat

The merging of two datasets from different microarray experiments run on different platforms, Affymetrix and Illumina, was performed applying the ComBat software (136). This method is based on the empirical Bayes method and estimates its parameters by the borrowing of information from genes and conditions of the experiment. To avoid introducing bias resulting from distinct expression level and probe sensitivity of each gene, this method first standardizes the variance and mean for each gene, resulting in mean and variance shrinkage. The batch effect parameter is then estimated from the standardized data. These estimators are used to correct for the batch effect in the data. Overall, this method is robust to small sample size and insensitive to outliers (61).

4 RESULTS

This thesis is based on results from four studies focusing on different research questions related to *in vitro* differentiation of hepatocytes. The samples that were investigated in each study were identical and sourced from the same batches of cells.

4.1 PAPER I: HIGHLY SYNCHRONIZED EXPRESSION OF LINEAGE-SPECIFIC GENES DURING IN VITRO HEPATIC DIFFERENTIATION OF HUMAN PLURIPOTENT STEM CELL LINES

To study DNA methylation and transcriptional patterns during hepatic differentiation, six XY hPSC lines, including three hESC lines and three hiPSC lines, were differentiated into hPSC-HEP (N=2).

To determine representative time points at which to collect samples for the developmental stages, DE, hepatoblasts, early hPSC-HEP cells, and late hPSC-HEP samples were collected daily during the hepatic differentiation from day 0 to day 35. The samples were split into two aliquots to be used for both RNA and DNA analysis. RT-qPCR was run on the RNA split for 16 lineage-specific genes, and correlation and clustering analyses were applied on the expression data. The results revealed that the expression of lineage-specific genes was synchronized across all six cell lines. While markers for early developmental stages, including pluripotent primitive streak, DE, and ventral foregut, showed a very strong correlation between cell lines, markers for later stages, including hepatoblast, fetal hepatocyte, and mature hepatocytes, displayed a varied correlation between cell lines ranging between weak to strong. The observed weak correlation regarding the expression of CYP*3A4* is attributed to the fact that *CYP3A4* is a polymodal metabolic enzyme. However, each gene showed the same pattern of upregulation and downregulation for all of the six cell lines. Results from a clustering analysis demonstrated the distribution of markers in clusters reflecting the different stages of hepatocyte differentiation. Deviations in the expression of some genes compared to *in vivo* hepatocyte development were also observed. Based on these results, the days selected to represent the typical developmental stages of hepatic differentiation were days 5, 14, 25, and 30, representing DE, hepatoblast, early hPSC-HEP, and late hPSC-HEP respectively.

In addition, results from xenobiotic-metabolizing enzyme (XME) activity assays revealed considerable XME functionality in hPSC-HEP compared to cryoplateable human primary hepatocytes such as CYP3A, CYP1A, CYP2C9, and CYP2C19. Moreover, results from immunocytochemistry revealed expression of CYP3A4 and the drug transporters BSEP, NTCP, MRP2, and OATP1B1, in addition to hepatic markers HNF4A, AAT, ALB, and CK18. Furthermore, hPSC-HEP showed an ability to store glycogen.

4.2 PAPER II: COMPARATIVE TRANSCRIPTOMICS OF HEPATIC DIFFERENTIATION OF HUMAN PLURIPOTENT STEM CELLS AND ADULT HUMAN LIVER TISSUE.

To investigate the transcriptional dynamics of the differentiation of hPSC-HEP, comprehensive transcriptome analysis was conducted applying genome-wide transcriptome microarrays. Based on results from Paper I four stages during the hepatic differentiation were selected as representative for DE, hepatoblasts, early hPSC-HEP, and late hPSC-HEP and these were investigated for transcriptomics analysis using microarrays. Based on the peak expression levels for the different developmental markers, day 5, day 14, day 25, and day 30 were good representations for DE cells, hepatoblasts, early hPSC-HEP, and late hPSC-HEP. In addition, the early developmental stages hPSC and DE clustered together, while hepatoblasts, early hPSC-HEP, and late hPSC-HEP huddled together. AL also clustered closer to the late developmental stages than to the early stages. Interestingly, no distinction was observed between cells originated from hESC and hiPSC. To identify differentially expressed genes across all investigated developmental stages, significance analysis of microarrays (SAM) was applied and the top 2,000 different genes were investigated further. Kmeans was performed on these top 2,000 differentially expressed genes to explore the transcriptional profile of the differentiation process. Nine distinct clusters were generated, representing the different developmental stages. In-depth pathway and gene ontology enrichment analyses on each of the clusters confirmed the involvement of the genes from the respective clusters in functions and processes characteristic for that developmental stage. Cell cycle and growth genes were enriched in early stages, while genes involved in hepatic functionality were enriched in the later stages. Hub proteins discovered in this study were mainly involved in essential cell processes such as cell cycle and growth. Moreover, identified modules also showed involvement in essential biological processes. However, a module

from the cluster characterizing the AL, showed particularly interesting patterns. This module included drug metabolism enzymes and Phase I and II genes (*CYP4A11, CYP2A6, CYP2B6, CYP2C9, CYP2E1, UGT2B4*, and *AOX1*). These genes are essential for biotransformation, but they were distinctly downregulated in the hPSC-HEP samples.

4.3 PAPER III: NOVEL TRANSCRIPTOMICS TARGETS FOR FUNCTIONAL IMPROVEMENT OF HEPATIC DIFFERENTIATION OF HUMAN PLURIPOTENT STEM CELLS.

In this study, transcriptomics results were further investigated and pairwise comparisons between samples from sequential developmental stages were performed. Furthermore, the transcriptomics dataset of this study was merged with another transcriptomics dataset retrieved from the Gene Expression Omnibus (GEO) database. This dataset included human adult (AL) and fetal (FL) liver tissues of different ages and developmental weeks respectively. This dataset was instrumental for mapping the developmental stages of the *in vitro* hepatic differentiation to those of *in vivo* hepatic development. Although the datasets are generated using different microarray platforms, they were successfully normalized applying the *ComBat* method in R. Of special note from this study is that the correlation and similarity analysis between hPSC-HEP, FL, and AL revealed a higher correlation and similarity of hPSC-HEP to AL than to FL. To the best of our knowledge this has not been demonstrated in previous studies.

Differentially expressed genes in the following comparisons: hPSC-HEP vs. DE, DE vs. hepatoblasts, hepatoblasts vs. hPSC-HEP day 25, hPSC-HEP day 25 vs. hPSC-HEP day 30, and hPSC-HEP day 25 vs. AL, were detected when applying the SAM algorithm. The results revealed clear distinction of developmental markers at the respective differentiation stages. A putative novel DE biomarker, *RP4-559A3.6,* was discovered. This transcript partially overlaps with the *LEFTY2* sequence. Furthermore, only 11 genes were differentially expressed between hPSC-HEP day 25 and hPSC-HEP day 30. These genes had no obvious connection to hepatic functionality and were not investigated further. Genes that were differentially expressed between hPSC-HEP day 25 and AL included 22.4% of the total list of ADME genes. Members of the functional module that was discovered in Paper II, *CYP2E1, CYP2C9*, *CYP2A6, CYP2B6*, and *AOX1*, were among the downregulated ADME genes in hPSC-HEP. Interestingly, the regulator *NR1I3* (CAR) and the coregulator

PPARGC1A were also among the downregulated genes in hPSC-HEP compared to AL.

Subsequently, overlap of differentially expressed genes from the different comparisons were investigated using a Venn diagram by applying the *Venndiagram* package in R. Notably, only four genes (*COL5A2, STC1, ACE2*, and *LINC00261*) were discovered to be differentially expressed across all developmental transitions. These were also differentially expressed between hPSC-HEP and AL; thus, the transcription factors that may regulate these genes were also investigated. The genes have binding sites for the following transcription factors: AR, BACH1, FOXA1, FOXA2, FOXM1, JUN, and SOX2, of which only *AR* was differentially expressed between hPSC-HEP and AL.

Gene ontology and pathway enrichment analysis were applied on the differentially expressed genes identified in the different developmental transitions. The results revealed that differentially expressed genes between hPSC and DE were involved in differentiation pathways. Differentially expressed genes between DE and hepatoblasts were involved in the cell cycle and differentiation pathways in addition to various hepatic function pathways.

Finally, enrichment analysis for binding site of transcription factors was also performed for the differentially expressed genes from the different comparisons. Moreover, a Venn diagram was constructed for visualization of the results. The results show that the transcription factors AR, BACH1, FOXA1, FOXA2, FOXM1, JUN, and SOX2 are putative regulators of the differentially expressed genes identified at the different comparisons. In addition, differentially expressed genes between hPSC-HEP and AL were enriched for binding sites of the transcription factors CDX2, CLOCK, HIF1A, and NR1H3.

4.4 PAPER IV: IDENTIFICATION OF HYPERMETHYLATED GENES INVOLVED IN HEPATIC FUNCTIONALITY IN HUMAN PLURIPOTENT STEM CELL-DERIVED HEPATOCYTES.

Here, the methylation status of the developmental stages hPSC, DE, hepatoblast, hPSC-HEP day 25, hPSC-HEP day 30, and AL was determined. A hierarchical clustering was constructed for the beta-values of the different samples, which revealed clustering of the biological replicates from each developmental stage together. Furthermore, hPSC-HEP day 25 and hPSC-HEP day 30 clustered together without any distinction between the biological replicates. AL clustered closer to hepatoblasts and hPSC-HEP than to hPSC and DE. Interestingly, cells derived from hiPSC in hPSC and DE clustered together in each category; however, this clustering diminished in later stages of the differentiation.

The methylation profiles of hPSC-HEP day 25 were compared to the methylation of AL, and the results were integrated with the transcriptome data from the previous study (Paper III). As the bisulfite conversion does not distinguish between 5mC and 5hmC, we applied the oxidative bisulfite treatment (oxBS) to one aliquot of the investigated samples to identify 5hmC as well. Firstly, differentially methylated probes (DMPs) were identified between bisulfite-treated hPSC-HEP versus oxBS-treated hPSC-HEP and bisulfite-treated AL versus oxBS-treated AL. The probes that were differentially expressed were considered hydroxymethylated and were removed from both hPSC-HEP and AL at this stage. Subsequently, differentially methylated probes were identified between hPSC-HEP and AL. These probes were found to be annotated to about 72% of the differentially expressed genes between these samples. A significant association between highly methylated and downregulated genes and vice versa was also demonstrated.

Interestingly, the members of the functional module identified in Paper II that were silenced in hPSC-HEP compared to AL (*AOX1, CYP2A6, CYP2B6, CYP2C9, CYP2E1, CYP4A11*, and *UGT2B4*), were found to be hypermethylated in hPSC-HEP compared to AL. Investigation of the regulators of these genes also revealed abnormal hypermethylation. While the regulators *AR, PPARGC1A*, and *NR1I3* were both downregulated and hypermethylated in hPSC-HEP compared to AL, *CDX2* and *IGF1R* were

upregulated and hypomethylated in hPSC-HEP. Notably, no miRNA regulating these genes were found to be abnormally expressed or methylated between hPSC-HEP and AL. Furthermore, two regulators *PITX2* and *β-catenin* were highly expressed in hPSC-HEP compared to AL. Despite its overexpression in hPSC-HEP, *PITX2* was also hypermethylated in hPSC-HEP compared to AL. These regulators, in addition to the above-mentioned regulators, may have key roles in rectifying the functionality of the module.

5 DISCUSSION

This thesis is based on comprehensive transcriptome profiling of the differentiation process of hPSC-HEP in comparison to liver tissues and further integration of the obtained results with DNA methylation investigation of hPSC-HEP compared to human adult liver.

Although hPSC-HEP are already applied in disease modelling, including infectious diseases such as Hepatitis C virus and malaria, and in drug screening where they demonstrated the capacity to predict hepatotoxins and the efficacy of a drug (44), the immature phenotype of these cells and other safety considerations prevent their application in advanced stages within regenerative medicine, cell therapy, drug discovery and development. However, the tremendous improvement in the production of functional hPSC-HEP in the last few years foretells the significant role that these cells will play in those fields in the near future.

5.1 DIFFERENTIATION OF HPSC-HEP

The robustness of the applied protocol demonstrated in Paper I showed impressive synchronicity of the differentiation process across the different hPSC lines. This property was critical for the comprehensive investigation of these cells throughout the thesis, as no adjustment in the differentiation was needed and the samples that were analyzed in the following studies could be acquired from the same time points for all cell lines. As all the samples could be treated equally and be exposed to the same conditions and days in culture, unnecessary confounds in the experiment settings were significantly reduced. As clearly shown in the hierarchical clustering of both the transcriptome and the methylome datasets, the biological replicates of each specific developmental stage clustered tightly together (Papers II, III and IV). Furthermore, the end-point product of the differentiation, hPSC-HEP, exhibited considerable functionality comparable with what is observed in cryoplateable primary human hepatocytes in terms of the activity of xenobiotic-metabolizing enzymes CYP3A, CYP1A, CYP2C9, and CYP2C19, in addition to protein expression of CYP3A4 and the drug transporters BSEP, NTCP, MRP2, and OATP1B1. Furthermore, they were demonstrated to express hepatic markers at both transcriptional and translational levels, along with the ability to store glycogen. Compared to previous products, hPSC-HEP shows profound improvement, particularly regarding XME activity and the homogeneity of the cell cultures (14).

The differentiation process can be divided into two main parts by the re-plating procedure, and these are the differentiation to DE and the subsequent hepatocyte differentiation. The generation of homogenous DE cells in addition to the re-plating procedure were claimed to have substantial effect on the efficiency of the hepatic differentiation (5, 14, 150). The best possible result to obtain functional hPSC-HEP has been achieved by mimicking the embryonic developmental process. Furthermore, migration is essential to allow the cells to move to a new environment and interact with different cells that promote their development toward the destined cell fate (38, 121). This process was mimicked by the re-plating of cells *in vitro*.

Another property illustrated in the results from Papers III and IV was the stability of hPSC-HEP in culture, as both the transcriptome and the DNA methylation demonstrated minimal variation between hPSC-HEP day 25 and day 30. These results demonstrate the capacity to maintain functional hPSC-HEP in culture and prevent dedifferentiation, which is a major problem with primary hepatocytes that dedifferentiate and lose functional properties upon isolation and culturing (110, 127). One factor that likely contributes to the observed functionality and stability of hPSC-HEP in culture is the maturation medium included in the hepatic differentiation kit. Importantly, when the maturation medium is applied, a thin overlay appears to cover the surface of the cells. This type of culturing system recalls the sandwich-cultured hepatocytes that was showed to allow hepatocytes to retain *in vivo* properties in culture (127). However, despite the superiority over the traditional culturing system of hepatocytes, sandwich-cultured hepatocytes start to dedifferentiate after one week in culture (110). In this study, the hPSC-HEP were kept in culture until day 35 after onset of the differentiation before the termination of the experiment, and during this time the expression of hepatocytes markers and transporters was retained as demonstrated in Paper I. Thus, they show stability during at least ten days in culture (35). Herewith, hPSC-HEP as a hepatocyte model may resolve the dedifferentiation issue in *in vitro* applications and provide opportunities for long-term studies of toxicity and pathology assays.

5.2 TRANSCRIPTOME OF HEPATIC DIFFERENTIATION

Transcriptomics of the *in vitro* hepatic differentiation was performed at five developmental stages, starting with hPSC and moving through DE, hepatoblast, early hPSC-HEP, and finally late hPSC-HEP. Human adult liver tissues were used as the control for the benchmarking of the final product. The embryonic development was mimicked in the *in vitro* differentiation of

hepatocytes and confirmed through the investigation of the transitions between the developmental stages. Most of our knowledge about the embryonic development of the liver is acquired from studies on mice. Cytokines and growth factors that guide and promote the development of the liver in the mouse embryo have been shown to encourage the development of different human hepatocyte differentiation protocols (150). Although the development of the liver is considered a conservative process, not all of the knowledge gained from research on mice could be translated for human application (148). Furthermore, considering the complexity of the human physiology, the mechanisms in which cytokines and growth factors control the developmental processes are barely elucidated, but certainly subjected to gradual increase with the continuous development of the omics field.

The transcriptomics results of this study identified sets of genes that are characteristics for the transitions between subsequent developmental stages. These sets of genes included, among others, well-known lineage markers as well as genes with unknown functionality or genes with known functionality that have an obscure contribution to the development process of the specific stage during hepatogenesis. Moreover, genes that were not supposed to be expressed in a specific cell lineage were also identified. These types of discoveries in a time-series developmental study are instrumental for revealing novel mechanisms underlying different developmental stages in hepatogenesis, in addition to mechanisms that lead to deviations from normal development of hepatocytes. Interestingly, a potential DE biomarker, *RP4-559A3.6*, that was highly differentially expressed in hPSC vs. DE and DE vs. hepatoblasts, was identified. *RP4-559A3.6* is a novel transcript that partially overlaps with the *LEFTY2* gene, which is a nodal antagonist that regulates the differentiation process towards the different germ layers (87). *LEFTY2* was also upregulated in DE cells; however, the expression of *RP4-559A3.6* was much higher (34). Moreover, a subset of genes that were abnormally upregulated in hPSC-HEP compared to AL were also identified. Interestingly, these genes were not expressed in hepatoblasts, implying that the activation of these genes occurred upon the application of the maturation or the maintenance medium of the hepatic differentiation kit. Interestingly, two transcription factors that regulated the genes *CDX2* and *KLF5* were also upregulated in hPSC-HEP compared to both AL and hepatoblasts. KLF5 and CDX2 are colon-specific transcription factors, which were also detected in other hPSC-HEP derived by different research groups (32, 36). CDX2 have an essential role in forming and maintaining the trophectoderm in early embryonic development (129). KLF5 is also expressed in stem cells, where it regulates self-renewal and pluripotency (98). Furthermore, many fetal liver genes were still found to be expressed in hPSC-HEP, such as *AFP, FMO1*, and *CYP1B1*.

These findings implicated the failure of hPSC-HEP to turn off genes from earlier developmental stages, which is not unexpected considering the short time that is required to produce hepatocytes *in vitro* compared to hepatocyte development *in vivo* that spans until after birth. In accordance with this statement, a recent research study revealed that *CDX2* was not expressed in hepatocytes produced by transdifferentiation from fibroblasts, while hepatocytes produced from hiPSC reprogrammed from the same batch of fibroblasts applied for transdifferentiation exhibit *CDX2* expression (32). Herewith, the mechanisms that regulate the expression of fetal genes need to be controlled for in the differentiation process, and the time-series transcriptomics study can aid in the identification of genes that may have key roles in these mechanisms.

Pathway and GO enrichment analyses conceptualized the differentiation procedure of the hPSC-HEP starting with the transition from hPSC to DE where pathways important for all three germ lines (endoderm, mesoderm, ectoderm) were present. The transition from DE to hepatoblast showed differential expression of genes involved in the cell cycle and versatile pathways of advanced hepatic functionality, which was unexpected for the hepatoblast stage. Although hepatoblasts express some hepatic markers such as *ALB* and *AFP*, they are not expected to express genes for advanced hepatocytes functionality, such as XME genes (89). The results showed that the hepatoblasts at day 14 in this study expressed many ADME genes in levels comparable to hPSC-HEP and AL. Thus, these cells showed strong commitment to the hepatic lineage, exhibiting early hPSC-HEP phenotype. Furthermore, the transcription profile of the genes involved in the cell cycle showed a decline in gene expression in hepatoblasts compared to hPSC and DE. The expression was in line with that of hPSC-HEP, which also was unexpected since hepatoblasts are supposed to have higher proliferative ability than hPSC-HEP (89). The transition from hepatoblast to hPSC-HEP included genes involved in advanced hepatic functionality. Although many ADME genes showed comparable expression levels between hepatoblast and hPSC-HEP, the core ADME genes, which are of great importance for toxicology and drug development studies, were more highly expressed in hPSC-HEP compared to hepatoblasts. These results confirm that hPSC-HEP were guided towards maturity.

To map the *in vitro* hepatic process to the *in vivo* liver development, the transcriptomics dataset generated from this study was integrated with a dataset including samples from human AL and FL retrieved from the GEO database. The FL used in that study were obtained from fetuses at gestation weeks ranging from 8 to 21. The AL used in this study were obtained from donors

aged 21-38 years old. Interestingly, neither of the *in vitro*-generated samples from the specified developmental stages showed high correlation to FL. On the other hand, the hepatoblasts and hPSC-HEP showed high correlation and similarity to AL. These results show that hPSC-HEP have transcriptional profiles that resemble AL rather than FL, which contradicts previous reports. Human PSC-HEP have generally been considered to have a fetal phenotype because they express *AFP* and have some impaired ADME functionality (36, 114, 150). Although some fetal hepatic genes are still expressed in hPSC-HEP, hPSC-HEP retain many mature hepatic features, as mentioned previously, including the activity of CYP3A4, which is predominant in adult liver (92). Furthermore, differential expression of *CYP3A7*, the fetal version of *CYP3A4* (92), was detected between hPSC-HEP and AL. Surprisingly, it showed downregulation in hPSC-HEP while it is rarely expressed in AL (92), which also advocates for the adult phenotype of hPSC-HEP. A possible explanation for the low correlation of hPSC-HEP to FL may be the fact that fetal liver is the major hematopoietic organ until gestational week 28 (39).

Despite the discovery of the high correlation of hPSC-HEP to AL, profound deviations were detected in gene expression, including low expression for about 22% of ADME genes. Importantly, a functional module including silenced substantial ADME genes (*AOX1, CYP2E1, CYP2C9, CYP2B6, CYP2A6*, *CYP4A11*, and *UGT2B4*) in hPSC-HEP, was identified. Moreover, the regulators of the modules *NR3I1* and *PPARGC1A* were also downregulated in hPSC-HEP. Herewith, several putative targets for improvement of the hepatic differentiation were identified, and the expression of these genes needs to be rectified to achieve higher resemblance to AL.

5.3 METHYLOME OF THE HEPATIC DIFFERENTIATION

An extensive methylome profiling was performed on the same samples that were used to generate the transcriptomics data. A hierarchical clustering showed grouping of samples according to the differentiation stages and remarkably, although no distinction between hESC-derived cells and hiPSC-derived cells was observed at the transcriptional level (33), on the methylome level, biological replicates derived from hiPSC clustered together at the hPSC and DE developmental stages. However, these interesting patterns diminished at the later developmental stages. This finding implies the existence of residual epigenetic memory that accompanies reprogrammed cells until it vanishes upon extended culturing (66). Furthermore, the results signify the elasticity and the pioneering properties of the transcriptional machinery to adapt to the

reprogramming process and control the phenotype of the resulting product, compared to DNA methylation which requires longer time to change. In this thesis, only parts of the results from the methylome study are presented, and further extensive data analysis of the methylome data is still ongoing.

5.4 INTEGRATION OF TRANSCRIPTOME AND METHYLATION RESULTS OF HPSC-HEP

The human genome is regulated at several levels, including both genetics and epigenetics. The ideal study would investigate molecules of different types related to the explored biological system to achieve better interpretation of the results and to uncover important interactions. This study applied both transcriptome and methylome techniques to explore the similarities and differences between hPSC-HEP and AL. Results from both analyses demonstrated important deviations of hPSC-HEP from AL, and the integration of the transcriptome and the methylome was performed aiming to clarify the mechanisms underlying these deviations.

Remarkably, 72% of the genes that showed deviations on the transcriptional level between hPSC-HEP and AL showed deviation in DNA methylation of all or some of the probes annotated to these genes. Furthermore, a significant correlation between highly methylated genes in hPSC-HEP compared to AL, to downregulated genes in hPSC-HEP compared to AL, and vice versa, were detected. The correlation was performed on the median of delta-beta of all differentially methylated probes annotated for the promoter and intergenic regions of genes. Probes annotated for the body of genes were excluded, as these are primarily associated with gene activation (62). Intergenic regions often contain promoters for non-coding RNA regions, enhancers and other regulatory motifs (96). Herewith, the results obtained from the correlation analysis are in accordance with other studies showing that methylation of regulatory motifs results frequently in the suppression of the regulated gene (24).

Investigation of the methylation status of the genes (*AOX1, CYP2E1, CYP2C9, CYP2B6, CYP2A6, CYP4A11*, and *UGT2B4*) included in the functional module that was identified in Paper II and further explored in Paper III, revealed hypermethylation of all the genes in this module, in addition to several transcription factors regulating these genes. Therefore, further regulators identified from the literature were explored and also these showed deviating methylation patterns in our data. Moreover, differentially methylated regions were identified for *CYP2E1, AR, PPARGC1, IGF1R*, and *CDX2*. A protein

interaction network containing the module and the different regulators is proposed (Figure 3) and may serve as a basis for selecting factors to be manipulated in order to rectify these abnormalities and produce hPSC-HEP with substantially improved functionality. One strategy would be to induce *PPARGC1* while repressing both *PITX2* and *CTNNB1* (β-catenin), applying chemical or/and genetic manipulations to promote upregulation of *NR1I3, AR, CYP2E1, CYP2C9, CYP2B6, CYP4A11,* and *CYP2A6*. Studies have shown that ectopic expression of transcription factors in somatic cells leads to a successful demethylation of the corresponding genomic genes in both reprogramming and transdifferentiation procedures (52). These observations suggest that epigenetic modifications could be reversed by overexpression of transcription factors targeting the epigenetically inhibited genes. This property could be utilized to rectify abnormal DNA methylations.

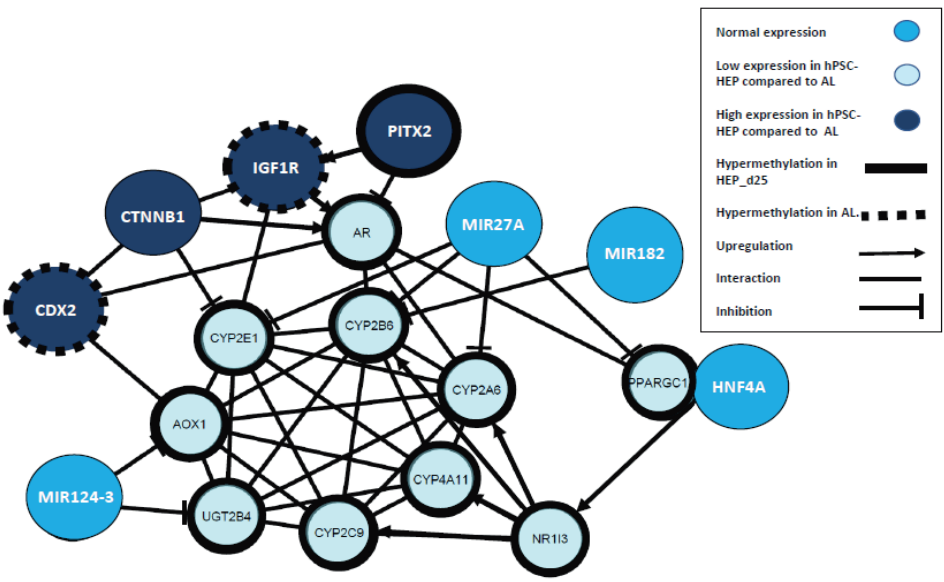

Figure 3. Protein interaction network proposed for the identified functional module of crucial drug-metabolizing enzymes CYP2E1, CYP2C9, CYP2B6, CYP2A6, CYP4A11, AOX1, and UGT2B4.

5.5 LIMITATIONS OF CURRENT STUDY

The study was designed to avoid confounds that may interfere with the obtained results. For example, all the biological replicates were of XY genotype to eliminate gender confounds, and hPSC at relatively low passages were used to avoid the risk of accumulation of abnormal methylation. Moreover, one standardized differentiation protocol was applied to all hPSC, and the cell sampling for the different developmental stages occurred at identical time points during the differentiation for all the biological replicates. Nevertheless, several limitations should be noted, which may also influence the interpretation of the results. The control for *in vivo* hepatocytes was AL, and although the majority of the cells in AL are hepatocytes which also occupy about 80% of the liver volume, other cell types exist in the liver and this was not considered in this work. To achieve accurate results, hPSC-HEP should have been compared to an *in vivo* counterpart without any signs of dedifferentiation or loss of functionality, but such cells are hard to find. The most suitable cell to be used as a control would be freshly isolated primary hepatocytes. However, the isolation procedure would also affect their performance.

Analyzing hPSC-HEP produced by only one differentiation protocol is also a limitation, since conclusions from this study can only be applied to cells differentiated with this specific protocol. Including several protocols in the study would have provided more general conclusions.

The conclusion that hPSC-HEP were highly correlated to AL rather than to FL was based on the comparison with FL form gestation weeks 8-21, which includes hematopoietic cells that may interfere with the results. Therefore, comparison with FL from later gestational weeks and neonatal liver tissue may provide more accurate results. However, access to such samples is limited due to ethical reasons and were not available for this study.

Further, samples from additional developmental stages would have strengthened the study, and data from proteomics, metabolomics and miRNA analysis would provide additional insights and a comprehensive understanding of the hepatic differentiation and deviations.

6 CONCLUSION

In this thesis, a comprehensive transcriptome and methylome analysis was conducted on different developmental stages during *in vitro* hepatic differentiation to follow the differentiation *in vitro,* identify deviations from *in vivo* counterparts, and search for solutions to rectify these deviations.

The differentiation process reflected *in vivo* hepatogenesis and could be used to study hepatocyte development and gain new insights into the mechanisms regulating this process in humans. However, methylation and transcriptional analysis identified profound deviations from AL that needs to be rectified for improved functionality of the hPSC-HEP. Interestingly, hPSC-HEP exhibited higher correlation to AL than to FL, which is a step forward compared to previous investigations performed by several research groups.

Finally, the targets identified in this study could be used to improve the differentiation process in order to achieve functional *in vitro* derived hepatocytes that are suitable for advanced applications in regenerative medicine, cell therapy, and for drug discovery and development.

7 FUTURE PERSPECTIVES

Hepatocyte-like cells generated from different cell types and protocols have already proven to acquire enhanced functionality upon transplantation in mice. Reports indicate that hPSC-HEP is a promising approach that will be useful in versatile advanced application in different fields, including regenerative medicine, cell therapy, toxicology, drug discovery, and development.

This and other studies have demonstrated that the differentiation of hPSC-HEP has improved significantly the last few years and hepatocytes with enhanced functionality can now be produced *in vitro*. However, to enable advanced applications of hPSC-HEP, several issues need to be rectified. Moreover, safety and efficiency assessments must be approved to ensure that no hazard could be introduced by the application of hPSC-HEP.

Another challenge that needs to be addressed is the large amount of cells that is required for performing cell therapies. About 200 grams of hepatocytes is required for each treatment with BAL for bridging ESLD patients to transplantation. Although hPSC-HEP can provide unlimited cell supply, new methods are needed to scale up the production of hPSC-HEP to meet the immense need. This may require 3D suspension culturing of hPSC-HEP to produce large amounts of cells.

Finally, continuous investigation of the methylome of the hepatic differentiation is ongoing, and these results should be compared to the transcriptome results and utilized for further development of novel strategies to improve current hepatic differentiation protocols. Considering the enormous advancements in stem cell technology over recent years, the aforementioned suggestions are likely achievable.

ACKNOWLEDGEMENT

This work was supported by Takara Bio Europe AB (Gothenburg, Sweden) and the University of Skövde, Sweden, under grants from the Knowledge Foundation [2012/0310], [2013/89] and [2014/0301].

I would like also to thank all the wonderful people that contributed to this achievement.

Thanks to my Supervisor **Jane Synnergren** for making this project possible, for accepting me in your extraordinary team, for all the support, help and the valuable time that you dedicated for me throughout the years.

Special thanks to my supervisors **Stina Simonsson** and **Helena Carén** for all the guidance, support, help and knowledge sharing during my doctoral studies.

Many thanks to all the co-authors that contributed to this work; **Barbara Küppers-Munther, Josefina Edsbagge, Annika Asplund and Mariska Van Giezen** for sharing your expertise in hPSC-HEP and DE cells and for all the help and support, **Björn Olsson** and **Benjamin Ulfenborg** for sharing your expertise in statistics and data analysis and for all the appreciated discussions, **Peter Sartipy** and **Petter Björquist** for all the knowledge sharing, guidance, support and great help, **Christian Andersson** for your appreciated cooperation and support, and **Tommy Andersson** for sharing your expertise in hepatocytes.

Special thanks for all the colleagues at the **Takara Bio Europe**, the **Cellartis** team and the **university of Skövde** for all the great help, support, kindness and warm atmosphere.

I would like also to thank my family; my mother **Nadia** and father **Wael** for their unconditional love and endless support, my beautiful and wonderful sister **Ousaifa** and my beloved brothers **Jacob** and **Josef** and their great families for all the love and being there for me always.

Thank you to my adorable son **Ronnie** for giving me the opportunity to be a mother of an incredible boy.

Finally, thank you to my adorable husband **Ali**, I am grateful to having you in my life, for your endless love, support and being there for better or worse.

REFERENCES

1. **Alison M, and Lin W**. *Regenerating the liver: not so simple after all? [version 1; referees: 3 approved]*. 2016.
2. **Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, and Irizarry RA**. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30: 1363-1369, 2014.
3. **Asgari S, Moslem M, Bagheri-Lankarani K, Pournasr B, Miryounesi M, and Baharvand H**. Differentiation and transplantation of human induced pluripotent stem cell-derived hepatocyte-like cells. *Stem cell reviews* 9: 493-504, 2013.
4. **Ashburner M**. Gene Ontology: tool for the unification of biology. 25: 25-29, 2000.
5. **Asplund A, Pradip A, van Giezen M, Aspegren A, Choukair H, Rehnstrom M, Jacobsson S, Ghosheh N, El Hajjam D, Holmgren S, Larsson S, Benecke J, Butron M, Wigander A, Noaksson K, Sartipy P, Bjorquist P, Edsbagge J, and Kuppers-Munther B**. One Standardized Differentiation Procedure Robustly Generates Homogenous Hepatocyte Cultures Displaying Metabolic Diversity from a Large Panel of Human Pluripotent Stem Cells. *Stem cell reviews* 12: 90-104, 2016.
6. **Asrani SK, Larson JJ, Yawn B, Therneau TM, and Kim WR**. Underestimation of Liver-Related Mortality in the United States. *Gastroenterology* 145: 375-382.e372, 2013.
7. **Bader GD, and Hogue CW**. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2, 2003.
8. **Bataller R, and Brenner DA**. Liver fibrosis. *Journal of Clinical Investigation* 115: 209-218, 2005.
9. **Bernal W, and Wendon J**. Acute Liver Failure. *New England Journal of Medicine* 369: 2525-2534, 2013.
10. **Bibikova M, Laurent LC, Ren B, Loring JF, and Fan JB**. Unraveling epigenetic regulation in embryonic stem cells. *Cell Stem Cell* 2: 123-134, 2008.
11. **Bolstad BM, Irizarry RA, Astrand M, and Speed TP**. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185-193, 2003.
12. **Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, Buurman WA, Deelen P, Greve JW, Ivanov M, Rensen SS, van Vliet-Ostaptchouk JV, Wolfs MG, Fu J, Hofker MH, Wijmenga C, Zhernakova A, Ingelman-Sundberg M, Franke L, and Milani L**. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics* 15: 860, 2014.

13. **Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, and Balasubramanian S**. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protocols* 8: 1841-1851, 2013.
14. **Brolen G, Sivertsson L, Bjorquist P, Eriksson G, Ek M, Semb H, Johansson I, Andersson TB, Ingelman-Sundberg M, and Heins N**. Hepatocyte-like cells derived from human embryonic stem cells specifically via definitive endoderm and a progenitor stage. *J Biotechnol* 145: 284-294, 2010.
15. **Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E, Oyolu CB, Schroth GP, Absher DM, Baker JC, and Myers RM**. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* 19: 1044-1056, 2009.
16. **Burry RW**. *Immunocytochemistry: A Practical Guide for Biomedical Research*. Springer, New York, NY, 2010.
17. **Byass P**. The global burden of liver disease: a challenge for methods and for public health. *BMC Med* 12: 2014.
18. **Camp JG, Sekine K, Gerber T, Loeffler-Wirth H, Binder H, Gac M, Kanton S, Kageyama J, Damm G, Seehofer D, Belicova L, Bickle M, Barsacchi R, Okuda R, Yoshizawa E, Kimura M, Ayabe H, Taniguchi H, Takebe T, and Treutlein B**. Multilineage communication regulates human liver bud development from pluripotency. *Nature* 546: 533, 2017.
19. **Cedar H, and Bergman Y**. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* 10: 295, 2009.
20. **Chen CC, Wang KY, and Shen CK**. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *J Biol Chem* 288: 9084-9091, 2013.
21. **Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, and Ma'ayan A**. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14: 128, 2013.
22. **Cheng Y, Xie N, Jin P, and Wang T**. DNA methylation and hydroxymethylation in stem cells. *Cell biochemistry and function* 33: 161-173, 2015.
23. **Crispe IN**. Immune tolerance in liver disease. *Hepatology* 60: 2109-2117, 2014.
24. **Deaton AM, and Bird A**. CpG islands and the regulation of transcription. *Genes Dev* 25: 1010-1022, 2011.
25. **Dhawan A, Puppi J, Hughes RD, and Mitry RR**. Human hepatocyte transplantation: current experience and future challenges. *Nature Reviews Gastroenterology &Amp; Hepatology* 7: 288, 2010.
26. **Dulak J, Szade K, Szade A, Nowak W, and Jozkowicz A**. Adult stem cells: hopes and hypes of regenerative medicine. *Acta biochimica Polonica* 62: 329-337, 2015.

27. **Edwards IR, and Aronson JK**. Adverse drug reactions: definitions, diagnosis, and management. *Lancet (London, England)* 356: 1255-1259, 2000.
28. **Ekman D, Light S, Bjorklund AK, and Elofsson A**. What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? *Genome biology* 7: R45, 2006.
29. **Ericka A. Pestana SB, Adama Diallo,, and John R. Crowther GJV**. Traditional PCR. In: *Real-Time PCR – The Basic Principles In: Early, rapid and sensitive veterinary molecular diagnostics - real time PCR applications*Springer Dordrecht, 2010.
30. **Friedman SL**. Liver fibrosis – from bench to bedside. *Journal of Hepatology* 38: 38-53, 2003.
31. **Fu Y, and He C**. Nucleic Acid Modifications with Epigenetic Significance. *Current opinion in chemical biology* 16: 516-524, 2012.
32. **Gao Y, Zhang X, Zhang L, Cen J, Ni X, Liao X, Yang C, Li Y, Chen X, Zhang Z, Shu Y, Cheng X, Hay DC, Lai D, Pan G, Wei G, and Hui L**. Distinct Gene Expression and Epigenetic Signatures in Hepatocyte-like Cells Produced by Different Strategies from the Same Donor. *Stem Cell Reports* 2017.
33. **Gebhardt R**. Metabolic zonation of the liver: Regulation and implications for liver function. *Pharmacology & Therapeutics* 53: 275-354, 1992.
34. **Ghosheh N, Kuppers-Munther B, Asplund A, Edsbagge J, Ulfenborg B, Andersson TB, Bjorquist P, Andersson CX, Caren H, Simonsson S, Sartipy P, and Synnergren J**. Comparative transcriptomics of hepatic differentiation of human pluripotent stem cells and adult human liver tissue. *Physiological genomics* physiolgenomics.00007.02017, 2017.
35. **Ghosheh N, Olsson B, Edsbagge J, Kuppers-Munther B, Van Giezen M, Asplund A, Andersson TB, Bjorquist P, Caren H, Simonsson S, Sartipy P, and Synnergren J**. Highly Synchronized Expression of Lineage-Specific Genes during In Vitro Hepatic Differentiation of Human Pluripotent Stem Cell Lines. *Stem Cells Int* 2016: 8648356, 2016.
36. **Godoy P, Schmidt-Heck W, Natarajan K, Lucendo-Villarin B, Szkolnicka D, Asplund A, Bjorquist P, Widera A, Stober R, Campos G, Hammad S, Sachinidis A, Chaudhari U, Damm G, Weiss TS, Nussler A, Synnergren J, Edlund K, Kuppers-Munther B, Hay DC, and Hengstler JG**. Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. *J Hepatol* 63: 934-942, 2015.
37. **Goldring Chris EP, Duffy Paul A, Benvenisty N, Andrews Peter W, Ben-David U, Eakins R, French N, Hanley Neil A, Kelly L, Kitteringham Neil R, Kurth J, Ladenheim D, Laverty H, McBlane J, Narayanan G, Patel S, Reinhardt J, Rossi A, Sharpe M, and Park BK**. Assessing the Safety of Stem Cell Therapeutics. *Cell Stem Cell* 8: 618-628.
38. **Gordillo M, Evans T, and Gouon-Evans V**. Orchestrating liver development. *Development* 142: 2094-2108, 2015.

39. **Gruppuso PA, and Sanders JA**. Regulation of liver development: implications for liver biology across the lifespan. *Journal of molecular endocrinology* 56: R115-125, 2016.
40. **Gu J, Stevens M, Xing X, Li D, Zhang B, Payton JE, Oltz EM, Jarvis JN, Jiang K, Cicero T, Costello JF, and Wang T**. Mapping of Variable DNA Methylation Across Multiple Cell Types Defines a Dynamic Regulatory Landscape of the Human Genome. *G3: Genes|Genomes|Genetics* 6: 973-986, 2016.
41. **Guguen-Guillouzo C, and Guillouzo A**. General review on in vitro hepatocyte models and their applications. *Methods Mol Biol* 640: 1-40, 2010.
42. **Gupta SK, Bhandari B, Shrestha A, Biswal BK, Palaniappan C, Malhotra SS, and Gupta N**. Mammalian zona pellucida glycoproteins: structure and function during fertilization. *Cell and tissue research* 349: 665-678, 2012.
43. **Han H, Cortez CC, Yang X, Nichols PW, Jones PA, and Liang G**. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* 20: 4299-4310, 2011.
44. **Hannoun Z, Steichen C, Dianat N, Weber A, and Dubart-Kupperschmitt A**. The potential of induced pluripotent stem cell derived hepatocytes. *J Hepatol* 65: 182-199, 2016.
45. **Harrington CA, Rosenow C, and Retief J**. Monitoring gene expression using DNA microarrays. *Current opinion in microbiology* 3: 285-291, 2000.
46. **Hay DC, Pernagallo S, Diaz-Mochon JJ, Medine CN, Greenhough S, Hannoun Z, Schrader J, Black JR, Fletcher J, Dalgetty D, Thompson AI, Newsome PN, Forbes SJ, Ross JA, Bradley M, and Iredale JP**. Unbiased screening of polymer libraries to define novel substrates for functional hepatocytes with inducible drug metabolism. *Stem Cell Res* 6: 92-102, 2011.
47. **Hayatsu H**. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — A personal account. *Proceedings of the Japan Academy Series B, Physical and Biological Sciences* 84: 321-330, 2008.
48. **Heins N, Englund MC, Sjoblom C, Dahl U, Tonning A, Bergh C, Lindahl A, Hanson C, and Semb H**. Derivation, characterization, and differentiation of human embryonic stem cells. *Stem Cells* 22: 367-376, 2004.
49. **Hengstler JG, Brulport M, Schormann W, Bauer A, Hermes M, Nussler AK, Fandrich F, Ruhnke M, Ungefroren H, Griffin L, Bockamp E, Oesch F, and von Mach M-A**. Generation of human hepatocytes by stem cell technology: definition of the hepatocyte. *Expert Opinion on Drug Metabolism & Toxicology* 1: 61-74, 2005.

50. **Hewitt KJ, Shamis Y, Hayman RB, Margvelashvili M, Dong S, Carlson MW, and Garlick JA**. Epigenetic and phenotypic profile of fibroblasts derived from induced pluripotent stem cells. *PLoS One* 6: e17128, 2011.
51. **Hirschi KK, Li S, and Roy K**. Induced pluripotent stem cells for regenerative medicine. *Annu Rev Biomed Eng* 16: 277-294, 2014.
52. **Holmberg J, and Perlmann T**. Maintaining differentiated cellular identity. *Nature reviews Genetics* 13: 429-439, 2012.
53. **Holmgren G, Sjogren AK, Barragan I, Sabirsh A, Sartipy P, Synnergren J, Bjorquist P, Ingelman-Sundberg M, Andersson TB, and Edsbagge J**. Long-term chronic toxicity testing using human pluripotent stem cell-derived hepatocytes. *Drug Metab Dispos* 42: 1401-1406, 2014.
54. **Huang K, and Fan G**. DNA methylation in cell differentiation and reprogramming: an emerging systematic view. *Regen Med* 5: 531-544, 2010.
55. **Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP**. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)* 4: 249-264, 2003.
56. **Ivanov M, Kacevska M, and Ingelman-Sundberg M**. Epigenomics and interindividual differences in drug response. *Clinical pharmacology and therapeutics* 92: 727-736, 2012.
57. **Ivanov M, Kals M, Lauschke V, Barragan I, Ewels P, Kaller M, Axelsson T, Lehtio J, Milani L, and Ingelman-Sundberg M**. Single base resolution analysis of 5-hydroxymethylcytosine in 188 human genes: implications for hepatic gene expression. *Nucleic Acids Res* 44: 6756-6769, 2016.
58. **Jang HS, Shin WJ, Lee JE, and Do JT**. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes* 8: 2017.
59. **Jenne CN, and Kubes P**. Immune surveillance by the liver. *Nature immunology* 14: 996-1006, 2013.
60. **Jensen J, Hyllner J, and Bjorquist P**. Human embryonic stem cell technologies and drug discovery. *J Cell Physiol* 219: 513-519, 2009.
61. **Johnson WE, Li C, and Rabinovic A**. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)* 8: 118-127, 2007.
62. **Jones PA**. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics* 13: 484-492, 2012.
63. **Jungermann K, and Katz N**. Functional hepatocellular heterogeneity. *Hepatology* 2: 385-395, 1982.
64. **Jungermann K, and Kietzmann T**. Oxygen: Modulator of metabolic zonation and disease of the liver. *Hepatology* 31: 255-260, 2000.
65. **Kia R, Sison RL, Heslop J, Kitteringham NR, Hanley N, Mills JS, Park BK, and Goldring CE**. Stem cell-derived hepatocytes as a

predictive model for drug-induced liver injury: are we there yet? *Br J Clin Pharmacol* 75: 885-896, 2013.
66. **Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LIR, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, and Daley GQ**. Epigenetic memory in induced pluripotent stem cells. *Nature* 467: 285-290, 2010.
67. **Kitade M, Kaji K, and Yoshiji H**. Relationship between hepatic progenitor cell-mediated liver regeneration and non-parenchymal cells. *Hepatol Res* 46: 1187-1193, 2016.
68. **Kogiso T, Nagahara H, Otsuka M, Shiratori K, and Dowdy SF**. Transdifferentiation of human fibroblasts into hepatocyte-like cells by defined transcriptional factors. *Hepatology international* 7: 937-944, 2013.
69. **Kolios G, Valatas V, and Kouroumalis E**. Role of Kupffer cells in the pathogenesis of liver disease. *World J Gastroenterol* 12: 7413-7420, 2006.
70. **Kotlyar M, Fuhrman S, Ableson A, and Somogyi R**. Spearman correlation identifies statistically significant gene expression clusters in spinal cord development and injury. *Neurochem Res* 27: 1133-1140, 2002.
71. **Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, and Ma'ayan A**. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44: W90-97, 2016.
72. **Kung JW, Currie IS, Forbes SJ, and Ross JA**. Liver development, regeneration, and carcinogenesis. *Journal of biomedicine & biotechnology* 2010: 984248, 2010.
73. **Lee JS, Ward WO, Liu J, Ren H, Vallanat B, Delker D, and Corton JC**. Hepatic Xenobiotic Metabolizing Enzyme and Transporter Gene Expression through the Life Stages of the Mouse. *PLoS One* 6: 2011.
74. **Lemaigre F**. Markers and signaling factors for stem cell differentiation to hepatocytes: lessons from developmental studies. *Methods Mol Biol* 640: 157-166, 2010.
75. **Li F, and Tian Z**. The liver works as a school to educate regulatory immune cells. *Cellular and Molecular Immunology* 10: 292-302, 2013.
76. **Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, and Schubeler D**. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43: 1091-1097, 2011.
77. **Liu H, Li S, Wang X, Zhu J, Wei Y, Wang Y, Wen Y, Wang L, Huang Y, Zhang B, Shang S, and Zhang Y**. DNA methylation dynamics: identification and functional annotation. *Briefings in Functional Genomics* 15: 470-484, 2016.

78. **Liu J, Hua P, Hui L, Zhang LL, Hu Z, and Zhu YW**. Identification of hub genes and pathways associated with hepatocellular carcinoma based on network strategy. *Experimental and therapeutic medicine* 12: 2109-2119, 2016.
79. **Liu L, Wei J, and Ruan J**. Pathway Enrichment Analysis with Networks. *Genes* 8: 2017.
80. **Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, McGlinchey RE, Milberg W, and Miller MW**. The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics* 9: 1363-1371, 2017.
81. **Ludwig G, Nejman D, Hecht M, Orlanski S, Abu-Remaileh M, Yanuka O, Sandler O, Marx A, Roberts D, Benvenisty N, Bergman Y, Mendelsohn M, and Cedar H**. Aberrant DNA methylation in ES cells. *PLoS One* 9: e96090, 2014.
82. **Lönnstedt I, and Speed T**. REPLICATED MICROARRAY DATA. *Statistica Sinica* 12: 31-46, 2002.
83. **Magnusson M, Larsson P, Lu EX, Bergh N, Caren H, and Jern S**. Rapid and specific hypomethylation of enhancers in endothelial cells during adaptation to cell culturing. *Epigenetics* 11: 614-624, 2016.
84. **Maksimovic J, Phipson B, and Oshlack A**. A cross-package Bioconductor workflow for analysing methylation array data. *F1000Research* 5: 1281, 2016.
85. **Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M, Sundberg CJ, Ekström TJ, Teschendorff AE, Tegnér J, and Gomez-Cabrero D**. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* 8: 333-346, 2013.
86. **Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, and Lander ES**. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454: 766-770, 2008.
87. **Meno C, Gritsman K, Ohishi S, Ohfuji Y, Heckscher E, Mochida K, Shimono A, Kondoh H, Talbot WS, Robertson EJ, Schier AF, and Hamada H**. Mouse Lefty2 and Zebrafish Antivin Are Feedback Inhibitors of Nodal Signaling during Vertebrate Gastrulation. *Molecular cell* 4: 287-298, 1999.
88. **Messerschmidt DM, Knowles BB, and Solter D**. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* 28: 812-828, 2014.
89. **Miyajima A, Tanaka M, and Itoh T**. Stem/Progenitor Cells in Liver Development, Homeostasis, Regeneration, and Reprogramming. *Cell Stem Cell* 14: 561-574.
90. **Mokdad AA, Lopez AD, Shahraz S, Lozano R, Mokdad AH, Stanaway J, Murray CJL, and Naghavi M**. Liver cirrhosis mortality in 187

countries between 1980 and 2010: a systematic analysis. *BMC Medicine* 12: 145, 2014.
91. **Montagnani S, Rueger MA, Hosoda T, and Nurzynska D**. Adult Stem Cells in Tissue Maintenance and Regeneration. *Stem Cells Int* 2016: 7362879, 2016.
92. **Moscovitz JE, and Aleksunes LM**. Establishment of metabolism and transport pathways in the rodent and human fetal liver. *International journal of molecular sciences* 14: 23801-23827, 2013.
93. **Mukaka M**. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi* 24: 69-71, 2012.
94. **Nachar N**. The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology* 4: 13-20, 2008.
95. **Nakatani K, Kaneda K, Seki S, and Nakajima Y**. Pit cells as liver-associated natural killer cells: morphology and function. *Medical electron microscopy : official journal of the Clinical Electron Microscopy Society of Japan* 37: 29-36, 2004.
96. **Nelson CE, Hersh BM, and Carroll SB**. The regulatory content of intergenic DNA shapes genome architecture. *Genome biology* 5: R25, 2004.
97. **Pang-Ning Tan MS, Vipin Kumar**. Cluster Analysis: Basic Concepts and Algorithms. In: *Introduction to Data Mining*. Boston: Addison-Wesley Longman Publishing Co., Inc., 2005.
98. **Parisi S, and Russo T**. Regulatory role of Klf5 in early mouse development and in embryonic stem cells. *Vitamins and hormones* 87: 381-397, 2011.
99. **Patil V, Ward RL, and Hesson LB**. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* 9: 823-828, 2014.
100. **Pestana E.A. BS, Diallo A., Crowther J.R., Viljoen G.J.** . Real-Time PCR – The Basic Principles. In: *Early, rapid and sensitive veterinary molecular diagnostics - real time PCR applications*Springer, Dordrecht, 2010.
101. **Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, Clark SJ, and Molloy PL**. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 8: 2015.
102. **Pfeiffer E, Kegel V, Zeilinger K, Hengstler JG, Nussler AK, Seehofer D, and Damm G**. Featured Article: Isolation, characterization, and cultivation of human hepatocytes and non-parenchymal liver cells. *Experimental biology and medicine (Maywood, NJ)* 240: 645-656, 2015.
103. **Phipson B, Lee S, Majewski IJ, Alexander WS, and Smyth GK**. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO

DETECT DIFFERENTIAL EXPRESSION. *The annals of applied statistics* 10: 946-963, 2016.

104. **Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, and Clark SJ**. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome biology* 17: 208, 2016.

105. **Pinney SE**. Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology* 3: 739-751, 2014.

106. **Pitt JJ**. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews* 30: 19-34, 2009.

107. **Racanelli V, and Rehermann B**. The liver as an immunological organ. *Hepatology* 43: S54-62, 2006.

108. **Raven A, Lu WY, Man TY, Ferreira-Gonzalez S, O'Duibhir E, Dwyer BJ, Thomson JP, Meehan RR, Bogorad R, Koteliansky V, Kotelevtsev Y, Ffrench-Constant C, Boulter L, and Forbes SJ**. Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. *Nature* 547: 350-354, 2017.

109. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK**. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015.

110. **Rowe C, Gerrard DT, Jenkins R, Berry A, Durkin K, Sundstrom L, Goldring CE, Park BK, Kitteringham NR, Hanley KP, and Hanley NA**. Proteome-wide analyses of human hepatocytes during differentiation and dedifferentiation. *Hepatology* 58: 799-809, 2013.

111. **Rumman M, Dhawan J, and Kassem M**. Concise Review: Quiescence in Adult Stem Cells: Biological Significance and Relevance to Tissue Regeneration. *Stem Cells* 33: 2903-2912, 2015.

112. **Saksouk N, Simboeck E, and Déjardin J**. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin* 8: 3, 2015.

113. **Sartipy P, and Bjorquist P**. Concise review: Human pluripotent stem cell-based models for cardiac and hepatic toxicity assessment. *Stem Cells* 29: 744-748, 2011.

114. **Schwartz RE, Fleming HE, Khetani SR, and Bhatia SN**. Pluripotent stem cell-derived hepatocyte-like cells. *Biotechnol Adv* 32: 504-513, 2014.

115. **Seita J, and Weissman IL**. Hematopoietic Stem Cell: Self-renewal versus Differentiation. *Wiley interdisciplinary reviews Systems biology and medicine* 2: 640-653, 2010.

116. **Serin EA, Nijveen H, Hilhorst HW, and Ligterink W**. Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci* 7: 444, 2016.

117. **Shanks N, Greek R, and Greek J**. Are animal models predictive for humans? *Philosophy, Ethics, and Humanities in Medicine : PEHM* 4: 2, 2009.
118. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T**. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
119. **Shen L, and Zhang Y**. 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Current opinion in cell biology* 25: 289-296, 2013.
120. **Shen Y, Chow J, Wang Z, and Fan G**. Abnormal CpG island methylation occurs during in vitro differentiation of human embryonic stem cells. *Hum Mol Genet* 15: 2623-2635, 2006.
121. **Si-Tayeb K, Lemaigre FP, and Duncan SA**. Organogenesis and Development of the Liver. *Developmental Cell* 18: 175-189.
122. **Smyth GK**. Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, edited by Gentleman R, Carey V, Dudoit S, Irizarry R, and Huber W. New York: Springer, 2005.
123. **Smyth GK**. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 3: Article3, 2004.
124. **Solter D, and Knowles BB**. Immunosurgery of mouse blastocyst. *Proceedings of the National Academy of Sciences of the United States of America* 72: 5099-5102, 1975.
125. **Spirin V, and Mirny LA**. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA* 100: 2003.
126. **Steuer R, Kurths J, Daub CO, Weise J, and Selbig J**. The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* 18 Suppl 2: S231-240, 2002.
127. **Swift B, Pfeifer ND, and Brouwer KLR**. Sandwich-Cultured Hepatocytes: An In Vitro Model to Evaluate Hepatobiliary Transporter-Based Drug Interactions and Hepatotoxicity. *Drug metabolism reviews* 42: 446-471, 2010.
128. **Szkolnicka D, Zhou W, Lucendo-Villarin B, and Hay DC**. Pluripotent stem cell-derived hepatocytes: potential and challenges in pharmacology. *Annu Rev Pharmacol Toxicol* 53: 147-159, 2013.
129. **Takakura Y, Hinoi T, Oue N, Sasada T, Kawaguchi Y, Okajima M, Akyol A, Fearon ER, Yasui W, and Ohdan H**. CDX2 regulates Multidrug Resistance 1 gene expression in malignant intestinal epithelium. *Cancer Res* 70: 6767-6778, 2010.
130. **Tan AKY, Loh KM, and Ang LT**. Evaluating the regenerative potential and functionality of human liver cells in mice. *Differentiation* 98: 25-34, 2017.

131. **Tan L, and Shi YG**. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* 139: 1895-1902, 2012.
132. **Tang Y, Yu P, and Cheng L**. Current progress in the derivation and therapeutic application of neural stem cells. *Cell death & disease* 8: e3108, 2017.
133. **Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, and Teschendorff AE**. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* btx513-btx513, 2017.
134. **Tognini P, Napoli D, and Pizzorusso T**. Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity. *Frontiers in Cellular Neuroscience* 9: 2015.
135. **Tu T, Calabro SR, Lee A, Maczurek AE, Budzinska MA, Warner FJ, McLennan SV, and Shackel NA**. Hepatocytes in liver injury: Victim, bystander, or accomplice in progressive fibrosis? *J Gastroenterol Hepatol* 30: 1696-1704, 2015.
136. **Turnbull AK, Kitchen RR, Larionov AA, Renshaw L, Dixon JM, and Sims AH**. Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. *BMC Medical Genomics* 5: 35, 2012.
137. **Tusher VG, Tibshirani R, and Chu G**. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 98: 5116-5121, 2001.
138. **Upadhyay M, Samal J, Kandpal M, Vasaikar S, Biswas B, Gomes J, and Vivekanandan P**. CpG Dinucleotide Frequencies Reveal the Role of Host Methylation Capabilities in Parvovirus Evolution. *Journal of Virology* 87: 13816-13824, 2013.
139. **Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, and Raval A**. Identifying Hubs in Protein Interaction Networks. *PLoS ONE* 4: 2009.
140. **Vazin T, and Freed WJ**. Human embryonic stem cells: Derivation, culture, and differentiation: A review. *Restorative neurology and neuroscience* 28: 589-603, 2010.
141. **Weiskirchen R, and Tacke F**. Cellular and molecular functions of hepatic stellate cells in inflammatory responses and liver immunology. *Hepatobiliary surgery and nutrition* 3: 344-363, 2014.
142. **Wen L, and Tang F**. Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics* 104: 341-346, 2014.
143. **Xu C, Li CY, and Kong AN**. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of pharmacal research* 28: 249-268, 2005.
144. **Xu H, Yi BA, Wu H, Bock C, Gu H, Lui KO, Park JH, Shao Y, Riley AK, Domian IJ, Hu E, Willette R, Lepore J, Meissner A, Wang**

Z, and Chien KR. Highly efficient derivation of ventricular cardiomyocytes from induced pluripotent stem cells with a distinct epigenetic signature. *Cell Res* 22: 142-154, 2012.
145. **Yanger K, Knigin D, Zong Y, Maggs L, Gu G, Akiyama H, Pikarsky E, and Stanger BZ**. Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. *Cell Stem Cell* 15: 340-349, 2014.
146. **Yildirimman R, Brolen G, Vilardell M, Eriksson G, Synnergren J, Gmuender H, Kamburov A, Ingelman-Sundberg M, Castell J, Lahoz A, Kleinjans J, van Delft J, Bjorquist P, and Herwig R**. Human embryonic stem cell derived hepatocyte-like cells as a tool for in vitro hazard assessment of chemical carcinogenicity. *Toxicol Sci* 124: 278-290, 2011.
147. **Yu M, Hon Gary C, Szulwach Keith E, Song C-X, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min J-H, Jin P, Ren B, and He C**. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* 149: 1368-1380.
148. **Yu Y, Wang X, and Nyberg SL**. Potential and Challenges of Induced Pluripotent Stem Cells in Liver Diseases Treatment. *J Clin Med* 3: 997-1017, 2014.
149. **Zhang Z, Liu J, Liu Y, Li Z, Gao WQ, and He Z**. Generation, characterization and potential therapeutic applications of mature and functional hepatocytes from stem cells. *J Cell Physiol* 228: 298-305, 2013.
150. **Zhao D, Chen S, Duo S, Xiang C, Jia J, Guo M, Lai W, Lu S, and Deng H**. Promotion of the efficient metabolic maturation of human pluripotent stem cell-derived hepatocytes by correcting specification defects. *Cell Res* 23: 157-161, 2013.
151. **Zhao R, and Duncan SA**. Embryonic development of the liver. *Hepatology* 41: 956-967, 2005.
152. **Zhao W, Ji X, Zhang F, Li L, and Ma L**. Embryonic stem cell markers. *Molecules (Basel, Switzerland)* 17: 6196-6236, 2012.
153. **Zhou WL, Medine CN, Zhu L, and Hay DC**. Stem cell differentiation and human liver disease. *World J Gastroenterol* 18: 2018-2025, 2012.
154. **Zorn A**. Liver Development. In: *Stem Book*, edited by Melton DaG L2009.
155. **Zorn AM, and Wells JM**. Vertebrate endoderm development and organ formation. *Annual review of cell and developmental biology* 25: 221-251, 2009.