Multi-OMICS analysis of stem cell-derived hepatocyte-like cells reveals a liver-intestine hybrid state that can be targeted by bioinformatics-guided interventions

Dissertation

Zur Erlangung des Grades des Doktors der Naturwissenschaften (Dr. rer. nat)
an der Fakultät für Chemie und Chemische Biologie der Technischen Universität
Dortmund vorgelegt

Vorgelegt von

Patrick Nell, M.Sc.

Dortmund, 2021

Erster Gutachter: Prof. Dr. Jan G. Hengstler

Zweiter Gutachter: Prof. Dr. Jörg Rahnenführer

Eidesstattliche Versicherung (Affidavit)

_____
Name, Vorname
(Surname, first name)

_____
Matrikel-Nr.
(Enrolment number)

_____
Ort, Datum
(Place, date)

_____
Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

_____

_____

_____

_____
Ort, Datum
(Place, date)

_____
Unterschrift
(Signature)

Publications and manuscripts in the context of this PhD thesis:

Ardakani, F. B. *et al.* (2020) 'Prediction of single-cell gene expression for transcription factor analysis', *GigaScience*. Oxford University Press, 9(11), pp. 1–14. doi: 10.1093/gigascience/giaa113.

Godoy, P. *et al.* (2018) 'Assessment of stem cell differentiation based on genome-wide expression profiles', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society Publishing. doi: 10.1098/rstb.2017.0221.

Lucendo-Villarin, B. *et al.* (2020) 'GENOME-WIDE EXPRESSION CHANGES INDUCED BY BISPHENOL A, F AND S IN HUMAN STEM CELL DERIVED HEPATOCYTE-LIKE CELLS', *EXCLI Journal*, 19, pp. 1459–1476. doi: 10.17179/excli2020-2934.

Sachinidis, A. *et al.* (2019) 'Road Map for Development of Stem Cell-Based Alternative Test Methods', *Trends in Molecular Medicine*. Elsevier Ltd, pp. 470–481. doi: 10.1016/j.molmed.2019.04.003.

Kattler, K *et al.* (202x) 'In vitro differentiated hepatocyte-like cells remain in an immature epigenetic state' [Manuscript draft]

Nell, P. *et al.* (202x) 'Identification of an FXR-modulated liver-intestine hybrid state in iPSC-derived hepatocyte-like cells' [Manuscript draft undergoing major revision at the Journal of Hepatology]

Content

SUMMARY

Human induced pluripotent stem cell (iPSC)-derived cells hold much promise for future cell therapy applications in the field of personalized regenerative medicine and offer an unlimited source of cell material for all branches of life-science associated research, including their use as model systems in pharmacological and toxicological studies. Differentiation of iPSC to hepatocyte-like cells (HLC) has been achieved in vitro by a number of differentiation protocols showing varying degrees of phenotypic maturity. However, their genome-wide expression patterns still differ strongly from primary human hepatocytes (PHH) in that they express liver- as well as intestine-associated genes. In this PhD thesis, differentiation of iPSC via definitive endoderm (DE) to HLC was characterized by single cell, as well as bulk RNA-sequencing with complementary epigenetic analyses. A supervised clustering strategy was developed to identify genes that exhibit similar expression dynamics during the differentiation of iPSC to HLC and assign them into differentiation pattern groups (DPG). These DPGs were analyzed by bioinformatics procedures to identify gene networks and regulatory factors that are likely to contribute to the hepatocyte (favorable) or non-hepatocyte (adverse) differentiation of genes in each DPG. Epigenetic analysis showed global hypermethylation of promoters in iPSC- derived cells compared to PHH and that there is a link between chromatin accessibility and favorable and adverse gene expression, suggesting that HLC fail to acquire an epigenetic landscape characteristic of PHH. Strikingly, single cell RNA-seq revealed that favorable and adverse gene expression occurs in a single population of HLC, rather than subpopulations. Thus, HLC exist in a hybrid state, where hepatocyte-associated genes are expressed in concert with genes that are not expressed in PHH or fetal hepatocytes (FH) - mostly intestinal genes - within the same cell. This finding highlights a substantial lack of developmental cues that are required to guide lineage-specific differentiation of iPSC to hepatocytes. Overrepresentation analysis at the bulk level, as well as regulon analysis at the single cell level, identified sets of regulatory factors that differ between HLC, FH, and PHH, hinting at a central role for the nuclear receptor FXR in the functional maturation of HLC. Combined FXR expression plus agonist exposure enhanced the expression of hepatocyte-associated genes and suppressed undesired non-liver gene expression, thereby improving HLC similarity to PHH.

ZUSAMMENFASSUNG

Humane induzierte pluripotente Stammzellen (iPSC) zeigen Potenzial für die therapeutische Applikation in der personalisierten Medizin, sowie die Verwendung zur Entwicklung von Modellsystemen in pharmakologischen und toxikologischen Studien. Die Differenzierung von iPSC zu Hepatozyten-ähnlichen Zellen wurde durch eine Vielzahl an Protokollen realisiert, wobei die resultierenden Zellen ein variierendes Ausmaß phänotypischer Reife aufweisen. Die genomweite Expression zeigt enorme Unterschiede zu primären Hepatozyten (PHH), veranschaulicht an Hand der Expression von Leber- wie auch Darm-assoziierten Genen. In dieser Arbeit wurde die Differenzierung von iPSC über definitive Endoderm (DE) zu HLC mittels Einzelzell- sowie Bulk-RNA-Sequenzierung und epigenetischen Analysen charakterisiert. Für die Auswertung wurde eine überwachte Strategie zur Gruppierung von Genen mit ähnlichen Expressionsverläufen in Differenzierungsmuster-Gruppen (DPG) entwickelt. Diese DPGs wurden mit bioinformatischen Methoden analysiert um Gen-Netzwerke und Regulatoren zu identifizieren, die zu gewünschter oder auch unerwünschter Differenzierung beitragen. Epigenetische Analysen konnten zeigen, dass eine Verbindung zwischen der Zugänglichkeit von Chromatin und der Expression erwünschter und unerwünschter Faktoren besteht und dass HLC im Vergleich zu PHH eine globale hyper-Methylierung ihrer Promoter Regionen aufweisen. Dies legt grundsätzlich nahe, dass HLC nicht die für Hepatozyten charakteristische epigenetische Landschaft ausbilden. Außerdem konnte mittels Einzelzell-Sequenzierung gezeigt werden, dass HLC eine einzige Zellpopulation darstellen, welche in einem Hybridzustand verharrt, in dem Hepatozyten-assoziierte Gene zusammen mit unerwünschten – hauptsächlich Darm-assoziierten - Genen, die weder in adulten noch fetalen primären Hepatozyten vorkommen, exprimiert werden. Diese Ergebnisse verdeutlichen, dass in der Differenzierung von HLC in vitro wichtige Signale fehlen, die der Differenzierung von Hepatozyten in vivo zu Grunde liegen. Analysen zur Überrepräsentation von Gen-Regulatoren, die zwischen HLC, fetalen und adulten primären Hepatozyten unterschiedlich exprimiert werden, deuteten auf eine zentrale Rolle für den Kernrezeptor FXR hin. Die Kombination von FXR Expression und Aktivierung des Rezeptors durch Agonisten konnte die Expression von Hepatozyten-assoziierten Genen fördern, während die Expression unerwünschter Darm Gene teils unterdrückt wurde. Auf diese Weise konnte gezeigt werden, wie die HLC Differenzierung durch gezielte Manipulation von Gennetzwerken verbessert werden kann.

ABBREVIATIONS

%	Percent
°C	Degree Celsius
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometer
2D	Two-dimensional
3D	Three-dimensional
AB	Antibody
AFP	Alpha-fetoprotein
ALB	Albumin
ATAC-seq	Assay for transposase accessible chromatin followed by sequencing
ATP	Adenosine triphosphate
BGBl	Bundesgesetzblatt
BMP	Bone morphogenetic protein
bp	Base pair
BSA	Bovine serum albumin
BSEP/ABCB11	Bile salt export pump
Ca	Calcium
CaCl	Calcium chloride
CDCA	Chenodeoxycholic acid
CDX2	Caudal type homeobox transcription factor 2
ChiPSC18	Cellartis human induced pluripotent stem cell at passage 18
Cl^-	Chlorine ion
cm^2	Square centimeter
CMFDA	5-chloromethylfluorescein diacetate
CO_2	Carbon dioxide
CPM	Counts per million

CSC	Cancer stem cell
CXCR4	C-X-C motif chemokine receptor 4
DAPI	4',6-diamidino-2-phenylindole
DAR	Differentially accessible region
DE	Definitive Endoderm
D_i	Developmental index
DILI	Drug induced liver injury
DiPaC	Differentiation Pattern Clustering
DKK1	Dickkopf 1
DMEM	Dulbecco's modified eagle medium
DMR	Differentially methylated region
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Nucleoside triphosphate
D_P	Developmental potency
DPBS	Dulbecco's phosphate-buffered saline
DTT	DL-Dithiothreitol
E	Embryonic day
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescent protein
EMT	Epithelial to mesenchymal transition
ENV	Envelope
ESC	Embryonic stem cell
ESchG	Embryonenschutzgesetz
EtOH	Ethanol
FBS	Fetal bovine serum
FCS	Fetal calf serum

FDR	False discovery rate
FGF	Fibroblast growth factor
FH	Fresh hepatocytes
FOXA2	Forkhead box A2
FXR	Farnesoid-X-receptor
G	Standard gravity
g	Gram
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GRN	Gene regulatory network
H_2O	Water
$HCO3_-$	Bicarbonate
HEPH	Hephaestin
HGF	Hepatocyte growth factor
HKG	Housekeeping gene
HLC	Hepatocyte-like cell
HNF1α/HNF1A	Hepatocyte nuclear factor 1 alpha
HNF4α	Hepatocyte nuclear factor 4 alpha
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
ICM	Inner cell mass
IF	Immunofluorescence
iPSC	Induced pluripotent stem cell
ITS	Insulin-transferrin-selenium
iTSC	Induced tissue-specific stem cell
ISX	Intestine-specific-X transcription factor
KCl	Potassium chloride
KH_2PO_4	Potassium dihydrogen phosphate
KLF5	Kruppel like factor 5
L	Liter

LRH1/NR5A2	Nuclear receptor subfamily 5 group a member
LSEC	Liver sinusoidal cell
LSM	Laser scanning microscope
MBD4	Methyl-CpG binding domain 4, DNA glycolase
MEP1A	Mephrin A subunit alpha
Mg	Magnesium
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
MOI	Multiplicity of infection
MRP	Multidrug resistance protein
mRNA	messenger RNA
NANOG	Homeobox transcription factor Nanog
NaOH	Sodium hydroxide
ng	Nano gram
NPC	Non-parenchymal cell
NTCP	Na^+-taurocholate co-transporting polypeptide
OAT	Organic anion transporter
OATP	Organic anion transporting polypeptide
ORF	Open reading frame
OSM	Oncostatin M
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGD	Pre-implantation genetic diagnosis
PFA	Paraformaldehyde
PHH	Primary human hepatocyte
POU5F1	POU domain, class 5, transcription factor 1

PPARα	Peroxisome proliferator-activated receptor alpha
PrE	Primitive Endoderm
PV	Portal vein
qRT-PCR	Quantitative real-time polymerase chain reaction
RLU	Relative light unit
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rpm	Revolutions per minute
RRBS	Reduced representation bisulfite sequencing
RT	Room temperature
RT-PCR	Reverse transcriptase PCR
s	Second
sc-RNA-seq	single cell RNA sequencing
SDS	Sodium dodecyl sulfate
SHP/NR0B2	Nuclear receptor subfamily 0 group b member
SI	Sucrase isomaltase
SNAI1	Snail family transcriptional repressor 1
SNARF-1	5-(and -6)-Carboxy-SNARF-1
SOP	Standard operating procedure
StZG	Stammzellgesetz
TCA	Tricarboxylic acid
TF	Transcription factor
TFGβ	Transforming growth factor β
TWIST1	Twist basic helix-loop-helix transcription factor 1
WIF1	WNT inhibitory factor 1
XME	Xenobiotic metabolizing enzyme

1. INTRODUCTION

1.1 Development of stem cell technology in life science

Stem cell research and technology gained much attention in recent years, looking back at more than half a decade of scientific investigation. First indications for hematopoietic stem cells (HSC) can be found in the works of Ray D. Owen in 1945, who described the existence of cells that 'are apparently capable of becoming established in the hematopoietic tissues (…) and (…) provide a source of blood cells (…), presumably throughout his life' [1]. Key properties of (hematopoietic) stem cells were subsequently described in 1961, in a series of bone-marrow transplantation experiments in irradiated mice [2], in which Till and McMulloch concluded that only a fraction of undifferentiated cells of a bone marrow suspension could give rise to proliferating colonies of primitive cells in the spleen of intravenously injected mice. Over the following years, other types of tissue-resident adult stem cells were identified, including intestinal crypt stem cells [3,4] and neural stem cells [5] among others. The term 'embryonic stem cell' was first coined by Gail R. Martin after the successful isolation of mouse in 1981 and subsequently human embryonic stem cells by Thomson in 1998 [6,7], describing cells that were capable of giving rise to all major tissues and cell types of an organism. More recently, the generation of in vitro reprogrammed pluripotent stem cells in 2006 and 2007 using Yamanaka's cocktail of transcription factors [8,9] revolutionized the use of embryonic stem cell-like cells in science.

In this comparably short period of time, stem cell research has made enormous contributions to our understanding of embryonic and post-natal development, organ- and tissue-homeostasis as well as mechanisms of disease. Some of this knowledge has been translated into applications in regenerative medicine, such as peripheral blood and bone marrow hematopoietic stem cell transplantation (HSCT) to reconstitute the blood cell populations of patients suffering from conditions such as leukemia and acute ionizing irradiation syndrome. Starting with low patient survival rates following transplantation using bone marrow from normal donors without HLA matching or major histocompatibility complexes in 1957, E. Donnall Thomas reported that more than 50% of acute non-lymphoblastic leukemia patients went into remission following autologous bone marrow HSCT in 1979 [10]. By 2018, 9.55 stem cell transplants have been performed per 100,000 inhabitants in Germany, representing the highest rate of stem cell transplants in Europe

[11]. Another important example of how concepts of stem cell biology are applied to solve issues of clinical relevance is the emergence of the cancer stem cell (CSC) hypothesis, which contributed to understanding of mechanisms underlying metastasis, drug- and radio-resistance, functional heterogeneity and cellular plasticity of cancer [12,13].

In addition to the development of clinically relevant procedures and therapies, the identification and isolation of embryonic and adult stem cells has helped scientists to understand the mechanisms that lie at the mere basis of the development of life and maintenance of organisms. With increasing knowledge on the potency of stem cells and the mechanisms that lead to their acquisition of positional identity, self-renewal, differentiation and organization in niches in different tissues of the adult body as adult stem cells, researchers have been able to shed light on the mechanistic basis of the 'cell theory' that dominated the theory of preformism in the late 19th century [14]. Today it is accepted that living organisms are made up of cells that form the basic structural units of our bodies and that they come from pre-existing cells, so that all parts of the developing body arise from the germ and stem cells sharing the same set of DNA. Based on these insights, the perspective of employing principles of embryonic development and cell specialization to develop clinical therapies attracted much attention in the scientific community. It is thus not surprising that the first isolated human embryonic stem cell lines prompted extensive investigation of their therapeutic potential, starting in the end of the 20th century and continuing to date [15–17]. However, to differentiate stem cells into specialized cell types or tissues that could be employed in cell therapy approaches such as HSCT or as tissue transplants, for example to support damaged heart muscle tissue after infarction, the mechanisms governing cellular differentiation into specialized cells based on the same genetic material need to be understood in order to recapitulate them in vitro.

In 1957, Conrad Hal Waddington described cellular differentiation in the course of development, comparing the unspecialized cell to a ball running down a hill with diverging branches (Figure 1 A) [18]. He showed that environmental conditions can 'canalize' the events that determine the balls' path, thus enforcing a specific path, in a series of fruit fly experiments. He further described that, once the ball has reached the bottom of a specific path (terminal differentiation), it cannot easily cross the landscape to settle in a different path or reverse the process and run up the hill.

Thus, Waddington described cellular differentiation as a unidirectional process in the course of development [19,20]. With his work, Waddington challenged the Neo-Darwinist view that mutations on the DNA level would be responsible for such phenotypic plasticity and proposed that the genes responsible for newly assimilated phenotypes are already present in a population, inferring that they just need to be assembled by influences above the DNA level [21]. With this observation, the scientific field of epigenetics emerged, investigating the mechanisms that act above DNA to establish phenotypic changes in cells, such as DNA methylation, the establishment of histone marks and regulation by micro-RNAs involved in nucleosome organization and remodeling. It can thus be concluded that every specialized cell type, be it a neuron, a blood cell, a hepatocyte or an intestinal enterocyte follows a different path down Waddington's epigenetic landscape. Assuming that all of those very different cell types arise from early stem cells that form the embryo, it is intuitive to assume that some cell types are more closely related than others. In this example, this can be illustrated at the hand of the hepatocyte and intestinal enterocyte, which are both highly metabolically active epithelial cells that arise from the early gut endoderm of the embryo. In contrast, the neuron, the blood cell and the hepatocyte have common progenitors that occur much earlier in embryonic development, in form of pluripotent stem cells in the inner cell mass (ICM) of the developing blastocyst.
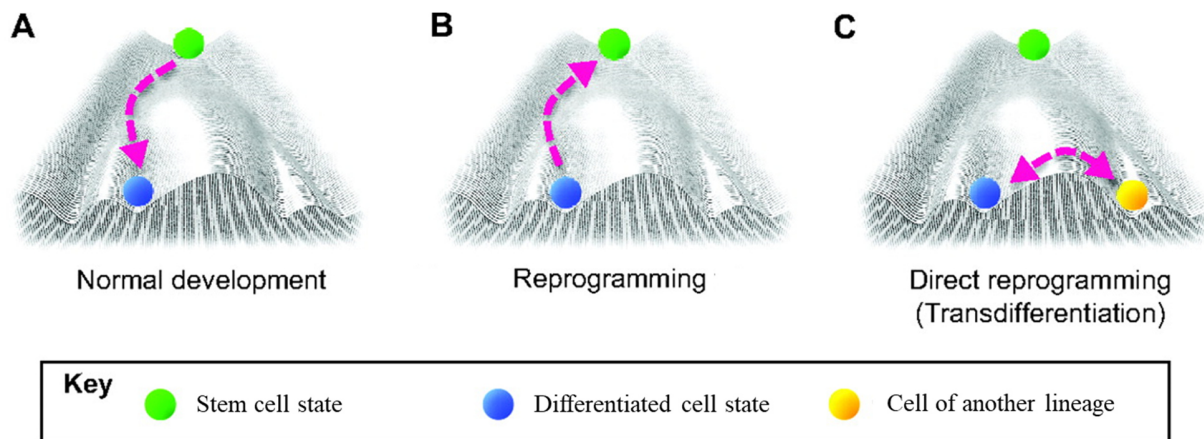
Figure 1: Waddington's epigenetic landscape illustrating developmental fate decisions of pluripotent stem cells. A) Canonical development by sequential cell fate decisions depicted as a path down the hill of Waddington's epigenetic landscape. B) Cellular reprogramming reversing the differentiated cells state to become pluripotent. C) The process of trans-differentiation is illustrated by conversion of the cell fate of one differentiated cell to become a differentiated cell of another lineage, without prior reprogramming to a pluripotent state. Adapted from [20].

First doubts concerning the unidirectionality of Waddington's concept of sequential establishment of the epigenetic landscape in cellular differentiation arose soon after it became a central dogma in developmental biology. In 1958, James Gurdon showed that the transfer of somatic cell nuclei into enucleated *Xenopus laevis* oocytes gave rise to sexually mature individuals [22], which showed that those somatic cell nuclei could be reprogrammed by factors in the enucleated oocyte (Figure 1 B). Almost 40 years later, this could be shown also in mammals at the famous example of the cloned sheep named Dolly [23]. Furthermore, experiments employing certain transcription factors, also known as master regulators, could show that terminally differentiated fibroblasts can be pushed over the hills dividing Waddington's epigenetic landscape to become myoblasts; a process that is now known as trans-differentiation (Figure 1 C) [24]. The fact that trans-differentiation is a process that also occurs in vivo, for example allowing cells to compensate tissue damage [25], highlights the true plasticity of cellular identities.

Although these insights fueled the development of (cell) therapeutic strategies, such as treatments for retinal degeneration [26] or spinal cord injury [27] and enabled ground breaking technology, such as the generation of transgenic mouse models by blastocyst transfer of gene-engineered embryonic stem cells [28], essential limitations remained that prohibited stem cell-based technology to live up to its long praised potential in clinical application. One disadvantage is that human embryonic stem cell (ESC)-based therapy potentially requires life-long immunosuppressive treatment. In addition, relatively limited amount of ESC lines carrying specific mutations that could serve as disease models are available, which is why their use fell short in studies related to mechanisms of disease [29]. Also, use of ESC lines that were isolated from the inner cell mass of blastocysts is considered a delicate ethical issue, as it involves the destruction of potential life. Even though it has been shown that ESC may also be isolated from pre-implantation stage blastomeres - that are technically not defined as ESC - and do not lead to destruction of the embryo since they are isolated in a procedure resembling that of preimplantation genetic diagnosis (PGD), the regulatory landscape for the use of ESC poses legal challenges to their application. In general, legal requirements for research involving ESC strongly vary internationally, with a wide range of accepted utilization of ESC in Europe (**Figure 2**).

Color	Policy Type	Countries
🟩	**Permissive** *(e.g., SCNT is specifically permitted under certain conditions)*	Belgium, Finland*, Spain, Sweden, United Kingdom
🟨	**Permissive Compromise** *(e.g., SCNT is prohibited; hESC research using supernumerary IVF embryos is specifically permitted or not prohibited)*	Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland*, France, Georgia, Greece, Hungary, Iceland, Latvia, Moldova, Netherlands, Norway, Portugal, Romania, Russian Federation, San Marino, Slovenia, Switzerland, Turkey
🟧	**Restrictive Compromise** *(e.g., hESC research only permitted using cell lines created before a certain date)*	Germany, Italy
🟥	**Prohibitive** *(e.g., research using embryos or cell products derived from embryos is prohibited)*	Austria, Lithuania, Poland, Slovakia

Figure 2: Overview of legal regulation of research involving human embryonic stem cells in Europe. Situation in April 2021 according to the Hinxton group (http://www.hinxtongroup.org/wp_eu_map.htmL, accessed 04-04-2021).

In Germany, the 'Embryonen-schutzgesetz' (ESchG 13th of December 1990 BGBl. I S. 2746 last changed 21st of November 2011 BGBl. I S. 2228) issued in 1990 and enforced in 1991 prohibits the use of human embryos for research purposes. The definition of embryos here includes the fertilized egg and every totipotent cell. While it is often argued that leftover embryos from in vitro fertilization (IVT) could be used for research, the ESchG also limits the number of in vitro fertilized eggs per IVT cycle to three, while demanding that all of those embryos are transplanted to the recipient. Furthermore, the German 'Gesetz zur Sicherstellung des Embryonenschutzes im Zusammenhang mit Einfuhr und Verwendung menschlicher embryonaler Stammzellen' (StZG Stammzellgesetz vom 28. Juni 2002 BGBl. I S. 2277, last changed 29th of March, 2017 BGBl. I S. 626) regulates the import and use of ESC lines that were isolated in other countries, prohibiting import and use of ESC lines that were isolated after May 7th of 2007. These legislative restrictions, in Germany, limit the potential of ESC in regenerative medicine irrespective of other ethical

considerations. However, even without legislative restrictions, the availability of embryos for personalized therapies would still be limited for practical reasons. This is also true for trans-differentiation approaches, because the number of cells that could be harvested from patients is limited.

Finally, with their finding that somatic cells, such as skin fibroblast, can be reprogrammed to acquire a pluripotent stem cell state, Takahashi and Yamanaka in 2006 and 2007 found a way to circumvent ethical restrictions of stem cell isolation, overcome limitations in availability and proved that indeed cellular differentiation could be reversed up to the point of a pluripotent stem cell state (Figure 1 B). Human induced pluripotent stem cells (iPSC) can be expanded, maintained and differentiated in vitro, which in combination with the progress in gene engineering technology has opened up exciting perspectives, such as the development of human- and patient-specific in vitro models for development and disease or potentially endless cell material availability for tissue engineering, transplantation and cell therapies. Several disease models have been developed, such as a model for the long QT syndrome [30] and Alzheimer's disease [31]. Examples for clinical applications of iPSC to date include treatments for retinal degeneration [32] and Parkinson's disease [33], among others. In recent years the development of iPSC-based three-dimensional cell aggregates that show self-organization into tissue-resembling structures, commonly referred to as organoids, has led to the establishment of bio-banks that store vast amounts of patient- and disease-specific organoids for research that is supposed to be translated into medicine in the near future [34]. However, in cell therapeutic scenarios as well as model systems, major issues have to be overcome in order to ensure biosafety, reproducibility of test results and accurate resemblance of iPSC-derived cells of their in vivo counterparts, all of which could be limited by genetic and phenotypic instability of iPSC as well as differentiation products. To date, a major concern for cell therapeutic treatments on the basis of iPSC is that they were shown to share numerous features with neoplastic cells, such as oncogene expression and reprogramming-associated DNA mutations, indicating their tumorigenic and immunogenic potential [35]. Furthermore, with rapid progress in stem cell technology and bioengineering, societies need to face ethical debates concerning, for example, the research and use of chimeric

animals as potential sources of human donor organs through embryo complementation with genetically engineered iPSC [36] or gene correction approaches in human embryos [37].

1.2 Stem cell types and developmental potency

In general, stem cells are classified as either embryonic or adult stem cells. While both types of stem cells share the property of self-renewal, meaning that cells divide either symmetrically, giving rise to two identical cells, or asymmetrically, where one daughter-cell is of a more specialized cell type, they differ in their developmental potency. The highest form of developmental potency is totipotency, where a cell has the potential to give rise to any tissue, intra- and extraembryonic, thus giving rise to an entire organism, which generates offspring [38]. In humans, only the zygote is proven to be totipotent, albeit some evidence suggests that cells of the 4-cell stage may be totipotent as well [39]. Totipotency of the human embryo is established by the removal of epigenetic marks from and remodeling of the parental chromatin [40].
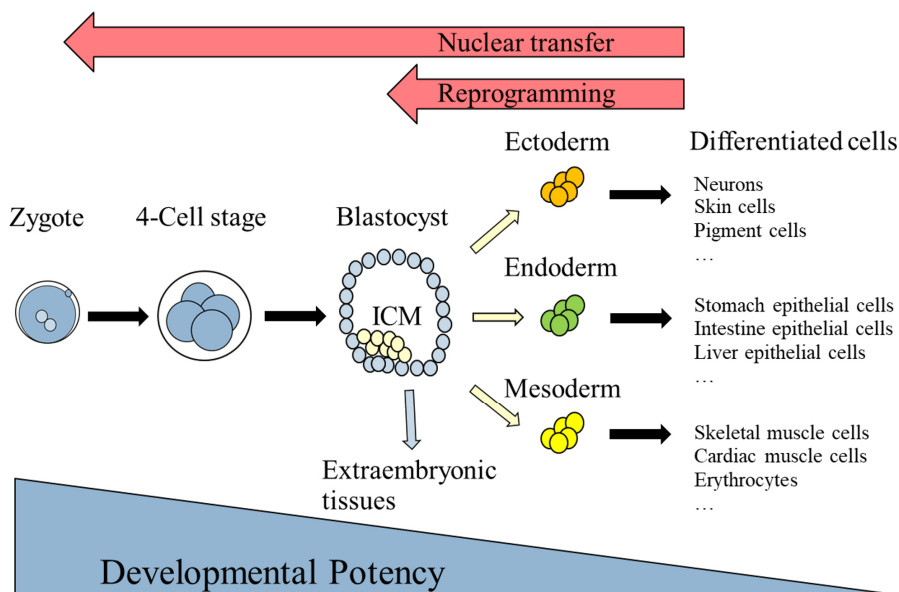
Figure 3: Developmental potency declines following successive cell divisions and differentiation in early embryonic development. With specialization and corresponding lineage choices, the ability to differentiate into other lineages becomes restricted. This ability can be restored by reactivation of e.g. pluripotency associated gene networks to varying degrees (illustrated by red arrows).

This involves the gradual demethylation of DNA at CpG dinucleotides - an epigenetic modification that is generally associated with transcriptional silencing [41] - and editing of histone marks

during pre-implantation development. In contrast, the restriction of developmental potency and differentiation is generally associated with de novo methylation during post-implantation development. Pluripotent stem cells, such as ESC of the inner cell mass (ICM) of the blastocyst, are limited to differentiation into the three germ layers, ecto-, meso-, endoderm and their derivatives and thus cannot form extraembryonic structures. Reprogramming of somatic cells to iPSC using the Yamanaka cocktail of retroviral vectors for delivery of the transcription factors OCT4 (POU5F1), NANOG, SOX2 and MYC (also known as OSKM method), reverses the differentiation state of the somatic cells to become pluripotent [9]. Although the reprograming to and exit from pluripotency remains incompletely understood, it has been shown that POU5F1 is associated with the regulation of DNA methyltransferases that need to be modulated in order to remove DNA methylation in pre-implantation development and allow re-methylation in pluripotent cells of the blastocyst [42]. To date, several non-invasive protocols for iPSC generation have been developed [43]. Furthermore, other combinations of transcription factors have been employed to improve the reprogramming process and gene delivery has been improved to avoid (random) integration into the host genome and improve efficiency, for example with synthetic RNA [44] or Epstein-Barr virus-derived sequences [45]. Such improved reprogramming approaches are thought to reduce the risk of teratoma formation, which is of major concern for the use of iPSC in regenerative medicine. Despite the successful acquisition of pluripotency, it has been reported that epigenetic memory of the reprogrammed cell type is retained in iPSC [46]. This phenomenon has been exploited to produce multipotent induced tissue-specific stem cells (iTSC), which are suggested to show more efficient differentiation into the lineage of their origin and reported to not form teratomas when subcutaneously injected into mice [47].

Multipotent stem cells, such as cells from the three germ layers, hematopoietic, neural and mesenchymal stem cells can differentiate into all cells of a specific lineage. They can be generated by differentiation of pluripotent stem cells or isolated from adult tissues with varying efficiency and have been used for HSCT [48]. They have shown potential for treatment of orthopedic disease [49] and cardiac infarction [50,51]. Of the least developmental potency are unipotent progenitor cells, such as germ line stem cells [52] or epidermal stem cells [53], which in addition to self-renewal are capable of differentiation into one cell type only.

At the moment, mesenchymal stem cells represent the stem cell type that is employed most in clinical trials with 1,199 listed trials in April of 2021 (https://www.clinicaltrials.gov). The number of clinical trials with iPSC- or ESC-based therapy is much lower, with 113 and 48 listed studies, respectively. However, regarding in vitro test systems, mesenchymal stem cells cannot be obtained, expanded, maintained and differentiated as effectively as pluripotent stem cells, which makes their use less feasible.

1.3 The liver

The human liver is the largest internal organ of the human body and is located in the right upper quadrant of the abdominal cavity. It receives oxygenated blood from the hepatic artery and oxygen depleted blood from the intestine, spleen and pancreas via the portal vein, accounting for about one third and two thirds of the total blood supply, respectively [54]. Among the various metabolic as well as endo- and exocrine functions exerted by the liver are bile production, which is essential for absorption of fat and lipophilic nutrients, metabolization of dietary compounds for carbohydrate and lipid homeostasis, biochemical detoxification, production of blood coagulation factors and maintenance of cardiovascular homeostasis, e.g. by elimination of old erythrocytes. Furthermore, the liver is involved in immune regulation by exerting local and systemic tolerogenic effects, evidenced by the success of porcine allografts of liver together with skin, kidney or heart from the same donor, without application of immunosuppressive drugs [55].

The basic microvascular unit of the liver is the hepatic acinus, represented by a cluster of parenchymal cells, the hepatocytes, around the terminal branches of the hepatic arteriole and portal venule that together with collecting bile ducts and lymph vessels form the space of Mall, also known as portal triads (Figure 1 A, B). The human liver is estimated to consist of about 100,000 acini [54]. Oxygenated blood enters through the hepatic arterioles and mixes with oxygen depleted blood from portal venules into hepatic sinusoids - thin-walled microvessels that are lined by highly fenestrated liver sinusoidal endothelial cells (LSEC), which allow blood plasma flow into the space of Disse between sinusoids and hepatocytes. Having passed the sinusoids, blood enters terminal hepatic venules that converge into the hepatic vein, bearing low oxygen levels. Six portal triads arranged in a hexagon formation around a central terminal hepatic venule constitute a liver

lobule; three terminal venules around a portal triad are referred to as a portal lobule (Figure 1 A). This architecture and associated gradients of oxygen, nutrients, hormones and signaling molecules such as WNT/β-catenin account for metabolic zonation of the liver that determines hepatocyte functionality along sinusoids. It has been shown that by reversing blood flow in vivo, liver zonation can be reversed [56]. Under physiological conditions, the periportal zone (zone 1) is characterized by oxidative energy metabolism, gluconeogenesis, tricarboxylic acid cycling and ureagenesis. Bearing relatively low levels of oxygen, the perivenous zone (zone 3) exhibits higher activity in xenobiotic metabolism, glycolysis and liponeogenesis. The intermediate zone (zone 2) is characterized by a mixture of zone 1 and 3.
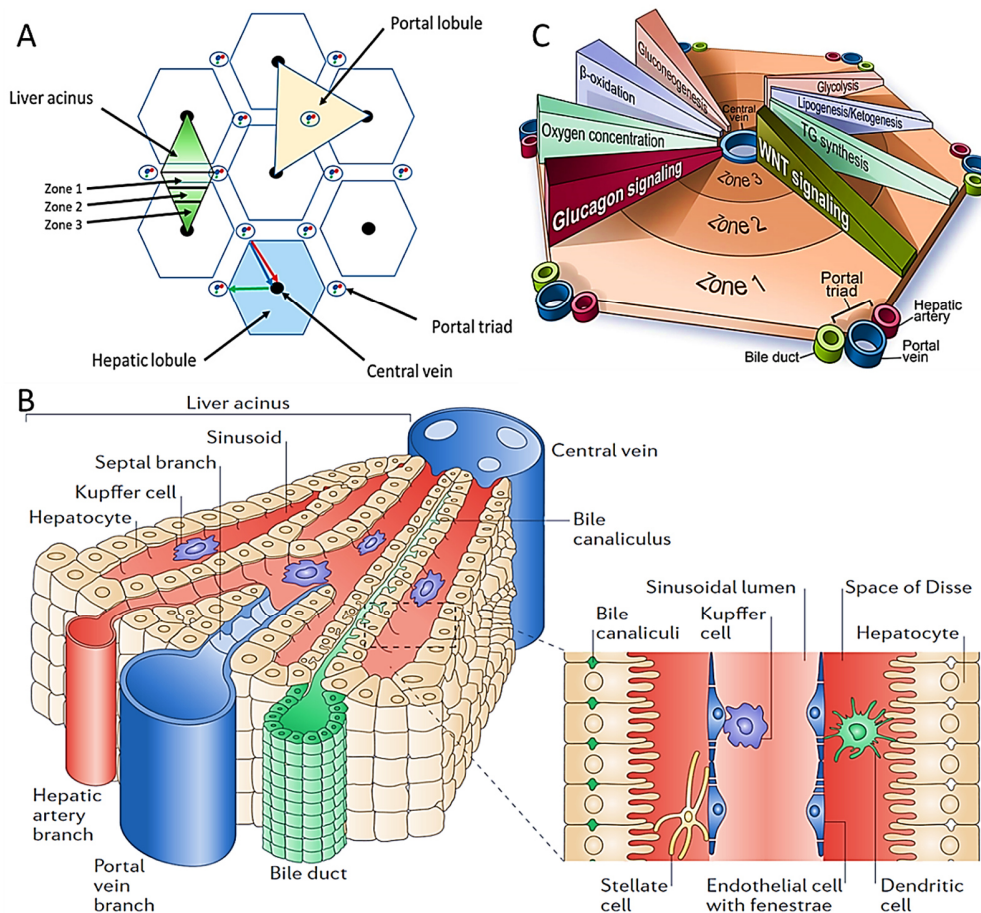

Figure 4: Liver microscopic architecture. A) Schematic of liver architectural units [57] **B)** Liver zonation and associated processes according [58] **C)** Schematic representation of the liver acinus architecture and close-up illustrating the resident cell populations in their respective compartments [59].

In the perisinusoidal space of Disse, plasma containing nutrients, glucose, bacteria, debris and toxins is available to the microvilli on the basolateral membranes of adjacent hepatocytes that are arranged in plates of 15-25 cells [60]. Hepatocytes are parenchymal cells that constitute about 80% of the liver, representing the major cell type responsible for its metabolic capacity [61]. The microvilli maximize the surface area of the basolateral membrane (Figure 1B), which harbors transporters such as the Na^+-taurocholate co-transporting polypeptide (NTCP), members of the organic anion transporting polypeptide (OATP) family and organic anion transporters (OAT) [62]. Compounds that are taken up by hepatocytes are modified, stored or degraded via different metabolic pathways, often under the influence of hormonal regulation, while excessive plasma in the space of Disse is drained into lymphatic vessels. Glucose, for example, under influence of insulin signaling, can be broken down to pyruvate via glycolysis, stored as glycogen, oxidized through the TCA cycle to produce ATP or converted into fatty acids or amino acids through de novo synthetic pathways [63]. In a fasting state, glycogen, pyruvate or lactate, can be employed for synthesis of glucose under influence of glucagon signaling. In addition to their metabolic repertoire, hepatocytes also exert important functions in innate immunity [64], by secreting acute phase proteins including complement and clotting factors. These are, together with other metabolic products such as lipids, proteins and carbohydrates, released into the blood stream, leaving the liver via the inferior vena cava [65]. On their apical side, hepatocytes secrete products of hepatocyte metabolism including conjugated bilirubin, bile salts, cholesterol, phospholipids, copper, proteins, ions and water into bile canaliculi via transporters such as the bile salt export pump (BSEP/ABCB11) and multidrug resistance proteins (MRP and MDR) [62]. The bile is transported in the opposite direction of the blood flow towards the canals of Hering, interlobular bile ducts of the portal triad [66] and larger ducts of the hepatic biliary tree. Cholangiocytes, which are closely related to hepatocytes, are non-parenchymal cells of the liver that form a polarized epithelial cell layer lining these ducts and are involved in modification of bile in a series of reabsorptive processes. These include secretion of Cl^-, HCO_3^-, and water from cholangiocytes and reabsorption of bile acids, amino acids, and glucose into cholangiocytes [67]. Subsequently, bile is transported via extrahepatic ducts to the gallbladder, where it is stored and concentrated, and to the intestine, where bile acids are involved in the digestion and absorption of fats and fat-soluble vitamins, before being recycled to the liver via the portal vein. Waste products are

excreted. The circulation of nutrient-rich blood to the liver and bile to the intestine is referred to as enterohepatic circulation and plays a major role in models of pharmacokinetics and toxicity of recirculating compounds [68].

A number of other non-parenchymal cells participate in specialized physiological functions of the liver. Within sinusoids, non-parenchymal cells play an important role in the innate immune response. Liver resident macrophages, also known as Kupffer cells, act as the first line of defense against particulates and immunoreactive compounds that enter via the portal circulation [69]. Adherent to LSECs, they effectively capture and phagocytose dying erythrocytes and other cells in the hepatic parenchyma and act as antigen-presenting cells that can either act immunogenic or tolerance-inducing, depending on innate immune signals [70]. Together with LSEC and Kupffer cells, professional antigen-presenting dendritic cells form the reticulo-endothelial system, which clears antigens, waste and toxins from sinusoidal blood by uptake through endocytic receptors. Some of these ligands can be transported to hepatocytes for metabolic conversion [65]. Kupffer cells can also interact with pit cells, which are large granular lymphocytes that originate from circulating natural killer cells of the immune system. They adhere to LSEC and Kupffer cells within sinusoids, where they are involved in the elimination of tumorigenic cells, thus representing a protective mechanism against cancer [71]. Another important non-parenchymal cell type of the liver are stellate cells, constituting up to 8% of the total cells in the liver [72]. They reside in the perisinusoidal space of Disse and due to their long cytoplasmic processes have been called 'Sternzellen'. These usually quiescent cells have been implicated in important homeostatic and regulatory functions in health and disease, including storage of 80% of retinoids in human body within lipid droplets and turnover of extracellular matrix compounds [72]. It has been reported that stellate cells exhibit significant contractile ability that contributes to the regulation of blood flow within sinusoids [73]. Under disease conditions, such as liver fibrosis, stellate cells become activated through various mechanisms involving cytokine and stress signals and contribute to disease progression through the production of larger amounts of extracellular matrix, while acquiring a (myo-) fibroblastic phenotype.

The liver has been shown to regenerate up to 70% of its mass after resection, a process which involves all cell types and an orchestrated process of growth factor and cytokine action and is

largely dependent on the plasticity of differentiated hepatocytes and biliary cells, rather than liver progenitor cells (LPC); stem cells that long have been assumed to replenish the hepatocyte pool in the 'streaming liver' model [74]. Despite this enormous potential, liver disease today represents a major burden worldwide with two million deaths per year. The diverse functions of the liver are central to our health – loss of liver functionality leads to chronic 'hepatic encephalopathy' and eventually coma [75]. Liver cirrhosis and hepatocellular carcinoma rank in the top 20 causes of death, which is fueled by the high prevalence of risk factors in the population world-wide, including alcohol consumption, obesity, diabetes and drug abuse. Viral hepatitis prevalence is high and the prevalence of acute hepatitis continues to rise on the basis of increasing incidence of drug induced liver injury (DILI), which represents the leading cause of withdrawal of pre-clinical studies of drugs in development. Top drugs linked to liver injury world-wide include antimicrobials, central nervous system agents, anti-epileptic drugs, cardio-vascular drugs and anti-neoplastic agents. In the UK and US, acetaminophen represents a top ranking cause for DILI [76]. Injuries of the liver are generally associated with inflammation. If left untreated, inflammation of the liver develops into fibrosis, with increased scar formation in the liver tissue, slowly replacing the healthy, functional liver tissue architecture. Liver fibrosis over time develops into cirrhosis, where the scar formation process becomes irreversible and liver functionality is strongly impaired. From here, it progresses to end-stage liver disease (ESLD), which can only be treated with liver transplantation. Liver transplantation represents the second most common solid organ transplantation although less than 10% of globally required donor organs are available. Hepatocyte transplantation has been applied to temporarily improve liver function, however, it is dependent on non-transplantable donor livers from which hepatocytes are isolated and associated with several disadvantages. Hepatocytes need to be isolated from donor livers, a process that is associated with loss of proliferative and metabolic capacities, decreased cell viability, poor engraftment and modest therapeutic benefit [77]. Yet, human liver material is not only required for transplantation, but also the source of human primary hepatocytes, which are currently the most widely accepted, golden standard of in vitro systems in use in pharmacologic and toxicity testing. Therefore, alternative cell sources for liver in vitro test systems, such as stem cell-derived hepatocyte-like cells (HLC), are attractive.

1.4 Stem cell-based in vitro test systems for liver and developmental toxicology

The wide use of in vitro systems for evaluation of compound toxicity prior to pre-clinical and clinical development has several advantages, since these systems can offer species specificity and successfully identify toxic compounds. This way the need for costly experiments in vivo, using animal models, can be reduced to those cases where compounds that passed initial in vitro toxicity screens are evaluated in the context of intact metabolic routes, i.e. in rodents and primates. However, investigating toxicity and the molecular mechanisms involved following exposure to test compounds requires competence of in vitro systems to not only be representative of cytotoxic mechanisms that are shared by diverse cell types, such as apoptosis mechanisms, but also tissue-, organ- or even cell type-specific molecular mechanisms. In liver toxicology, xenobiotic metabolizing enzyme (XME) induction assays require expression and activity of those enzymes in order to generate representative results. These assays serve to investigate the risk of unwanted drug-drug interactions posed by administration of test compounds. They were carried out in rats, dogs and sometimes monkeys up to the early 1990s, until they were replaced by in vitro XME induction assays using cryopreserved hepatocyte cultures [78]. However, even though in vitro test systems have expanded on the existing repertoire of available methods to assess drug safety and efficiency in treating disease, effects involving mechanisms such as the adsorption into the blood, distribution to tissues and organs, metabolism and excretion via the kidneys or bile are difficult to predict with standard in vitro methods. It is therefore not realistic that all animal experiments will be replaced by in vitro models in the near future. Rather, a combination of in vivo and in vitro experiments is advised. Nevertheless, developing suitable in vitro systems and assays that can reliably predict compound effects holds enormous potential for resolving ethical concerns and avoiding the cost of animal experiments, inter- and intra-species accuracy of results and shortages of primary tissue and cells. Induced pluripotent stem cell-based test systems are especially promising in this regard: (i) iPSC can be generated from somatic cells of any patient in unlimited numbers, (ii) they can theoretically be differentiated into any cell type constituting human tissues and (iii) they may serve as (patient-specific) model systems mimicking development and disease conditions in the future.

Regardless of whether they will be used for the assessment of cyto- or developmental toxicity, stem cell-based, organ-specific test systems will need to be able to produce mature cell types representative of the respective organ. To date, the development of stem cell-based liver and, more specifically, hepatocyte differentiation systems has progressed to a point where protocols from different laboratories allow for differentiation of HLC or liver organoids in vitro [79–85]. Despite this progress, the use of primary human hepatocytes (PHH) represents the gold standard for liver toxicology mostly due to the fact that stem cell-based hepatocyte differentiation fails to model hepatocyte-specific (xenobiotic) metabolism to an extent that is truly comparable to PHH. Notably, also primary human hepatocytes ex vivo were shown to de-differentiate rapidly when taken into culture, especially in a 2D monolayer culture format [86,87]. Culturing PHH in a sandwich culture, where PHH are cultured in between collagen gel layers, or spheroids, where cells are aggregated in hanging drop culture systems, has been shown to improve the stability of their phenotype in vitro and allow for higher and more durable expression of metabolizing enzymes [88,89]. Still, creating a phenotype-stabilizing microenvironment in vitro continues to pose a challenge that needs to be overcome by the currently available culture methods. Moreover, whether mono-cellular 2D test systems, 3D spheroids or even multi-cellular organoids are best applicable in toxicological testing largely depends on the type of assay that is used, the reproducibility of the results and acceptance as an industry standard. Since the results of toxicity tests can depend on the levels of drug-metabolizing enzymes, deviations in their expression by orders of magnitude in HLC compared to primary hepatocytes is an obstacle to their implementation.

Since iPSC require induction of developmental signaling pathways in order to differentiate them into specialized cell types, they are also considered a promising tool for assessment of developmental toxicity. Test compounds that interfere with such developmental pathways can be expected to negatively affect iPSC differentiation in vitro, thus indicating developmental toxicity. A critical consideration for testing developmental toxicity is the discrimination of cytotoxic substances and developmentally harmful teratogens. Therefore, the ratio by which developmental genes are enriched among genes that become differentially expressed as a consequence of the compound treatment, compared to what can be randomly expected, has

been proposed as an index of developmental toxicity (D_i) (**Figure 4**) [90]. A more direct but less sensitive measure of whether a compound has the potential to induce developmental effects is given by the overlap of developmental genes and genes altered by a test compound divided by the number of developmental genes, which was described as developmental potency (D_P).
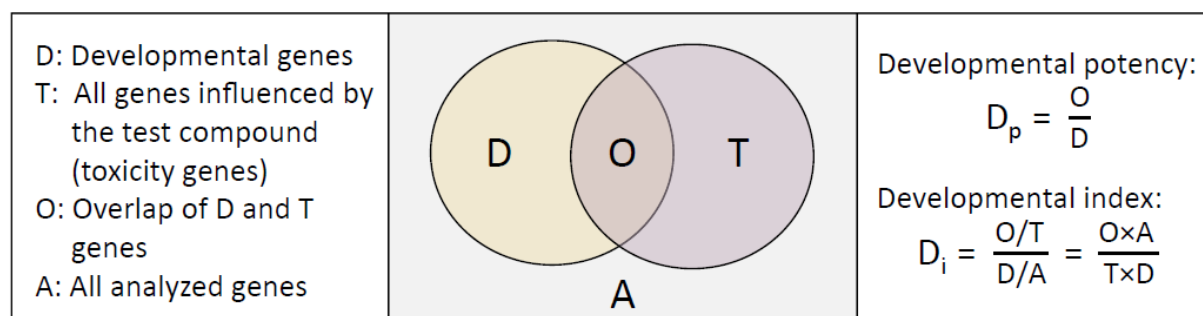
D: Developmental genes
T: All genes influenced by the test compound (toxicity genes)
O: Overlap of D and T genes
A: All analyzed genes

Developmental potency:
$$D_p = \frac{O}{D}$$

Developmental index:
$$D_i = \frac{O/T}{D/A} = \frac{O \times A}{T \times D}$$

Figure 5: Quantification of the developmental effects of test compounds in stem cell differentiation. Developmental potency (D_p) gives the fraction of all developmental genes (D genes) that are affected by a compound. The developmental index (D_i) informs how strongly D genes are enriched among all genes that are affected by a compound (T genes). [78]

Even though in vitro differentiation cannot yet fully recapitulate embryonic development with regard to its duration and the complexity of micro-environmental cues that establish positional identities, guiding cell lineage choices and organogenesis into fully functional tissues, this technology represents a considerable leap forward in our capacity to investigate developmental toxicity mechanisms of compounds at sub-toxic concentrations. However, because of the vast number and incomplete understanding of processes that are involved in human embryonic development, it remains to be seen if stem cell based in vitro test systems will allow to successfully identify critical concentrations of teratogens that induce congenital abnormalities in vivo. Therefore, comprehensive studies employing stem cell-based test systems representative of various developmental pathways are required that are exposed to large numbers of substances including appropriate controls with low and high cytotoxicity, but no developmental toxicity.

1.4.1 Hepatocyte differentiation in liver development and PSC differentiation strategies

In order to establish stem cell-based alternative cell sources as a replacement for primary human hepatocytes, stem cell-based differentiation approaches in vitro aim at recapitulating known developmental mechanisms in vivo. Detailed knowledge on human embryonic liver development is limited but required in order to achieve proper differentiation into hepatocytes. Although availability of studies on human liver development is scarce, conserved key mechanisms were identified in mice. It is well known that hepatocytes are epithelial cells that originate from endoderm lineage, which is established as the result of one of the first cell fate decisions in mammalian development.
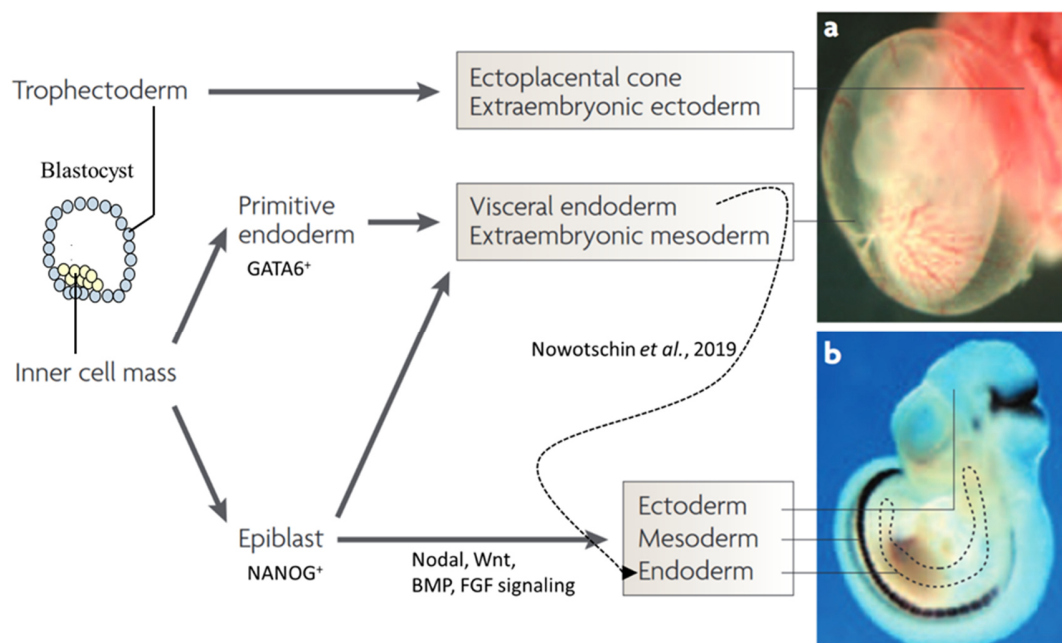

Figure 6: Early lineage decisions in the developing mouse embryo lead to formation of extra-embryonic tissues and germlayer formation. **A)** Extraembryonic tissues arise from trophectoderm and contribute to placenta formation (a). **B)** Pluripotent embryonic stem cells from the inner cell undergo the first linage choice of embryonic development involving Nanog and Gata6 expression, to either become primitive endoderm (PrE) or pluripotent epiblast cells. The figure was adapted from [91].

The endoderm is one of the three primary germ layers - ectoderm, mesoderm and endoderm - that originates from pluripotent stem cells of the inner cell mass (ICM) of blastocysts and gives rise to various epithelial cell types in organs lining the respiratory and digestive tract, including thyroid, thymus, lungs, liver, biliary system, and pancreas [92]. Two types of endoderm cells

emerge during mammalian development and contribute to formation of the early gut tube from which the future liver bud will be formed (Figure 6). First, primitive endoderm (PrE) cells, which were long thought to exclusively contribute to the extraembryonic parietal and visceral yolk sac, arise at the blastocyst stage at embryonic day (E) 3.5-4 from cells of the ICM. The second type, the cells of definitive endoderm (DE) arise from another fraction of the ICM, called epiblast, during gastrulation. Lineage decision of the ICM cells towards either PrE or epiblast cells is not fully understood, but involves downregulation of the pluripotency-associated transcription factor POU5F1 and 'salt and pepper' expression of Nanog and GATA6, where Nanog primes cells for the epiblast and GATA6 for the PrE lineage, respectively (Figure 6). Additionally, FGF4, BMP4 and PDGFA signaling has been shown to be required for maturation of the PrE lineage. These signaling molecules are produced by epiblast cells and received by neighboring cells that have been specified as PrE [93]. In PrE, endoderm markers including SOX17, FOXA2, CXCR4, GATA4a and DAB2 become expressed [93,94].

The formation of DE in the process of gastrulation likewise involves a complex interplay of different signaling pathways and cell migration that sequentially specifies the (positional) identity of the epiblast cells. Gastrulation leads to the formation of the three primary germ layers - ectoderm, mesoderm and (definitive) endoderm - and starts around E6.5 with the formation of the primitive streak, which in the course of its elongation of the epiblast defines the anterior-posterior and medio-lateral axes of the embryo [95]. Acquisition of migratory properties by a fraction of epiblast cells through epithelial-to-mesenchymal transition (EMT) facilitates their ingression [96] through the primitive streak to give rise to mesoderm and definitive endoderm at E7.5 [91,92,95]. These transient, migratory cells represent common progenitors of mesoderm and DE, referred to as mesendoderm [97]. Epiblast cells that do not undergo EMT and remain at the anterior pole of the epiblast contribute to ectoderm formation. In contrast to the long prevailing view that the future gut endoderm arises solely from DE cells that originate from the epiblast, it was shown recently in a single cell sequencing approach that the future gut tube (indicated with a dotted line in the in situ hybridization image of Figure 6 B) is formed through intercalation of visceral endoderm that arises from PrE and definitive endoderm cells of epiblast origin [98].

Several morphogens are required for the control and fine tuning of these processes including WNT, more specifically WNT3a, BMP, FGF and Nodal. Their signaling controls initiation of the primitive streak and the formation of DE, mesoderm and ectoderm, as well as specification of different derivatives of the gut tube, based on duration and concentration of signaling gradients [99,100]. The complexity of these processes is illustrated by Nodal signaling in DE specification, where high Nodal activity is initially required but needs to be repressed for endoderm maturation [91,101]. Since Nodal was shown to control the activity of a number of endoderm genes it is considered to be the main regulator for endoderm development [100]. Additional transcriptional regulators associated with mesoderm and endoderm formation are TBXT and FOXA2, respectively. The pioneer transcription factor FOXA2 has been linked to the opening of chromatin and induction of endoderm-associated target genes, such as HEX, OTX2, CER, SHH, and SOX17, many of which are transcription factors implicated in endoderm patterning [102].
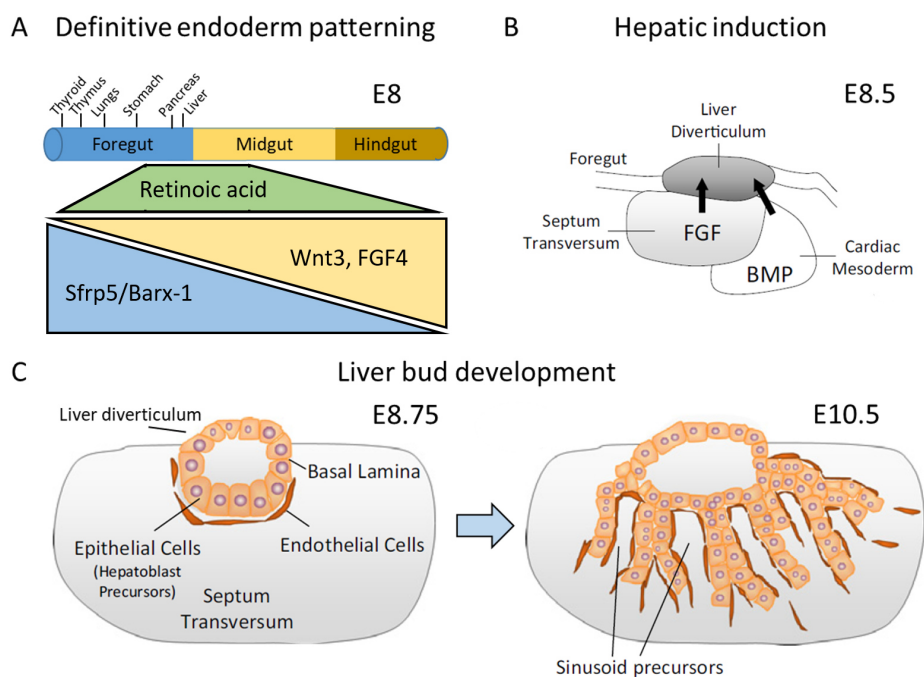

Figure 7: Key steps of early liver bud development in vivo. A) The early gut tube made of DE is patterned by morphogen signaling into foregut, midgut and hindgut regions. For specification of the foregut, WNT-antagonism enables hepatic induction around E8.5. Adapted from [78,103]. **B)** Under influence of BMP signaling from the cardiac mesoderm and FGF signaling from the septum transversum, the liver diverticulum is formed. Adapted from [78]. **C)** Hepatoblast precursors from the liver diverticulum start invasion of the septum transversum mesenchyme around E9.5, forming the early liver bud, where hepatoblasts are lined with endothelial sinusoid precursors at E10.5. Adapted from [78].

Gut endoderm is patterned into the foregut, midgut, and hindgut around E8 under control of WNT, FGF and BMP cytokines secreted by the surrounding mesoderm, as well as retinoic acid signaling [78,103,104] (Figure 7 A) . WNT and FGF ligands are expressed at higher concentrations in the mesoderm surrounding the hind- and midgut compared to the foregut [105]. By contrast, the specification of the foregut requires the action of WNT antagonists, such as secreted frizzled related proteins (SFRPs), which was shown to be dependent on BARX1 expression [106–108]. This gradient with low WNT and FGF signaling is critical for early foregut patterning. At E8.5, bone morphogenetic proteins (BMPs) and FGF from the septum transversum and the cardiac mesoderm induce thickening of the ventral foregut, which forms the liver diverticulum [109–112] (Figure 7 B). The liver diverticulum consists of columnar epithelial cells surrounded by a basal lamina and endothelial precursor cells (Figure 7 C). At E9.5, the epithelial cells break through the basal lamina and invade the septum transversum. Under the influence of the microenvironment of the septum transversum mesenchyme, the epithelial cells of the liver diverticulum differentiate into hepatoblasts. The latter already express numerous genes of adult hepatocytes; for example, albumin and hepatocyte nuclear factor 4α (HNF4α), although not yet at the level of the mature cells and also some genes that will be downregulated in the adult liver, such as alpha-fetoprotein. It has been shown that HNF4α acts as an indispensable master regulator of hepatocyte differentiation, acting from the top of the hierarchy to regulate expression of other hepatic transcription factors, such as HNF1α and NR1I2 (PXR) [113]. At E 13, some periportal hepatoblasts differentiate to cholangiocytes under the influence of the periportal mesenchyme [114,115], while the majority of hepatoblasts differentiate to hepatocytes, under influence of signaling molecules such as oncostatin M (OSM), which is secreted from hematopoietic cells and acts in a paracrine fashion, repressing alpha-fetoprotein expression and inducing hepatocyte phenotype [116,117]. Another essential factor in liver development and hepatocyte maturation is the hepatocyte growth factor (HGF) that also is known for its role in liver regeneration, among others [118]. Experiments in mice have shown that embryos lacking the morphogenic effects of HGF signaling in liver development die in utero [119]. Besides the complexity of the microenvironment generated at the different developmental days, namely the early (E8) mesoderm, septum transversum (E8.5), and cardiac mesoderm (E8.5), the existence of narrow time windows of less than 6 h should be considered, when the same cytokines may have opposite effects. For example,

WNT factors have to be antagonized for correct foregut patterning but are later needed for maturation of hepatoblasts [104]. Although a considerable number of - in part - well-studied signaling pathways have been uncovered in studies liver development, fine-tuning of cellular fates through the integrated known and unknown pathways is far from being understood and represents an exciting challenge in life sciences. Influences of hepatic differentiation are also contributed by extrahepatic tissues and intrahepatic cell types. Especially the process of hepatic induction, as well as the induction of other organ primordia from the gut tube represent challenging research targets as these complex and tightly temporally controlled processes are not fully understood at a high enough resolution, especially considering influences from surrounding tissues, such as the splanchnic mesoderm. Recent advances with single cell sequencing combined with computational signaling network inference uncovered complex relationships between different mesodermal subtypes that interact with epithelial cells in liver organogenesis and offer new insight on the underlying relationships [120]. Furthermore, hematopoietic stem cells that populate the liver around E11.5-E16.5 have shown capacity to differentiate into hepatocytes in vivo and may exert important functions in development of hepatocytes [121]. Therefore, additional efforts are required to optimally recapitulate the complexity and the quickly changing dynamics of a developing liver in vivo [78].
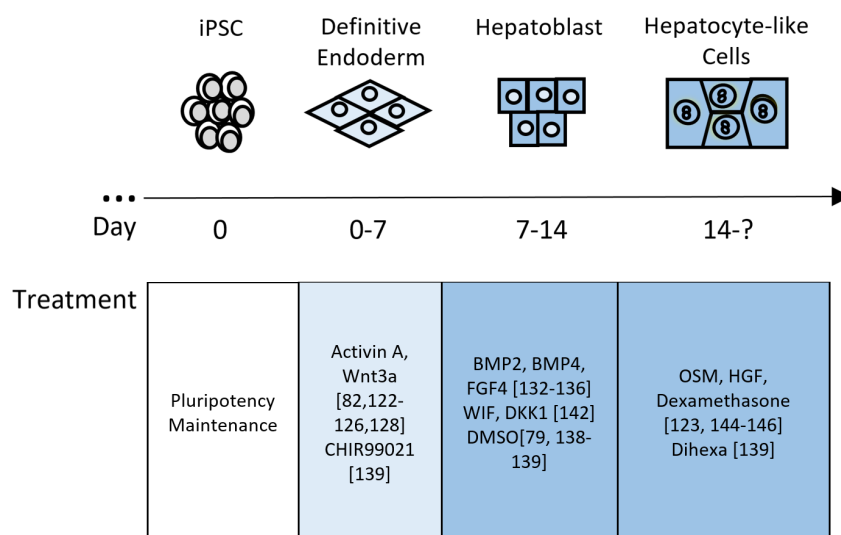

Figure 8: In vitro differentiation of pluripotent stem cells into hepatocytes. Stem cells are commonly induced by a combination of WNT3a and Activin A to differentiate into DE within a time span of maximal 7 days. Differentiation towards hepatoblasts is achieved for example by application of BMP-2 and FGF-2,

WNT-antagonists or sub-toxic DMSO concentrations. Maturation into hepatocyte like cells is induced mainly by HGF and OSM growth factors and often combined with interventions involving transcription factors or matrix compounds. Time windows for treatments can vary for every protocol.

In vitro, differentiation of ESC or iPSC to DE is achieved by employing the knowledge generated in developmental studies. Common protocols for induction of definitive endoderm [82,122–126] use Activin A to mimic Nodal signaling, both of which are members of the transforming growth factor beta (TGFβ) superfamily of morphogens and act through type II activing receptors and the downstream effectors Smad2 and Smad3 [100]. Smad2/3 complexes can recruit different activating and inhibiting co-factors of transcription leading to cell-type specific responses. The strategy of employing Activin A in DE differentiation makes use of this signaling cascade to induce EMT, regulated by activity of the transcription factor SNAI1 [127]. In addition to Activin A, WNT3a is often used to achieve endoderm differentiation in the early differentiation phase. While WNT-signaling has been established to endow competence of cells to respond to Activin A signaling [128], long exposure with WNT3a was shown to limit some of the differentiation potential of the presumptive DE cells. Alternative approaches use small molecule GSK3 inhibitors such as CHIR99021 to favor accumulation of beta-catenin, its translocation to the nucleus and activation of target genes [129–131]. Presumptive DE cells produced by these approaches have been shown to express genes representative of DE, including SOX17, FOXA2 and CXCR4 and have been used for further differentiation into different endoderm derivatives, including pancreas, liver, and intestinal cells, also in combination with stem cell derived meso- and ectoderm [101].

To establish foregut identity and hepatic induction of definitive endoderm, in vitro protocols often involve BMP-4 and FGF-2 [132–136] in order to mimic signaling from the cardiac mesoderm and septum transversum, based on the knowledge generated in developmental biology studies [112]. These morphogens have varying effects based on the applied concentration. In the case of FGF2, increasing dose induces intestinal or pancreatic differentiation [137]. Other approaches for specification of DE and differentiation of hepatoblasts avoid use of morphogens and instead use DMSO for this purpose [79,138,139], which has been shown to downregulate pluripotency genes and enhance the hepatic differentiation potential [140,141]. As reported for foregut specification in vivo, the use of WNT inhibiting factors such as WNT inhibitory factor (WIF1) and Dickkopf

(DKK1) has been shown to aid hepatic differentiation [142]. The resulting populations of hepatoblast-like cells show expression of genes representative of embryonic hepatoblasts, including AFP, HNF1α and HNF4α [124,143]. These immature hepatoblast-like cells can be further matured using growth factors, most importantly OSM and HGF [123,144–146]. The addition of OSM in this case compensates for the lack of paracrine signaling from hematopoietic cells, which would provide OSM under physiological conditions. Together with dexamethasone, OSM induces hepatocyte marker expression via the OSM receptor, as well as hepatocyte-like morphology and functional maturation, i.e. glycogenesis, and suppression of alpha-fetoprotein [117,147]. HGF in turn is normally produced by stromal cells that support liver regeneration and hepatocyte maturation [118] and therefore commonly added to differentiation media as a recombinant protein. Because of the variability associated with use of recombinant proteins and undefined media components, such as fetal bovine serum, some protocols try to establish hepatic differentiation only by the use of small molecules. Use of CHIR99021 for DE differentiation, DMSO for hepatic induction and dihexa for hepatic maturation is an example of such a protocol, where dihexa is thought to potentiate HGF signaling [139]. Although a great number of differentiation protocols for the generation of HLC have been reported to date, the great majority of all protocols exploit the here mentioned strategies, using different concentrations of the differentiating compounds and different time points or culture matrices [148]. In addition, some studies investigated the influence of transduction of HLC with transcription factors [149,150], the differentiation in 3D spheroid or organoid structures [151–153] or targeted gene activation with CRISPR-Cas [85]. Protocols for HLC differentiation thus have achieved a remarkable progress towards alternative sources of PHH and are under investigation for their use in cell therapeutic applications, showing promising results in some respects, e.g. for the treatment of patients with primary hyperoxaluria with gene-corrected HLCs [154]. However, numerous studies suggest that the state of maturity of generated cells lacks well behind that of PHH [80,88,155–158], illustrating the knowledge gap and technical limitations that are yet to be overcome to generate a fully competent alternative to PHH.

1.5 Current limitations of hepatocyte-like cells

Despite the progress that has been made towards generating HLC from stem cells, including the expression of important markers, studies have shown that these cells are not yet fully comparable to PHH. It was observed that characteristics of HLC often are subject of overoptimistic interpretation, requiring unbiased procedures for HLC analysis [88]. In general, HLC generated by different protocols and laboratories have been shown to exhibit low expression of important gene clusters associated with hepatocyte functions (up to 100-fold lower than in PHH) [78,80], including for instance genes involved in hepatocyte metabolism, especially related to the processing of xenobiotics, and the major transcription factors associated with these processes, such as NR1I3 (CAR), NR1H4 (FXR), and NR1I2 (PXR). Other transcriptional regulators that are highly expressed in HLC are characteristic of dividing cells (E2F2, MYB), cells undergoing EMT (SNAI1, TWIST) or hepatic progenitors (AFP) and do not decrease to levels observed in primary human hepatocytes during differentiation to HLC. Importantly, some genes that are not expressed in PHH increase in expression, indicating misguided differentiation into other tissue cell types, such as intestinal epithelial cells, together with associated transcriptional regulators (CDX2, KLF5). This has two major implications in relation to the application in toxicological or pharmacological tests: First, the immature character of HLC, especially regarding the lack of enzymes involved in metabolic functions, limits their capacity to faithfully predict adverse effects of investigated compounds. Second, the presence of unwanted features associated with other cell types such as stem cells and intestinal cells presents a risk of responses that are not representative of hepatocytes, but other cell types, thus inviting misinterpretation of results. With regard to their application in cell therapeutic approaches, the presence of non-hepatocyte associated gene expression and immature state raises questions with regard to the phenotypic stability of HLC in such settings. In the past, multiple studies relied on selected markers of hepatic differentiation for the evaluation of HLC. More recently, studies increasingly relied on genome-wide analysis using microarray or next generation sequencing (NGS) to assess the transcriptome of differentiated cells, supported by algorithms specifically designed to compare differentiated cells and target cells. CellNet [85,159,160] and liver-specific gene expression algorithms [161] objectify the differentiation status of stem cell-derived cell types. It was determined that HLC generated by different protocols reach between 50 and 80% of the Gene Regulatory Network

(GRN) status of PHH and reference liver tissue [78,80,85,162]. CellNet processes gene expression data from HLC as input and compares them to large data sets of expression profiles of different cell types [160]. Using this platform, it was demonstrated that the relatively large difference between HLC and PHH is not only explained by the low expression of numerous liver-associated genes in HLC, but also by the expression of 'undesired genes' that are normally expressed in e.g. endodermal and intestinal cells but not in adult hepatocytes [80,139]. It is therefore vital to thoroughly characterize gene expression changes during differentiation of iPSC to HLC in comparison to PHH in a time-dependent manner on the population and single-cell level so that molecular mechanisms can be identified that may be employed to improve HLC differentiation and to get insight into the heterogeneity of HLC populations.

1.6 Aim

Human hepatocytes are required for clinical therapy as well as for studies in pharmacology and toxicology. Stem cell-derived HLC offer the perspective of an unlimited supply of cells that closely resemble human hepatocytes and therefore may serve in toxicological and pharmacological model systems or cell therapy approaches in the future. However, genome-wide and functional analyses demonstrated that HLC still show features of misguided differentiation and functional immaturity, which need to be understood and overcome to enable their use in representative model systems. This PhD thesis aims to analyze the differentiation process of iPSC towards HLC, employing a state-of-the art differentiation strategy together with genome-wide transcriptomics on bulk and single cell level and epigenomics procedures to develop bioinformatics-guided intervention strategies for improving HLC differentiation.

The specific aims were:

1. To study the HLC phenotype using qRT-PCR, immunocytochemistry and a bile canalicular secretion assay to confirm the previously reported [80] hepatic as well as non-hepatic features of HLC.
2. To characterize HLC differentiation at the stages of iPSC, DE, HLC and PHH using bulk transcriptomics and epigenomics, and develop an approach to identify and describe favorable and adverse gene expression changes induced by the differentiation protocol.

3. To investigate if features of hepatic and intestinal differentiation arise in separate HLC subpopulations or within the same (hybrid) cells using high-throughput immunocyto-chemistry and single-cell RNA-sequencing.

4. To identify possible targets for intervention to improve HLC differentiation using bioinformatics analysis.

5. To establish proof-of-principle that an intervention by lentiviral transduction of HLC with NR1H4 (FXR) and its activation by chenodeoxycholic acid (CDCA) and GW4064 can shift the balance of liver and intestinal identity in HLC towards liver.

2. MATERIAL AND METHODS

2.1 Material

2.1.1 Equipment

Table 1: Equipment

Item	Manufacturer
5075 ELV Autoclave	Tuttenauer
7500 Fast & 7500 Real-Time PCR System	Applied Biosystems
AF 100 ice machine	Scotsman
Agilent 2100 Bioanalyzer	Agilent Technologies
Balance EW	Kern
Biofuge Heraeus Fresco 21	Thermo Scientific
Centrifuge 5424R	Eppendorf
Centrifuge MEGA STAR 1.6R	VWR
Chemical safety cabinet Herasafe™	Heraeus
CO2 Incubator C150 R Hinge 230	Binder
Diaphragm Vacuum Pump	Vacuumbrand
Dual-Action Shaker KL 2	Edmund Bühler
Freezer (-20 °C)	Siemens
Freezer (-80 °C)	Thermofisher Scientific
HiSeq2500 Sequencer	Illumina
IKAMAG RCT magnetic stirrer	IKA
ImageXpress Micro XLS	Molecular Devices
Laminar flow hood (Electronics FAZ2)	Waldner
LSM 880 confocal laser scanning microscope	Zeiss
Megafuge 1.0R	Heraeus
MiniSpin®/ MiniSpin® plus	Eppendorf
NanoDrop 2000	PeQLab Biotechnologie GmbH
Neubauer cell counting chamber	VWR

PH-meter CG 842	Schott
Pipette Reference	Eppendorf
Pipette Research plus	Eppendorf
Pipetteboy	Integra
Precision balance EW 150-3M	Kern
Rocking Platform Shaker	VWR
Shaker KS 260 basic	IKA
SONOPULS mini20 sonicator	Bandelin
Testtube shaker	VWR
Vortex-Genie 2	Bender&Hobein
Waterbath GFL 1083	Gesellschaft für Labortechnik

2.1.2 Consumables

Table 2: Consumables

Item	Manufacturer	ID
Biosphere Filtered Tip 100 µl	Sarstedt	70.760.212
Biosphere Filtered Tip 1000 µl	Sarstedt	70.762.211
Biosphere Filtered Tip 20 µl	Sarstedt	70.1116.210
Biosphere Filtered Tip 200 µl	Sarstedt	70.760.211
Cell scraper	Sarstedt	83.183
Cover glass	VWR International (Argos Technologies)	631-0711
Falcon ® 24-well TC-treated Plate	Corning	353504
Falcon tube 15 mL	Sarstedt	62.554.512
Falcon tube 50 mL	Sarstedt	62.547.254
Filtropur v100 1000 mL vacuum filtration unit	Sarstedt	83.1824.001
Filtropur v25 250 mL vacuum filtration unit	Sarstedt	83.1822.001
Filtropur v50 500 mL vacuum filtration unit	Sarstedt	83.1823.001
Kimtech Science Delicate Task Wipes	Kimberly-Clark Professionals	7216

MicroAmp Optical 96-well Reaction Plate	Applied Biosystems	N801-0560
MicroAmp Optical Adhesion Film	Applied Biosystems	4311971
PCR SingleCap 8er-SoftStrips 0.2mL	Biozym Scientific GmbH	710988
Pipette Tips 1000 µL	Sarstedt	70.762
Pipette Tips 20 µL	Sarstedt	70.1116
Pipette Tips 200 µL	Sarstedt	70.760.002
RNase-free Microfuge Tubes 1.5 mL	Thermo Fischer Scientific	AM12400
RNaseZap® RNase Decontamination Solution	Thermo Fischer Scientific	AM9780/AM9782
SafeSeal 0.5 mL microtube	Sarstedt	72.699
SafeSeal 1.5 mL microtube	Sarstedt	72.706
SafeSeal 2.0 mL microtube	Sarstedt	72.695.500
Serological Pipette 10 mL	Sarstedt	86.1254.001
Serological Pipette 25 mL	Sarstedt	86.1685.001
Serological Pipette 5 mL	Sarstedt	86.1253.001
Serological Pipette 50 mL	Sarstedt	86.1689.001
Tissue culture flask 25 cm²	Sarstedt	83.3910.502
Tissue culture flask 75 cm²	Sarstedt	83.3911.502
Tissue Culture Plate Flat-Bottom 12-Well Plate	Sarstedt	83.3921
Tissue Culture Plate Flat-Bottom 24-Well Plate	Sarstedt	83.3922
Tissue Culture Plate Flat-Bottom 6-Well Plate	Sarstedt	83.1839
WypAll L30 wipes	Kimberly-Clark Professionals	7301

2.1.3 Reagents

Table 3: Reagents

Item	Company	ID
(2'Z,3'E)-6-Bromoindirubin-3'-oxime	Sigma-Aldrich	B1686-5MG
10Z-Heptadecenoic Acid	Sigma-Aldrich	H8896-500MG
5-(and-6)-Carboxy SNARF™-1	ThermoFischer Scientific	C1270
Acetic acid	Carl Roth	3738.5
AG® 501-X8 Mixed Bed Resin	Bio-Rad	1437424
Alexa Fluor® 488 Donkey anti-Mouse	Dianova	715-545-150
Alexa Fluor® 488 Donkey anti-Rabbit	Dianova	711-545-152
Amino acid solution (Costumer formulation)	PAN Biotech GmbH	SO-33100
Ampure XP beads	Beckman Coulter	A63881
Anti-AGR2 polyclonal antibody	Atlas Antibodies	HPA007912
Anti-albumin polyclonal antibody	Atlas Antibodies	HPA031024
Anti-alphafetoprotein monoclonal antibody	R&D Systems	MAB1369
Anti-CDX2 recombinant antibody	Abcam	ab114247
Anti-DPPIV monoclonal antibody	Abcam	ab215711
Anti-HNF4a polyclonal antibody	Atlas Antibodies	HPA004712
Anti-ISX polyclonal antibody	Atlas Antibodies	HPA060328
Arachidonic Acid	Sigma-Aldrich	10931-250MG
beta-Mercaptoethanol	ThermoFischer Scientific	31350010
Bovine Albumin Fraction V (BSA)	Carl Roth	8076.4
BSA fatty acid free	Sigma-Aldrich	A8806-1G
Calcium chloride (CaCl2)	Sigma-Aldrich	5239.1
Cellartis® DEF-CS500 Culture System	Cellartis (TaKaRa)	Y30010

Cellartis® Definitive Endoderm Differentiation system	Cellartis (TaKaRa)	Y30035
Cellartis® Hepatocyte Differentiation System	Cellartis (TaKaRa)	Y30050
Cellartis® iPS Cell to Hepatocyte Differentiation System	Cellartis (TaKaRa)	Y30055
CellTracker™ Green CMFDA Dye	ThermoFischer Scientific	C2925
CGP 52608	Sigma-Aldrich	C5749
Chenodeoxycholic acid	Sigma-Aldrich	C9377
Chloroform	Carl Roth	7331.2
Cholecalciferol	Sigma-Aldrich	C9756-1G
Collagenase from Clostridium hystolyticum	Sigma-Aldrich	C2674
cOmplete™ Protease Inhibitor Cocktail	Sigma-Aldrich	11697498001
Cy3-AffiniPure Donkey anti-Rabbit	Jackson ImmunoResearch	AB_2307443
DAPI	Invitrogen	D3571
DEPC Treated Water	Invitrogen	750024
D-erythro-Sphingosine	Santa Cruz Biotech	sc-3546
Dexamethasone	Sigma-Aldrich	D4902-25MG
Direct-zol RNA MiniPrep Plus kit	Zymo Research	R2071
DMSO	Sigma-Aldrich	34869
DPBS	Sigma-Aldrich	
EGTA	Carl Roth	3054.2
Ethanol	Merck	100983
Ethanol 70%	Walter CMP	WAL10506
Ethidium Bromide	Invitrogen	15585-011
Gentamycin	PAN Biotech GmbH	P06-13001
Gibco™ Recombinant AOF Insulin	ThermoFischer Scientific	A11382II

GlutaMAX Supplement	ThermoFischer Scientific	35050061
GW4064	Sigma-Aldrich	G5172
HEPES	Carl Roth	9105.4
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems	4368813
IGEPAL CA-630	Sigma-Aldrich	I8896
Insulin-transferrin solution (IST) 100x	Sigma-Aldrich	I3146
Knockout DMEM	Life Technologies	10829-018
Knockout Serum Replacement	ThermoFischer Scientific	10828010
L- Glutamine	Sigma-Aldrich	G3126
L- Glutamine	PAN Biotech GmbH	P04-82100
Laminin 111	BioLamina	LN111-03
Laminin 521	BioLamina	LN521-25
Lyophilized rat-tail collagen	Roche Diagnostics	11179179001
MEM Essential Amino Acids Solution (50x)	ThermoFischer Scientific	11130051
MEM Non-Essential Amino Acids Solution (100x)	ThermoFischer Scientific	11140050
MinElute PCR Purification Kit	Qiagen	28004
Nextera DNA library prep kit	Illumina	20025519
Nextera XT library preparation kit	Illumina	FC-131-1096
Nonident P-40	Sigma-Aldrich	I8896-50ML
Penicillin/Streptomycin	PAN Biotech GmbH	P06-07100
Phalloidin-rhodamine	ThermoFischer Scientific	R415
Polyvinylpyrrolidone	Sigma-Aldrich	PVP360
Potassium Chloride (KCl)	Carl Roth	6781 .1
Potassium dihydrogen phosphate (KH2PO4)	Carl Roth	3904.1

Oleic acid	Sigma-Aldrich	O1383
Qiazol Lysis Reagent	Qiagen	79306
Rho-associated kinase (ROCK) inhibitor Y27632	Sigma-Aldrich	Y0503-1MG
Roti Histofix 4%	Carl Roth	P087.5
Roti-CELL Dulbecco's PBS	Carl Roth	9131.2
Sera Plus (Special Processed FBS)	PAN Biotech GmbH	3702-P103009
Sodium chloride (NaCl)	Carl Roth	3957.2
TaqMan® Universal Master Mix II	Applied Biosystems	4440038
Transferrin human	Sigma-Aldrich	T8158-100MG
Triton X-100	Carl Roth	3051
TRIzol™ regaent	ThermoFischer Scientific	15596026
Trypan Blue	Sigma-Aldrich	T6146
TrypLE	ThermoFischer Scienticfic	12563011
Tween 20	Sigma-Aldrich	P7949
Tween 80	Sigma-Aldrich	P8074

2.1.4 Cells

Table 4: Primary cells and cell lines

Item	Company	ID
ChiPSC18 induced pluripotent stem cells (male)	Cellartis (TaKaRa)	Y00305
Primary human hepatocytes (male) AFJ, DJJ, IAN	BioReclamationIVT	M00995-P

2.2 Methods

All cell culture procedures were performed under sterile conditions.

2.2.1   Preparation and maintenance of human induced pluripotent stem cell cultures

Induced pluripotent stem cells were obtained from Cellartis (ChiPSC18, Takara Bio Europe, Cat. No. Y00300) and stored in liquid nitrogen gas phase at -150°C (~3 million cells/vial). Culture vessels were prepared with COAT-1 solution (Cellartis) diluted 1:20 in ice-cold DPBS containing Ca^{2+} and Mg^{2+}, using 0.1 mL per cm² of the culture vessel. The culture vessels were incubated for at least 20 min at 37°C. Cells were thawed in a 37°C water bath and transferred to DEF-CS medium (Cellartis) containing DEF-CS growth factor 1, 2 and 3 reagents (Cellartis) diluted 1:333, 1:1000 and 1:1000, respectively (referred to as full DEF-CS medium hereafter). After centrifugation at 300g for 5 min, supernatant was discarded and the cell pellet resuspended in 5 mL of full DEF-CS medium. The coating solution was aspirated from prepared culture vessels just before transfer of the cell suspension. Finally, cells were left for attachment at 37°C and 5% CO_2 until the following day. ChiPSC18 cells were cultivated for at least 5 passage intervals until further use, with daily medium changes using Cellartis DEF-CS supplemented with growth factors 1 and 2 (referred to as DEF-CS maintenance medium hereafter) diluted 1:333 and 1:1000, respectively. Cultures were routinely assessed for morphological signs of differentiation by light microscopy and pluripotency markers assessed by qRT-PCR and/or immunostaining for/against NANOG and POU5F1.

For passaging, culture vessels were prepared with Cellartis COAT-1 as described before. Before inducing cell dissociation, the fully confluent monolayer of ChiPSC18 cells was washed once using warm DPBS without Mg^{2+} and Ca^{2+}. DPBS was aspirated and replaced with 0.015 mL per cm² of room-tempered TrypLE select (ThermoFisher), followed by incubation at 37°C and 5% CO_2 for approximately 5 min to obtain a single cell suspension. Upon loss of cell contacts, at least 9 volumes of full DEF-CS medium were added to the cell suspension, followed by centrifugation at 300g for 5min and resuspension in fresh full DEF-CS medium as required, before cell counting using Neubauer hemocytometer. In total, 40,000 cells were seeded in 0.2-0.3 mL of full DEF-CS medium per cm² of the culture vessels. Further details on the Cellartis® DEF-CS culture system, Cat no. Y30010, can be found in the manufacturer's instructions.

2.2.2 Differentiation of ChiPSC18 to definitive endoderm

Culture vessels to be used for the derivation of definitive endoderm from ChiPSC18 cells were coated with Definitive Endoderm Differentiation Coating Reagent (Cellartis), using 0.1 mL per cm² of the culture dish and incubation at RT for 60 min. Upon reaching full confluence and showing a dense monolayer, ChiPSC18 cells were harvested and counted. Next, the cell suspension was aliquoted as required for seeding of 34,000 cells/cm² and centrifuged at 300g for 5 min, the supernatant was aspirated, the pellet was dissolved in the required amount of Cellartis Definitive Endoderm D0 medium needed for inoculation (0.2 mL/cm²). The cell suspension was transferred and cells left for incubation at 37°C until the next day. Endoderm differentiation was induced according to schedule (Figure 1 A) over the course of 7 days, until further use in the HLC differentiation procedure. Further details on the Cellartis® Definitive Endoderm Differentiation system, Cat.no. Y30035, can be found in the manufacturer's instructions.

2.2.3 Differentiation of ChiPSC18-derived definitive endoderm to hepatocyte-like cells

Culture vessels were prepared either with Laminin 521 and Laminin 111 (Biolamina) in a 1:3 ratio, followed by a 1:20 dilution in PBS with Ca^{2+} and Mg^{2+}, or Cellartis® hepatocyte differentiation coating, using 0.1 mL per cm² of the culture vessels. Upon completion of the definitive endoderm differentiation schedule (day 7), prospective definitive endoderm cells were harvested by first washing the confluent cell layer with warm DPBS without Mg^{2+} and Ca^{2+}, followed by incubation with 0.1 mL/cm² TrypLE Select at 37°C, 5% CO_2 for approximately 2 min. To collect the cell suspension, one volume of DPBS containing 10 % FBS was added and the number of viable cells was determined after preparation of a 1:1 mixture of cell suspension and trypan blue, using a Neubauer chamber. Cells were centrifuged at 300g for 5 min and resuspended in Thawing and Seeding Medium (TS Medium, Cellartis) at $2.6*10^5$ cells/mL. Cells were seeded at a density of $1.3*10^5$ cells/cm² in 0.5 mL of TS medium/cm² and incubated at 37°C and 5% CO_2 for 24 h. Over the course of the following 18 days, cells were treated according to schedule (Figure 1 A). Upon completion of day 25, HLC were harvested for downstream applications. Further details on the Cellartis® Hepatocyte Differentiation System, Cat.no. Y30050, can be found in the manufacturer's instructions.

2.2.4 Cultivation of primary human Hepatocytes

Cryopreserved primary human hepatocytes (PHH) were purchased from BioIVT. Three different donors were used for single-cell sequencing (Donor 1, 2 and 3), two for qRT-PCR (Donor 2 and 3) data generation and one for immunostaining (Donor 2). Cryopreserved PHH were cultured according to a published standard operating procedure (SOP) (Gu et al. 2018 supplement 2), with modifications. Lyophilized rat-tail collagen (Roche Diagnostics) was dissolved overnight in 40mL 0.2% acetic acid at 4 C. Each well of Sarstedt Standard Tissue Culture F 12-well plates was coated with 1 mL (250μg/mL) collagen solution. Collagen was then aspirated and plates were left to dry overnight under sterile conditions. Before use, coated plates were washed three times with DPBS. PHH were thawed in a water bath at 37 °C and transferred to a Falcon tube with culture medium (Williams E, 1% P/S, 10 μg/mL gentamycin, 10 ng/mL ITS, 2 mM L-glutamine, 100 nM dexamethasone, 10% FCS). After cell counting using Trypan blue to determine viability, approximately 65k cells/cm² were plated followed by incubation at 37 C for at least 3h. Culture dishes were gently moved to facilitate equal distribution during the first 15 minutes of attachment. After 3h, cells were washed with warm PBS three times, followed by addition of 2 mL culture medium per well of a 12-well plate. Cells were then incubated at 37 °C, 5% CO_2 with medium changes every other day until further use.

2.2.5 Cultivation of CACO-2 cells

CACO-2 cells were obtained from ATCC. Cryopreserved cells were thawed in a water bath at 37 °C and resuspended in 5 mL of Dulbecco's modified eagle medium, followed by centrifugation at 300g for 5 minutes. Medium was aspirated and the cell pellet dissolved in 10 mL of fresh medium. Cell count was determined in a Neubauer chamber using trypan blue to account for viable cells. In a T75 cell culture flask, 50k cells/cm² were seeded, followed by incubation at 37°C and 5% CO2 until further used. Fresh medium was provided twice per week and cells passaged before reaching 80% confluence. Before processing, cells were allowed to reach more than 80% confluence while avoiding overgrowth.

2.2.6 Immunocytochemistry

HLC were fixed with 4% PFA for 15 min at 37°C. Fixed cells were washed with PBS on an orbital shaker 3 times for 5 min, membranes were permeabilized with Triton-X 100 for 10 min, washed 3 times for 5 min with PBS and blocked for 1 h with blocking solution containing 10% BSA and 0.1% Tween-20 in PBS. Subsequently, fixed cells were treated overnight with antibody solution containing 3% BSA, 0.1% Tween-20 and primary antibody. Primary antibodies used were anti-*ALB* (HPA031024, 1:200, Sigma), anti-AFP (MAB1369), anti-*CDX2* (ab114247, 1:200, Abcam), anti-*ISX* (HPA060328, 1:200, Atlas Antibodies), anti-*AGR2* (HPA007912, 1:200, Atlas Antibodies), anti-*DPPIV* (ab215711, 1:200, Abcam) and anti-HNF4A (HPA004712, Atlas Antibodies) The next day, cells were washed 3 times for 5 min on an orbital shaker, secondary antibody solution containing 3% BSA, 0.1% Tween-20 and secondary antibody in PBS was added, followed by 2 h incubation under light protection. Secondary antibodies used were Alexa Fluor® 488 Donkey anti-Rabbit (1:100), Alexa Fluor® 488 Donkey anti-Mouse and Cy3-AffiniPure Donkey anti-Rabbit (all Jackson ImmunoResearch Laboratories). Nuclear staining was performed using DAPI solution 1:10000 either alone or in combination with actin staining using phalloidin-rhodamine (ThermoFischer) 1:5000 for 10 min at RT. Cells were washed 3 times with PBS before imaging and storage at 4°C.

For high throughput screening of the immunohistochemistry time series of HLC differentiation, cell populations were fixed as described above at the respective time points. After fixation, cells were washed three times with PBS, followed by addition of 0.2 mL/cm² PBS with 10% FBS. The preparations were then stored at -20°C until simultaneous staining according to the described procedure, to reduce technical variability. High throughput screening was performed on the ImageExpress Micro XLS system (Molecular Devices) at 20x magnification. Image analysis was performed using MetaXpress software (Molecular Devices) and processed using R for statistical programming. Briefly, per antibody a setup was chosen to isolate stained area of at least 32 sites (16 sites per well of a 12 well plate) per time point of the imaging series. In addition, ICC controls without primary antibody were analyzed at 2 time points of the series (iPSC day 0, HLC day 25),

to guarantee primary antibody specificity. Intensities at the wavelength corresponding to secondary antibody emission (520 nm for Alexa Fluor 488) were extracted from masked areas for
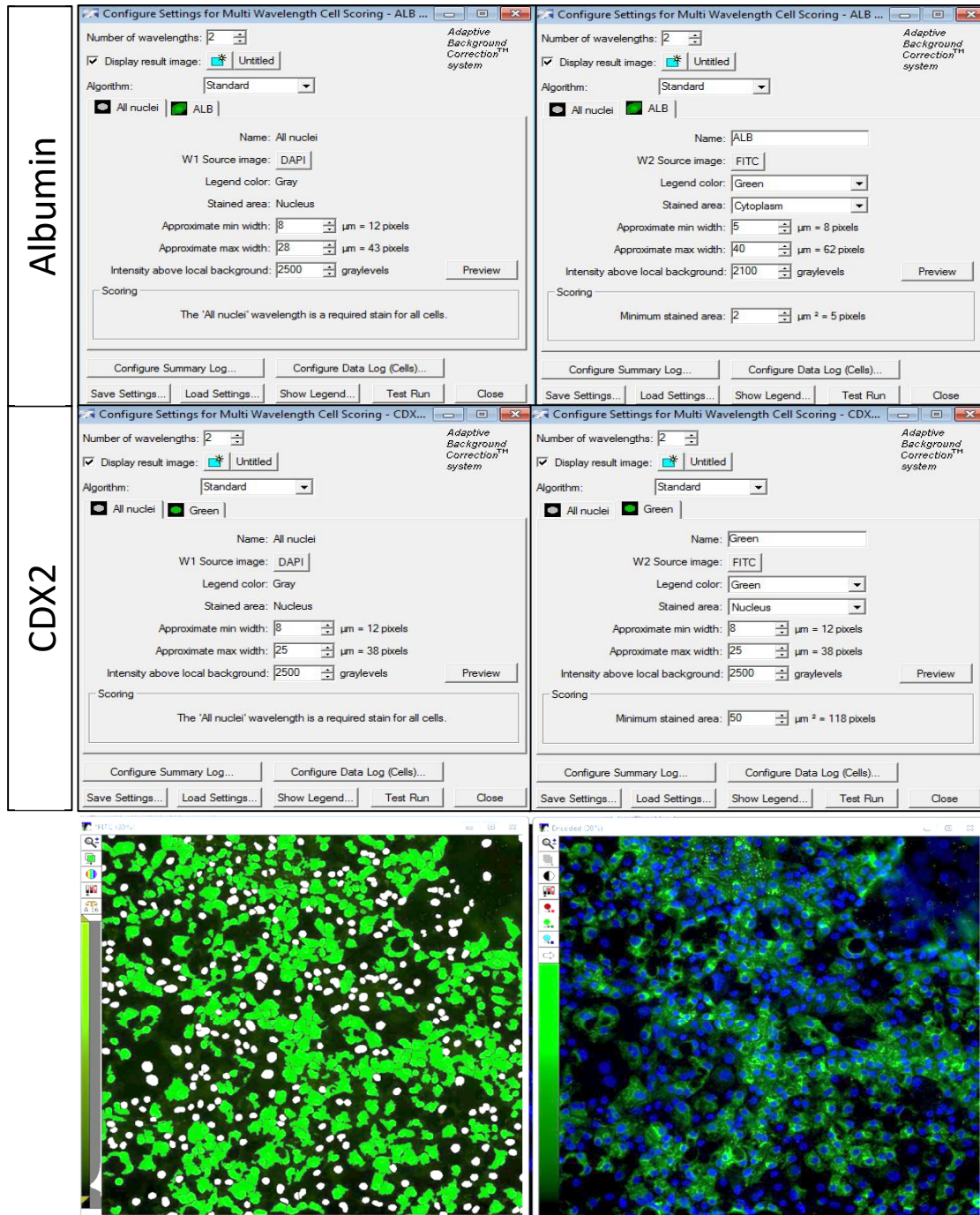
Figure 9: Image processing for high-throughput analysis. A) Masking cutoff parameters for image processing with the multi-wavelength analysis tool of the MetaXpress software (Molecular Devices), showing minimum and maximum width for nuclear signals, as well as cytoplasmic signal in albumin (ALB) and CDX2 signal quantification. **B)** Example of the masking result for albumin.

cytoplasm or nucleus respectively and control intensity was subtracted. Fractions of positive (intensity >= 0.001 RLU) and negative (Intensity = 0 RLU) cells were determined per time point.

2.2.7 Bile canalicular excretion assay live cell imaging

HLC were differentiated and PHH were seeded according to the previously described procedures. On day 25 for HLC and day 3 for PHH, cells were loaded onto a Zeiss LSM 880 confocal imaging system into an incubation chamber providing stable 37°C and 5% CO_2 under humid conditions. Cells were left for equilibration for at least 15 min before treatment with 5 µM 5-chloromethylfluorescein diacetate (CMFDA) and 5 µM 5-(and -6)-Carboxy-SNARF-1 from 10 mM stock solutions in DMSO diluted 1:2000 in medium upon addition to the cell cultures, corresponding to 2 mL of medium in total per well of a 12-well Sarstedt Standard Tissue Culture Plate. Imaging was started immediately after addition, with an excitation wavelength of 488 nm for CMFDA and 561 nm for SNARF-1. Emission was monitored at 525 nm and 629 nm for CMFDA and SNARF-1, respectively, in order to avoid bleed-through of the emission spectra. A time-series of 50 images was captured in 15 min intervals in 5 z-planes for PHH and HLC respectively.

2.2.8 Preparation of oleic acid solution and cell treatments

Oleic acid (OA) and fatty acid-free bovine serum albumin (BSA) were purchased from Sigma Aldrich. OA was complexed to BSA at a 6:1 molecular ratio by mixing 3 mL of a 60 mM OA solution with 12 mL of a 2.5 mM BSA solution as previously described (Cadenas et al., 2019). Cells were incubated with OA/BSA or BSA-only at the indicated concentrations for 48 h by adding the corresponding amounts to fresh cell culture medium.

2.2.9 Lipid droplets staining

Cells were plated on 12-well plates on a collagen-coated coverslip. The cell culture medium was removed, and cells were fixed for 20 min with 4% PFA and subsequently permeabilized with 0.3% Triton-X-100 for 10 min. To label the actin cytoskeleton cells were incubated with rhodamine-labeled phalloidin (Invitrogen), diluted 1:400 in 0.3% BSA, for 2 h. To visualize lipid droplets, cells were incubated with BODIPY® 493/503 (Life technologies), diluted 1:1000, for 45 min. The cell

nuclei were labeled with DAPI, diluted 1:5000, for 10 min. All incubation and washing steps were performed at room temperature. 2-3 washing steps of 5 min each between treatments were performed with PBS. Fluorescence was visualized by confocal laser scanning microscopy (FluoView™FV1000 Scanning Unit with BX61 Automated Research Microscope, Olympus, Hamburg, Germany).

2.2.10 Lentiviral transduction

The self-inactivating lentiviral vector construct pLV[Exp]-EGFP:T2A:Puro-EF1A>hNR1H4[NM_001206993.1] carrying ORFs for EGFP and FXR as well as the resulting lentiviral particles were generated in collaboration with VectorBuilder (Chicago, USA) as previously described (Tiscornia et al., 2006. Nature Protocols, 1(1), 241–245). A vector summary and vector map are given in Figure 2. By quantitative PCR of a fragment in the ENV region of the lentiviral vector amplified from genomic DNA of transduced HEK 293 cells viral titer was determined to be 2.38×10^9 transducing units / mL. Cells targeted for transduction were differentiated according to the described protocol. On day 13 of the differentiation HLC were exposed to lentiviral particles for 16 h at a multiplicity of infection (MOI) of 10 and a polybrene concentration of 5 µg / mL. Transduction efficiency was assessed by EGFP fluorescence. At day 25 of the differentiation, cells were collected according to the methods for total RNA isolation.

Vector ID	VB180912-1194han
Vector Name	pLV[Exp]-EGFP:T2A:Puro-EF1A>hNR1H4[NM_001206993.1]
Date Created (Pacific Time)	2018-09-11
Size	10835 bp
Vector Type	Lentivirus gene expression vector (3rd generation)
Inserted Promoter	EF1A
Inserted ORF	hNR1H4[NM_001206993.1]
Inserted Marker	EGFP:T2A:Puro
Plasmid Copy Number	High
Antibiotic Resistance	Ampicillin
Cloning Host	Stbl3 (or alternative strain)

Vector Map
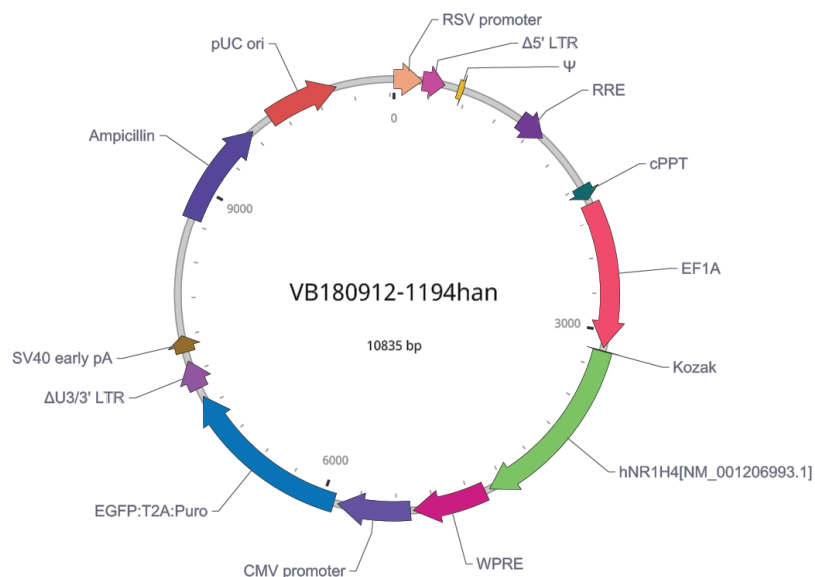
Figure 10: Vector information and map for exogenous expression of FXR in HLC.

2.2.11 Agonist treatment

For agonist treatments of FXR in HLC, a 100 mM stock solution of chenodeoxycholic acid (CDCA, Sigma) in DMSO and a 1.5 mM stock solution of GW4064 (Sigma) in DMSO were prepared. With medium changes on day 22 and day 24 cells were exposed to CDCA and GW4064 by adding 1.85 µL of stock solution to 1.85 mL of cell culture medium to achieve final concentrations of 100 µM and 1.5 µM, respectively. At day 25 of the differentiation cells were collected in Qiazol (Qiagen), sonicated and prepared for RNA isolation.

2.2.12 Total RNA isolation and cDNA synthesis

Total RNA was collected using Qiazol (Qiagen) and sonicated at 5 sec intervals with 2 sec break at 50 % intensity with a sonicator.(SONOPULS mini20, Bandelin). RNA was isolated from iPSC (D0, n=5), DE (D7, n=5), HLC (D9, D11, D14, D16, D18, D21, D23, D25, n=5) and PHH (D0, D1, D3, n=2, corresponding to two human donors). For cDNA synthesis, 1 µg of RNA was transcribed using a high-capacity cDNA Reverse Transcription Kit (ThermoFischer Scientific) according to manufacturer's instructions.

2.2.13 Quantitative Real-Time PCR

Quantitative real-time PCR (qRT-PCR) was performed on a 7500 Real-Time PCR System (Applied Biosystems) using 25 ng of cDNA, TaqMan Universal Master Mix (ThermoFischer Scientific) and the following primer probes from ThermoFischer Scientific: *ALB* (Hs00609411_m1), *HNF4A* (Hs00604435_m1), *HNF1A* (Hs00167041_m1), *NR1H4* (Hs01026590_m1), *ABCB11* (Hs00994811_m1), POU5F1 (Hs04260367_gH), *NANOG* (Hs04399610_g1), *CXCR4* (Hs00607978_s1), *SOX17* (Hs00751752_s1), *FOXA2* (Hs00232764_m1), *CDX2* (Hs01078080_m1), *KLF5* (Hs00156145_m1), *ISX* (Hs01368145_m1), *TWIST* (Hs01675818_s1), *SNAI1* (Hs00195591_m1). All qRT-PCR reactions were performed under the following conditions: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C for all PCRs. For qRT-PCR analysis the $2^{-\Delta\Delta Ct}$ method was applied, using GAPDH as a housekeeping gene and normalizing expression levels in reference to the iPSC population at day 0 of the differentiation. The results correspond to the mean of 5 independent biological replicates unless otherwise indicated.

2.2.14 Preparation of single cell suspensions and isolation of single cells

Cryopreserved primary human hepatocytes (PHH, male, obtained from BioIVT) were thawed in a water bath at 37 °C, transferred into 5 mL PBS supplemented with 10 % FBS. Day 25 adherent HLC were washed twice with Dissociation Buffer 1 (29 mM glucose, 10 mM NaCl, 1.8 mM KCl, 0.9 mM KH_2PO_4 (pH 7.4), 24 mM HEPES (pH 8.5), 14 % MEM Non-Essential Amino Acids Solution, 4.5 mM

glutamine, 0.5 mM EGTA (pH 7.6)) and incubated for at least 15 min in Dissociation Buffer 2 (31 mM glucose, 12% MEM Non-Essential Amino Acids Solution, 2.4 mM glutamine, 100 mM NaCl, 2 mM KCl, 1 mM KH_2PO_4, 25 mM HEPES (pH 8.5), 5 mM $CaCl_2 \times H_2O$) containing 1.5 mg/mL collagenase from *Clostridium histolyticum* (> 125 CDU / mg, Sigma) at 37 °C. As soon as the cells of the monolayer began to detach, they were scraped with a cell scraper and pipetted up and down 10 times to support disintegration of cell chunks. Finally, HLC suspensions were transferred to 1.5 mL tubes, centrifuged for 5 min at 300xg and 4°C, washed twice in cold PBS and resuspended in PBS containing 0.1 % polyvinylpyrrolidon. Single cells were manually collected by Konstantin Lepikhov (Saarland University) into 2 µl lysis buffer consisting of 2 U RNasin, (Promega), 0.2% Triton X-100 and 1 mM DTT.

2.2.15 Single cell RNA-seq library preparation

Preparation of cDNA was performed as previously described [163] with modifications by Kathrin Kattler. To each well 1 µL 10 mM dNTPs, 0.5 µL Oligo-dT primer (AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN) and 0.5 µL ERCC RNA Spike-In Mix (Invitrogen, diluted 1:400,000) were added. Then, samples were incubated at 72°C for 3 min and immediately placed on ice. Reverse transcription was carried out using 0.5 µL SuperScript II RT (200 U / µL, Invitrogen) supplemented with 0.25 µl RNasin (40 U / µL, Promega), 2 µL Superscript II first-strand buffer (5 X), 0.48 µL 100 mM DTT, 2 µL 5 M betaine, 0.12 µL 0.5 M $MgCl_2$, 0.1 µL 100 µM TSO (Biotin-AAGCAGTGGTATCAACGCAGAGTACAT997) and 0.25 µL nuclease-free water and incubation at 42°C for 90 min, followed by 10 cycles of 50°C for 2 min and 42°C for 2 min. Enzyme deactivation was achieved by incubation at 70°C for 15 min cDNA was pre-amplified by addition of 12.5 µL KAPA HiFi HotStart ReadyMix PCR Kit (Roche), 0.25 µL IS PCR primer (AAGCAGTGGTATCAACGCAGAGT) and 2.25 µL nuclease-free water with the following thermo-cycling conditions: 98 °C for 3 min, 18 cycles of 98°C for 20 sec, 67°C for 15 sec, 72°C for 6 min and final elongation at 72°C for 5 min. cDNA was purified using 0.8 X AMPure XP Beads (Beckman Coulter). Library preparation was performed using the Nextera XT kit (Illumina) in a reduced volume (0.25 ×) with 9 cycles for library amplification.
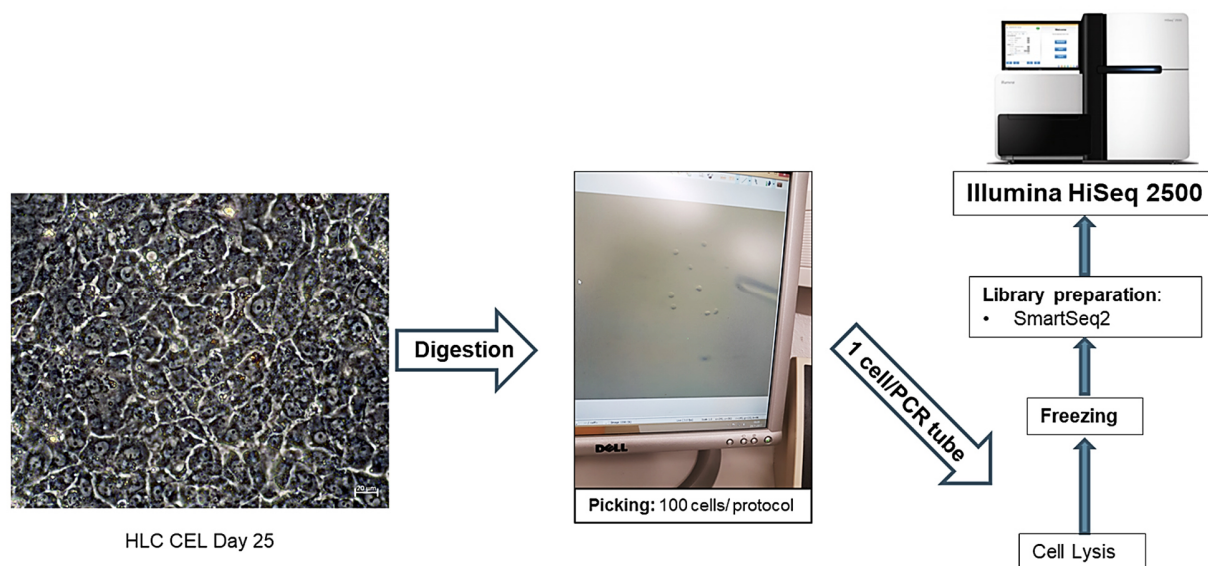
Figure 11: Single cell isolation and sequencing workflow. HLC were first digested, washed and then manually picked under the microscope. Individual cells were transferred to PCR tubes for cell lysis and freezing before library preparation and sequencing.

2.2.16 mRNA-seq library preparation

Total RNA was extracted from cells stored in 1 mL TRIzol (ThermoFisher) using the Direct-zol RNA MiniPrep Plus Kit (Zymo Research) including DNaseI treatment. Library preparation was performed as described for single cells with modifications. Briefly, at least 100 ng total RNA were used for reverse transcription with 5 cycles of pre-amplification. Libraries were prepared using the Nextera DNA Library Prep Kit (Illumina) with 8 cycles of enrichment PCR followed by 0.9 × Ampure XP Beads (Beckman Coulter) purification.

2.2.17 Reduced Representation Bisulfite Sequencing (RRBS)

Genomic DNA was isolated by phenol chloroform isoamyl alcohol extraction and purified by ethanol precipitation. RRBS libraries were prepared as previously described (Boyle et al., 2012. Genome Biology, 13(10), R92.), with modifications. In brief, at least 180 ng DNA were digested using 1 µL HaeIII (50 U, NEB, covers approximately 7 million CpGs) in a 30 µL reaction volume supplemented with 3 µL 10 × CutSmart Buffer (NEB) at 37 °C. After 2 hours 0.5 µL HaeIII (50 U/µl) were added and incubation was completed for additional 16 hours. A-tailing was performed by

addition of 1 μL Klenow Fragment (3`→5` exo-, 5 U/μL, NEB) and 1 μL10 mM dATP at 37°C for 30 min, followed by enzyme inactivation at 75°C for 20 min Ligation of sequencing adapters (10 μM, TruSeq DNA Single Index Set B, Illumina) was achieved by incubation with 2 μL adapter, 1 μL T4 Ligase (2,000,000 U per μL, NEB), 1 μL 10 X Cutsmart Buffer (NEB) and 4 μl ATP (10 mM, NEB) at 16°C for 22 h. After bisulfite conversion with the EZ DNA Methylation-Gold Kit (Zymo Research) DNA was eluted in 10 μL M-elution buffer (Zymo Research). Enrichment PCR of adapter ligated fragments was performed using 3 μl PCR Buffer (10×, Qiagen), 1.2 μl MgCl$_2$ (25 mM, Qiagen), 2.4 μl dNTPs (10 mM), 0.5 μl Hot Start Taq (5 U/μl, Qiagen), and each 0.5 μL primer 1 and 2 (10 μM, 1: AATGATACGGCGACCACCGAGATCTACAC, 2: CAAGCAGAAGACGGCATACGAGAT) in a final volume of 30 μl. After initial denaturation at 95°C for 15 min, thermocycling was carried out for 22 cycles at 95°C for 30 sec, 60°C for 30 sec and 72°C for 1 min, followed by final elongation at 72°C for 7 min Final library purification was carried out with 0.8 × Ampure XP Beads (Beckman Coulter).

2.2.18 ATAC-seq library preparation

ATAC-seq was performed as described earlier [164] with modifications. Fresh cells were resuspended in 1 mL Permeabilization Buffer A (60 mM KCl, 15 mM Tris-HCl (pH 8.0), 15 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine (free base), Protease Inhibitor Cocktail Tablets (Roche) supplemented with 0.05 % Nonidet P-40 and incubated on ice for 15 min. Approximately 50,000 nuclei were centrifuged, washed with Permeabilization Buffer A, and finally resuspended in 50 μl tagmentation mix containing 2.5 μl Tn5 (Nextera DNA Library Prep Kit, Illumina) in 1 X TD buffer. Tagmentation was carried out at 37°C for 30 min Libraries were amplified for 9 cycles and purified using 1 × Ampure XP Beads (Beckman Coulter).

2.2.19 Next Generation Sequencing (NGS)

Libraries were sequenced by Gilles Gasparoni (Saarland University) on a HiSeq2000 sequencer (Illumina) using TruSeq SBS Kit v3 – HS Chemistry in single read runs with read lengths of 94 bp (RNA-seq, ATAC-seq and RRBS) or 88 bp (scRNA-seq, dual index).

2.2.20 NGS Data processing

Sequencing data was pre-processed by Abdulrahman Salhab (Saarland University) and Kathrin Kattler. Reads were trimmed using Trim Galore! (v0.4.2) http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove 3' ends with base quality below 20 and adapter sequences. RRBS reads were trimmed in RRBS mode and aligned to GRCh38 with the BWA [165] wrapper methylCtools [166]. Samtools [167] and Picard tools (v1.115) (http://broadinstitute.github.io/picard/) were utilized for converting, merging and indexing of alignment files. Bis-SNP was used for SNP (dbSNP, v138, http://www.ncbi.nlm.nih.gov/SNP) aware realignment, quality recalibration and computation of methylation calls. ATAC-seq reads were aligned to GRCh38 with the GEM mapper [168], and samtools were used to convert SAM to BAM format. MarkDuplicate (version 1.115) from Picard tools (http://broadinstitute.github.io/picard/) was used to mark PCR duplicates. ATAC-seq peaks were called using MACS2 [169], applying the following parameters: --nomodel, --shift -125, --extsize 250. Coverage files normalized for library size were generated using bamCoverage from deepTools [170]. Normalized read counts in 100 bp and 10,000 bp bins were counted using featureCounts [171]. Bulk RNA-seq and scRNA-seq reads were aligned to GRCh38 using STAR [172] with per sample 2-pass mapping strategy. PCR duplicates were detected using MarkDuplicate from Picard tools (version 1.115; http://broadinstitute.github.io/picard/). Gene-wise read counts for RNA-seq data were estimated based on Gencode release 30 (GRCh38.p12) using RSEM [173].

2.2.21 RNA-seq data analysis

Expression values were normalized as log CPM + 1 (counts per million). Genes with average log CPM < 0.1 were excluded. The 1000 most variable genes were used for Principal Component Analysis (PCA), in which data are projected to a 2-dimensional space representing the two directions with largest variance. DEseq was used for calculation of scaling factors and robust estimation of dispersion in order to detect differentially expressed genes [174]. Supervised clustering of DPGs was performed based on differentially expressed genes. Briefly, DPGs were obtained by defining gene expression change patterns that describe the change and directionality

of log$_2$ fold change over PHH that occurs during iPSC to HLC differentiation, thus relating to a certain biologically interpretable behavior, e.g. 'insufficient upregulation' or 'adverse upregulation'. Figure 3 illustrates supervised clustering with the DiPa procedure. Tissue identity enrichment analysis was performed using TissueEnrich v1.8 for gene sets determined by the supervised clustering approach [175]. Gene ontology enrichment analysis was performed using enrichGO from the package Clusterprofiler v3.16 [176]. Overrepresentation analysis for transcriptional regulators was performed using the web-based RegulatorTrail pipeline using the entire database for regulator-target interactions [177]. Reactome pathway enrichment for gene sets was performed using ReactomePA v 1.16.2 [178]. CellNet analysis based on the reconstruction of cell type-specific gene regulatory networks (GRNs) was performed by Birte Hellwig using the R package CellNet [159], with a modification regarding zero counts. In CellNet, the GRN status is calculated based on a weighted sum of z-scores of the genes belonging to the gene regulatory networks. The value of GRN status ranges from 0 to 1. A high value means that the activity of genes in the analyzed dataset is similar to that in the respective tissue type. To estimate the importance of transcriptional regulators for dysregulated GRNs, the Network Influence Score (NIS) is used. This is also a weighted sum of the z-scores of the genes of the networks of a tissue type, with an additional weighting of the transcription factor of the network by the number of target genes in the network. Thus, a high absolute value of the NIS of a GRN indicates a large deviation in gene expression of that network, compared with the target tissue. A problem with the use of z-scores arises when genes of a network are not expressed in a tissue type. These genes then have a standard deviation of 0 and the z-score cannot be calculated accordingly, as in the denominator one divides by 0. However, these networks are of interest, as their expression is obviously highly divergent from the target tissue. Thus, a fudge factor approach is used in the calculation of z-scores. Here, a fixed value 1 is added to all standard deviations. Fisher tests were performed by Birte Hellwig to quantify the statistical significance for comparisons of fractions of differentially expressed genes in DPGs. The overlap ratio was used to quantify to which degree genes in an overlap are overrepresented. For pairwise overlap ratios the score is calculated as $OR=(O*N)/(n\_1*n\_2)$, where N represents the total number of genes, n_1 and n_2 the number of differentially expressed genes under condition 1 and condition 2, respectively, and O the number of genes in the overlap. A value of 1.0 indicates a random overlap,

and values higher than 1.0 are indicative of an overlap that is higher than expected by chance in case of independence.

2.2.22 RRBS data analysis

Detection of differentially methylated regions was carried out by Kathrin Kattler using MethylKit [179], which was modified to treat missing sites as zero-covered sites in the aggregation step of tiled analysis (500 bp tiles). After merging both strands, calls were filtered to sites with at least 10× coverage present in at least 75% of samples per group. Significant DMRs were defined as 500 bp tiles covering at least 3 CpGs and displaying a maximum FDR adjusted p-value of 0.01 and minimal methylation difference 20%. Mitochondrial, X and Y chromosomes were excluded. DMRs were annotated to closest genes and genomic features based on Gencode release 30 gene models [180] using GenomicRanges [181].

2.2.23 ATAC-seq data analysis

Bin-wise counts were normalized as log (CPM + 1) and the 50,000 most variable 100 bp bins were selected for PCA. For differential chromatin accessibility analysis, the mitochondrial chromosome and a set of blacklist regions were excluded. Differential analysis was performed using csaw [182] combined with edgeR [183]. To adjust background differences between samples, 10,000 bp bins were used for calculation of normalization factors. After initial edgeR based differential analysis, windows belonging to the same peak region were merged, allowing a maximal peak size of 3,000 bp and a maximum distance of adjacent bins of 150 bp. Then, FDR for the merged window was calculated based on Benjamini Hochberg adjustment. Annotation to the nearest gene was achieved as described for DMRs.
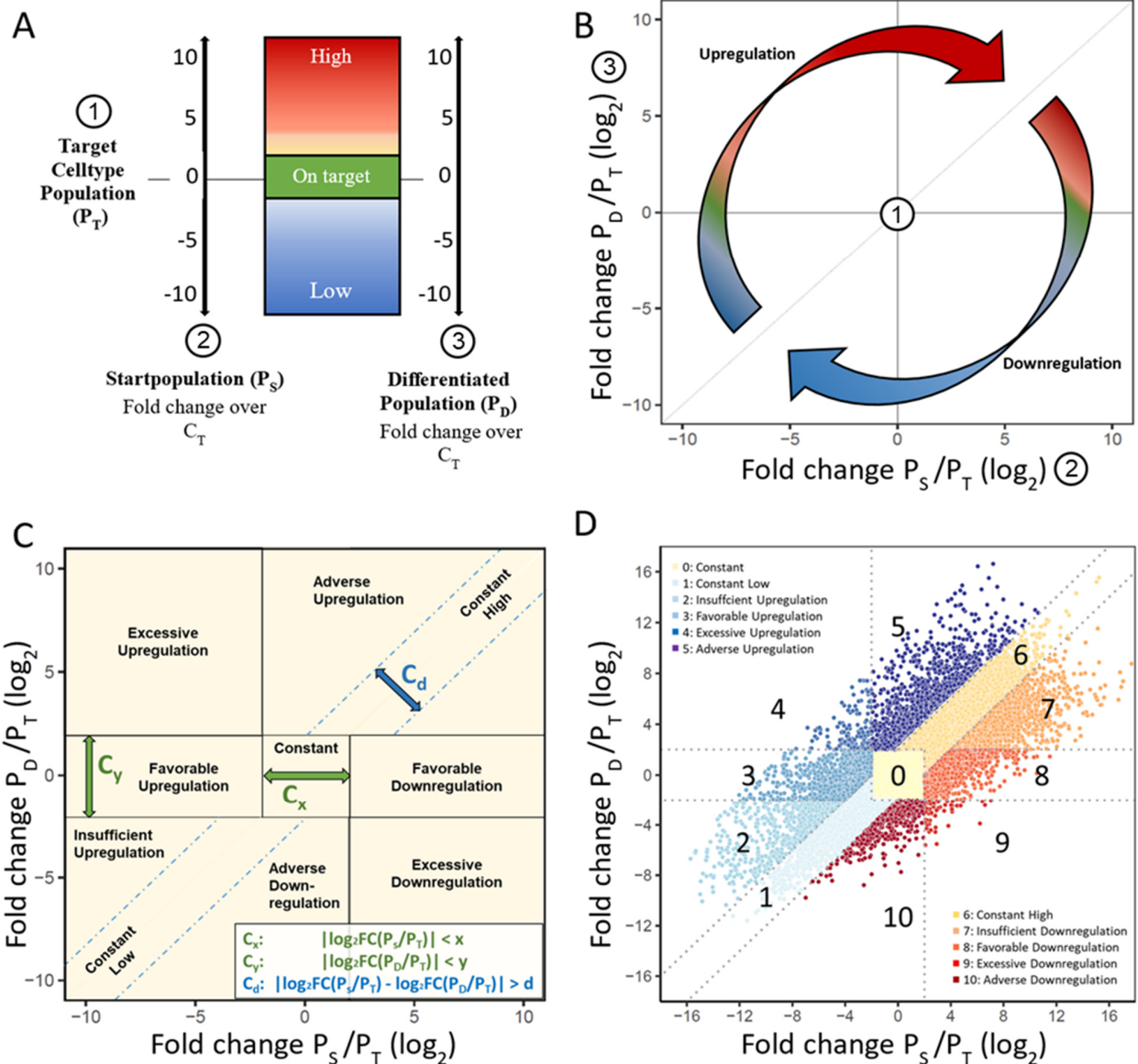
Figure 12: Supervised clustering of differentiation patterns (DiPa). A) Differentially expressed genes are determined for the starting population, and differentiated population, in comparison to the reference (here PHH). **B)** Log2 fold changes over the target cell population are compared in two dimensions, where the starting population is presented on the x-axis and the differentiated population on the y-axis. On the left side of the diagonal, genes are upregulated during the differentiation process, while genes on the right side of the diagonal are downregulated. **C)** Supervised clustering is established by applying combinations of cutoffs that discriminate genes whose expression levels were close to the target population before differentiation (green, cutoff C_x) genes that after differentiation reach expression levels close to those of the target cell population (red, cutoff C_y) and genes whose expression values do not show considerable changes from starting population to differentiated population. **D)** Example DiPa plot from the current study, where the starting population (human induced pluripotent stem cells; hiPSC) and the differentiated population (hepatocyte-like cells; HLC) are compared to the target cell type population (primary hepatocytes; PHH).

2.2.24 Single cell RNA-seq data analysis

The dataset was extended by an external dataset [82], which we processed accordingly by Kathrin Kattler. PHH from both datasets were well-mixed in dimensionality reduction plots (Figure 1 C) and PCA regression did not indicate batch effects. Thus, no batch correction was performed. Quality filtering (gene count > 500, read count > 200,000, ratio of mitochondrial genes < 0.25), normalization, calculation of highly variable genes and ranked genes, and visualization was achieved using scanpy [184]. Hybrid state genes, were derived from the present sc-RNAseq data by the following definition: $\log_2 FC[HLC/iPSC] > 1$ and $\log_2 FC[HLC/PHH] > 1$. Regulon analysis was done using the python implementation of SCENIC [185] pySCENIC (https://github.com/aertslab/pySCENIC) as outlined in the package manual. Regulons with activity scores above AUC score threshold in at least 25 % of single cells in at least 1 celltype were considered further. Binarized regulon activities were used for hierarchical clustering and correlation based t-SNE visualization of single cells.

2.2.25 Integrative data analysis

Clustered heatmap of gene expression, DNA methylation and chromatin accessibility was generated using pheatmap (https://CRAN.R-project.org/package=pheatmap). Z-scores of expression (log CPM + 1), accessibility (average log CPM + 1) and DNA methylation (average methylation percentage were calculated by Kathrin Kattler separately for each assay (**Supplement 6**). Regulator analysis of clusters was performed using REGGAE (Kehl et al., 2018. Bioinformatics, 34(20), 3503–3510) based on expression logFC of associated genes.

2.2.26 Statistical analysis

Statistical analysis in this thesis was performed in close cooperation with the statisticians Birte Hellwig and Jörg Rahnenführer from the Technical University of Dortmund and Kathrin Kattler from Saarland University.

3. RESULTS

The results presented in this PhD thesis are part of a manuscript with the title 'Identification of an FXR-modulated liver-intestine hybrid state in iPSC-derived hepatocyte-like cells', which is currently undergoing revision at the Journal of Hepatology.

3.1 Stem Cell-derived Hepatocyte-like cells show hepatic and intestinal gene expression

Differentiation of human induced pluripotent stem cells (iPSC) via definitive endoderm (DE) to hepatocyte-like cells (HLC) was performed according to the proprietary protocol of Cellartis (former Cellectis), which is part of Takara Bio Europe, with modifications. Stem cells, DE and HLC were analyzed at different time-points by a comprehensive program of RNA- and protein-based techniques, as well as transcriptomics (RNA-seq and sc-RNA-seq) and epigenomics (RRBS, ATAC-seq) analysis to characterize the differentiation status of HLC in comparison to cryopreserved and cultivated primary human hepatocytes (PHH). These assays formed the basis for the design of a bioinformatics-guided intervention strategy to improve HLC differentiation (**Figure 13 A**).

HLC showed hepatocyte-like morphology after differentiation using the Cellartis® protocol (**Figure 13 B**). The cells were fixed on day 25 and immunostained with antibodies directed against the adult hepatocyte marker albumin (ALB) and fetal hepatocyte marker alpha-fetoprotein (AFP), which both showed strong cytoplasmic signals in the majority of HLC. The apical hepatocyte marker DPPIV visualized thin canaliculi between the HLC, which were also marked by F-actin as visualized by phalloidin-rhodamine staining (**Figure 14 A**). It was thus an interesting question whether HLC could demonstrate transport across the bile canalicular (apical) membrane and thus exhibit polarity, which represents a key attribute of PHH.

Canalicular excretion was active, as evidenced by 5-chloromethylfluorescein diacetate (CMFDA) and carboxy-SNARF-1 [186], which was comparable to observations made in cultivated PHH (**Figure 14 B**). CMFDA and CM-SNARF1 are ester-caged fluorophores that show up to 400-fold increase in fluorescence upon uncaging by cytoplasmic esterases. Primary hepatocytes excrete CMFDA and CM-SNARF1 into canaliculi through active transport [186] mediated by multiple hepatocyte-specific transporters such as MDR2 and BSEP. Export was readily present after 10 min
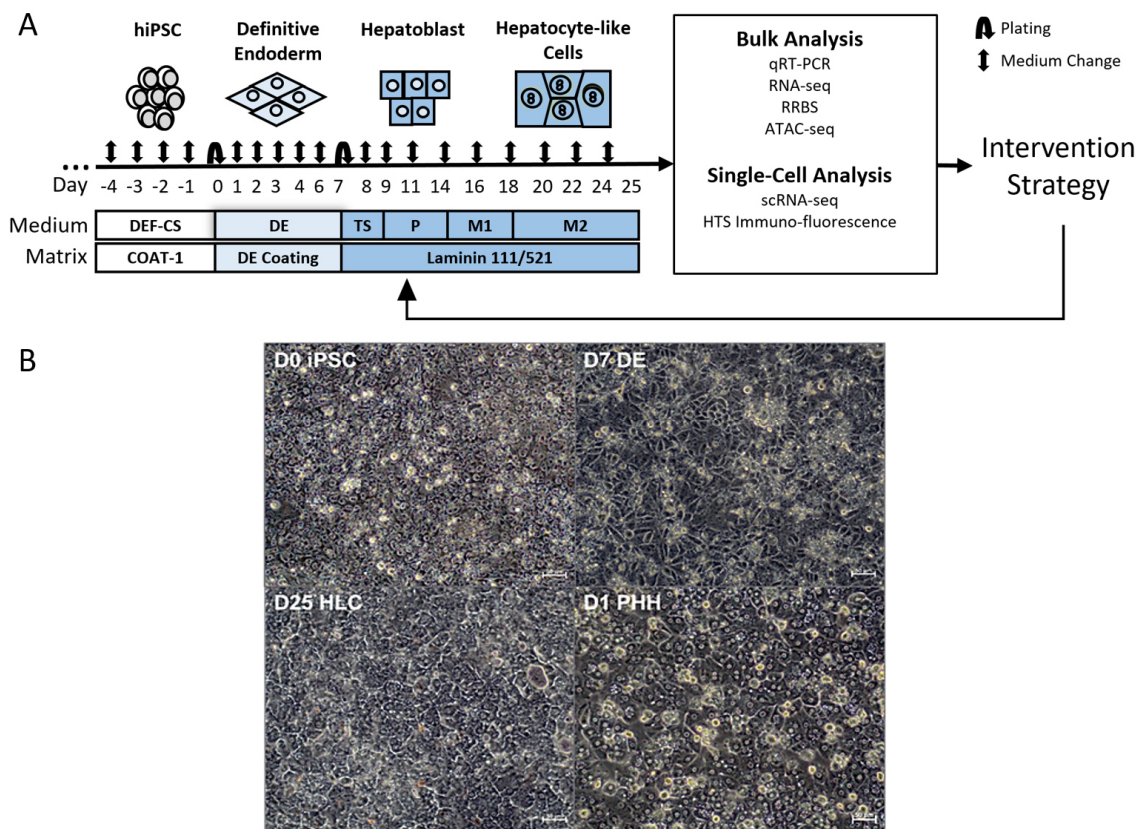
Figure 13: Experimental design and cell types. A) Schematic of the iPSC to HLC differentiation protocol and design of the study. Cells were differentiated for 25 days and collected at varying time-points for processing prior to bulk and single-cell level analysis that enabled development of a knowledge-based intervention strategy. **B)** Light microscopy of representative iPSC (Day 0), DE (Day 7), HLC (Day 25) and PHH (Day 1). Scale bar indicates 50 μm.

post addition in both HLC and PHH. Incubation of these compounds with hepatocyte-like cells led to efficient de-esterification and specific enrichment into canalicular structures, with kinetics similar to those of primary hepatocytes, indicating the presence of active transporters on their canalicular membrane. Corresponding video material for CMFDA and SNARF1 excretion into HLC and PHH canaliculi is available in **Supplementary file 1.1** and **1.2** of the digital appendix. HLC also formed structures with large lumina surrounded by multiple cells that showed accumulation of CMFDA and SNARF1, as well as network-like arrangements of CMFDA and SNARF1 accumulating ducts that spread throughout the HLC cultures (**Supplementary figure 1**, **Supplementary file 1.3** and **1.4**). Moreover, upon the addition of oleic acid to the culture medium, HLC formed lipid droplets, which demonstrated that HLC are capable of storing triacylglycerides (**Figure 14 C**) [187].
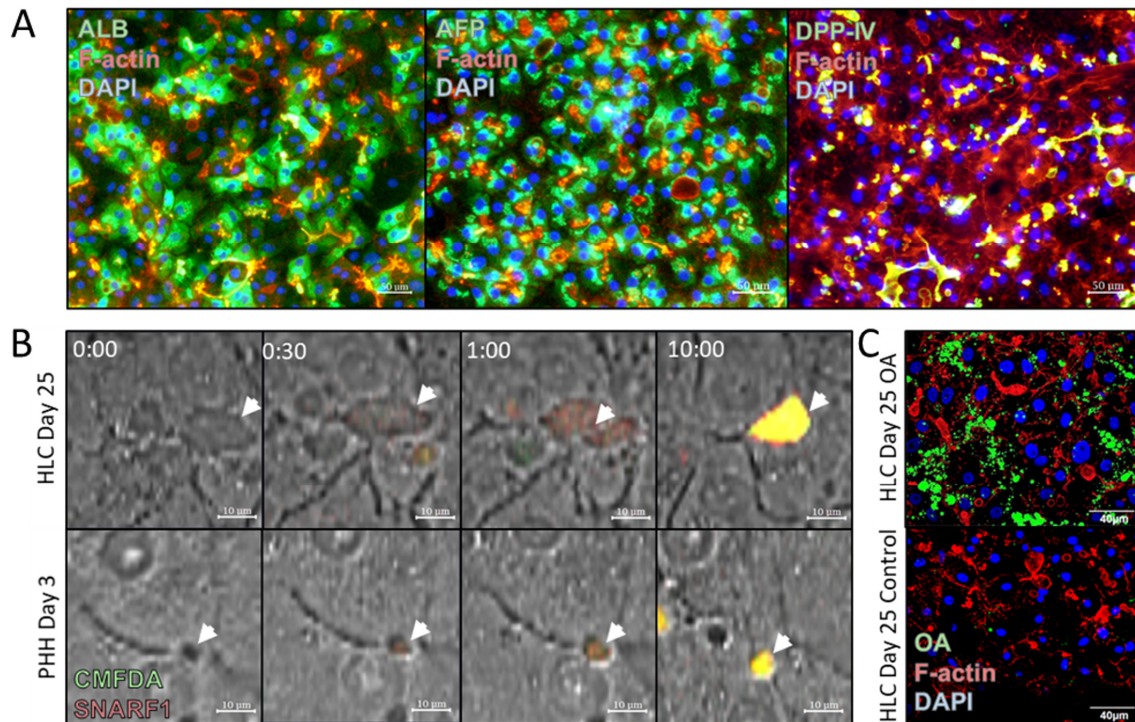
Figure 14: Phenotype and functional characterization of HLC. A) Fluorescence imaging of immunocytochemical stainings of ALB, AFP and DDP-IV (green), F-actin (phalloidin-rhodamine; red) and DAPI (blue) in day 25 HLC. **B)** Canalicular export of CMFDA and SNARF1. The time after addition of fluorophores is indicated. Exemplary bile canaliculi indicated by arrows. **C)** Lipid droplet accumulation visualized by fluorescent imaging of day 25 HLC treated with 0.8 mM oleic acid (OA) and untreated day 25 control HLC stained with BODIPY, phalloidin-rhodamine and DAPI.

HLC were further characterized by qRT-PCR using well-established marker genes (**Figure 15**). The pluripotency-associated transcription factors *POU5F1* and *NANOG* showed the highest expression levels in iPSC, which decreased during differentiation to levels below that observed in PHH, while markers of DE (*CXCR4, SOX17 and FOXA2*) showed stage-specific upregulation. In HLC, increased RNA expression was observed for the liver markers *ALB, HNF4, HNF1A, NR1H4* (FXR), and *ABCB11* (BSEP), although levels remained below those of freshly isolated and cryopreserved PHH. It should be noted that the expression of hepatocyte markers, particularly *ALB* and *ABCB11*, decreased rapidly upon the cultivation of PHH, demonstrating the importance of using freshly isolated cells as a reference. The expression of the intestine-associated transcription factors *CDX2, KLF5* and *ISX* significantly increased compared to the levels observed in PHH. A comparison of intestinal marker expression in the adenocarcinoma cell line CACO-2 and HLC showed similar expression

levels in contrast to PHH. Moreover, the fibroblast-associated genes *TWIST* and *SNAIL* decreased during differentiation; however, *SNAIL* still remained at higher levels in HLC than PHH.
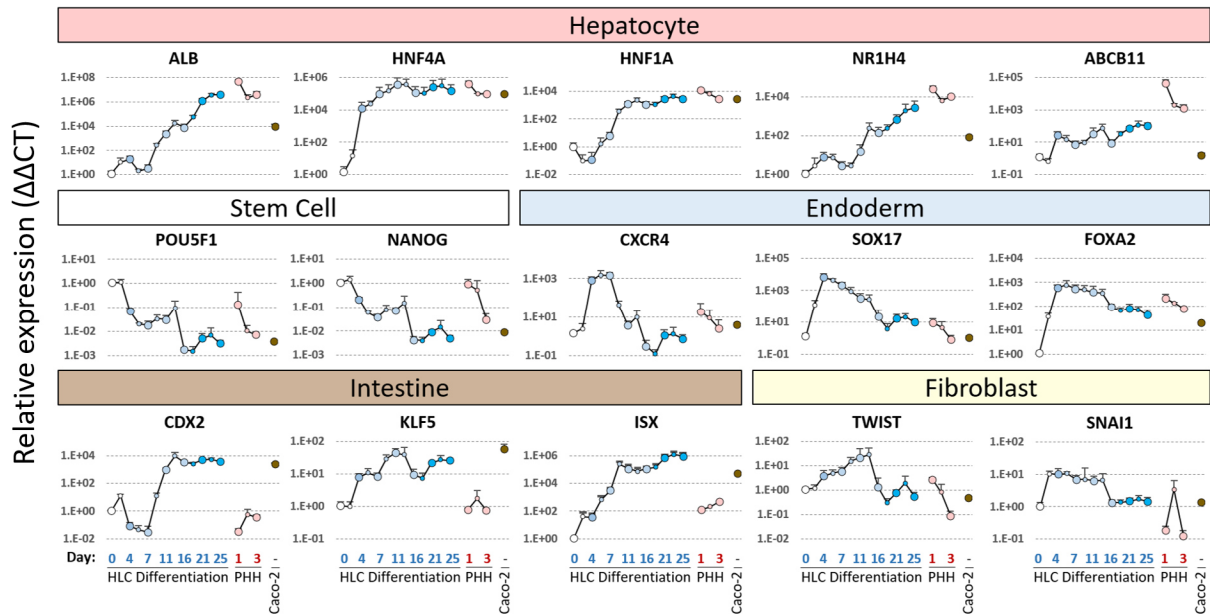
Figure 15: Time-resolved qPCR data for selected markers of hepatocytes, pluripotent stem cells, definitive endoderm, intestinal tissues and fibroblasts. The x-axis depicts the time period of HLC differentiation (d0-d25), freshly isolated hepatocytes (PHH) and PHH after cultivation periods of 1 and 3 days, as well as Caco-2 cells. The small data points represent day 2,6,9,14,18 and 23 of HLC differentiation and day 2 of PHH culture.

The characterization of HLC on RNA level thus suggests that iPSC have first successfully acquired the expression of definitive endoderm related genes and subsequently the expression of hepatocyte markers. In addition, HLC acquired functional properties characteristic of human hepatocytes, including canalicular excretion and lipid droplet formation. Nevertheless, the analysis also confirmed that unrelated intestinal marker expression was induced in the course of *in vitro* differentiation, as was previously described by Godoy et al. in 2015 in HLC differentiated in three laboratories employing different protocols [80]. Furthermore, the data suggests that upregulation of hepatic as well as intestinal markers may be induced as early as day 2 of the differentiation, as evidenced by the detection of an increase in *HNF4A* and *ISX* transcript expression.

3.2 Genome-wide analysis identifies target gene regulatory networks to improve HLC differentiation

Having characterized the phenotype of the stem cell-derived HLC on the basis of selected markers, RNA-seq on iPSC, DE, HLC and PHH bulk populations was performed in order to obtain global transcriptome profiles and characterize expression changes induced by the differentiation protocol. In a principal component analysis (PCA) of the top 1000 most variable genes, HLC located more than half-way from iPSC along PC1 towards primary hepatocytes, while DE clustered closely with iPSC (**Figure 16 A**).
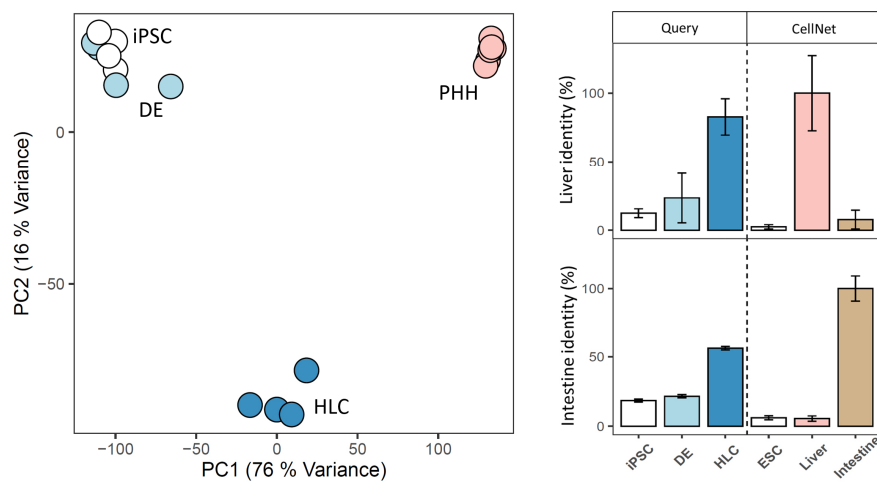
Figure 16: PCA analysis of variable genes and classification of differentiation state by CellNet. A) Principal component analysis (PCA) of the 1000 most variable genes among iPSC, DE, and HLC obtained by RNA-sequencing. **B)** CellNet analysis of RNA-sequencing data for liver and intestine tissue identities (GRNs) of the indicated cell populations in comparison to ESC, liver and intestine CellNet training datasets.

The similarity of iPSC, DE and HLC transcriptomes to the liver and intestine (colon) transcriptome training data was next analyzed using the CellNet algorithm [160] (**Figure 16 B**). Compared to the liver reference, iPSC, DE and HLC showed a gene regulatory network (GRN) status similarity of 12%, 23% and 83%, respectively. For the intestinal phenotype, iPSC showed 18%, DE 22% and HLC 56% similarity, further confirming that HLC express a substantial number of intestine-associated genes. The phenotype-specific reference data from CellNet illustrates that the high scores for the liver and intestine identity in HLC are not the result of shared genes between the liver and intestine gene sets.

Next, the DEseq2 pipeline was applied to identify differentially expressed genes (available as **Supplementary table 1**) and the resulting $\log_2$ fold changes were used to establish a novel supervised clustering approach (Differentiation Pattern Clustering; DiPaC), visualized in the Differentiation Pattern (DiPa) plot (**Figure 17**). Here, genes were plotted based on their expression fold changes of iPSC (x-axis) and HLC (y-axis) compared to the target population, PHH. Genes with similar expression changes between iPSC and HLC in relation to PHH were then categorized into eleven differentiation pattern groups (DPG0-10) (**Figure 17, Table 5)**, describing constant gene expression and insufficient, favorable, excessive or adverse gene up- and downregulation. This supervised strategy thus allowed clustering of the expression changes of genes in HLC in comparison to PHH based on the directionality and extent of the changes of the same genes during differentiation from iPSC to HLC.

Genes expressed at a similar level in iPSC, HLC and PHH are found in DPG0. These genes did not display expression changes induced by the differentiation protocol. Genes in DPG1 and 6 were expressed at a similar level in both iPSC and HLC and thus unaltered by the differentiation protocol, but expressed at too low (DPG1) or too high (DPG6) levels in HLC. In contrast, genes from DPG3 and 8 show favorable up- and downregulation, respectively, whereby favorable means that gene expression levels in HLC were close to that of PHH. The differentiation protocol thus achieved successful induction or inhibition of these genes. DPG2 and 7 genes were insufficiently up- and downregulated. Their expression is correctly induced (DPG2) or inhibited (DPG7) by the differentiation protocol, but fails to approach PHH levels. These DPGs thus represent especially interesting gene sets for identification of potential molecular intervention opportunities that aim at gene regulatory network activation, because HLC already demonstrated an initial positive response to the differentiation protocol. DPG4 and 9 contain genes that were excessively up- and downregulated, respectively. Genes in DPG5 and 10 were adversely up- or downregulated and therefore represent misguided differentiation. The DPG annotation for all DEGs is available in **Supplementary table 2**. Interactive versions of the DiPa plot in 2D and 3D, where DE is included, can be found in **Supplementary figures 2** and **3** of the digital appendix.
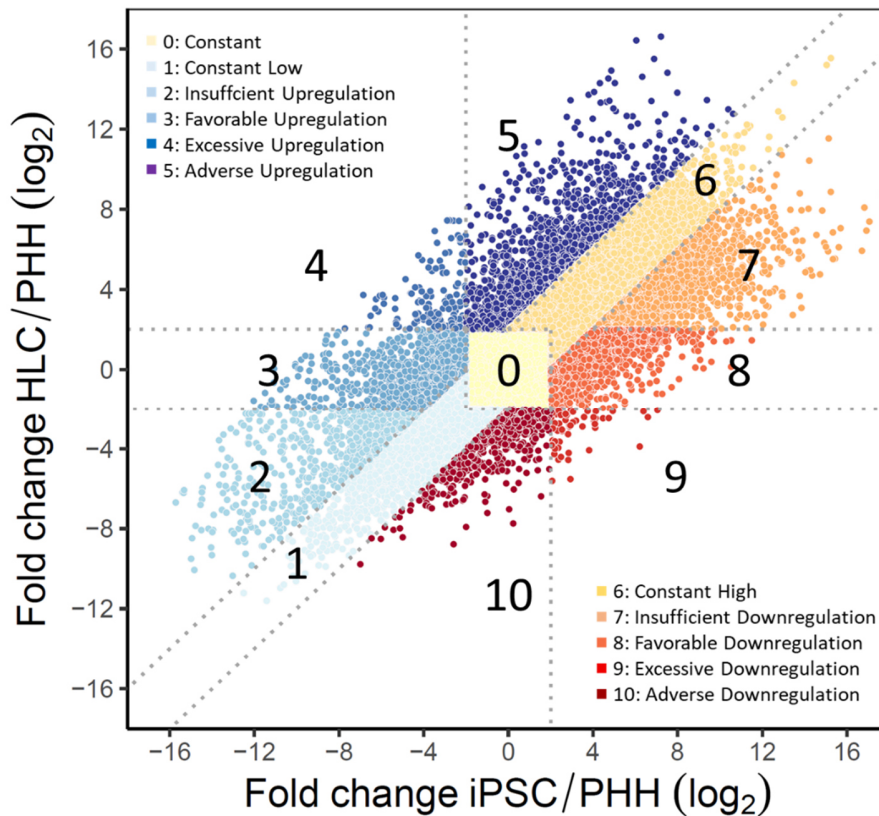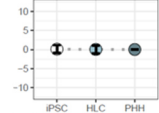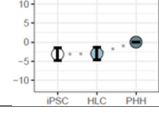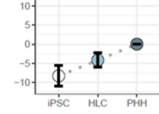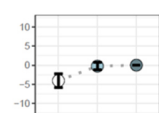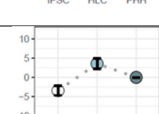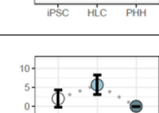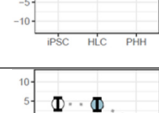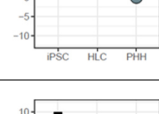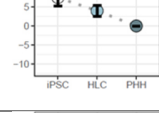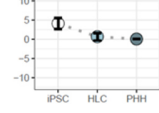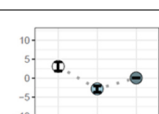
Figure 17: Differentiation Pattern Clustering (DiPaC) of differential gene expression data obtained from bulk RNA-sequencing, visualized in the differentiation pattern plot (DiPa plot). Colors indicate differentiation pattern groups (DPG). The x-axis represents $\log_2$ fold changes of iPSC over PHH, the y-axis indicates $\log_2$ fold changes of HLC over PHH. Dotted lines represent cutoffs of the clustering approach.

Several techniques were applied to characterize the genes in the individual DPGs, summarized in **Table 5**. An overrepresentation of liver-associated genes was observed in DPG1 (3,695 genes), 2 (624 genes) and 3 (1,330 genes) using TissueEnrich [175], in agreement with an overrepresentation of Gene Ontology (GO) [176] terms related to xenobiotic metabolism and Reactome pathways related to xenobiotic and bile acid metabolism in DPG2. In contrast, an enrichment of genes associated with intestine and gallbladder was found in DPG4 (142 genes), and RegulatorTrail overrepresentation analysis [177] indicated a strong enrichment of HNF1B binding sites, a transcription factor implicated in cholangiocyte differentiation [188]. DPG5 (1,712 genes) also showed enrichment of intestinal genes, with overrepresentation of transcription factor binding sites for CDX2, KLF5, PGR *and* GRHL2, as well as binding sites for the epigenetic regulator MBD4 and enrichment for digestion related GO terms and Reactome pathways.

Table 5: Enrichment analysis in DPGs obtained by DiPa supervised clustering. (Ranks)

	Mean Log$_2$ Fold Change over PHH	Tissue Group Enrichment (p-Value < 0.01)	Gene Ontology p.Adj. < 0.001	Regulatortrail Enrichment p-Value < 0.001	Reactome Pathway Enrichment p.Adj. < 0.01
DPG 0 „Constant"	(plot: iPSC, HLC, PHH)	----	ncRNA processing (1), Mitochondrial gene expression (2), Mitochondrial translation (3), mRNA catabolic process (4)	GLI2 (1) PPARG (2) RUNX1T1 (3) KDM2B (4)	Translation (1), Infectious disease (2), Eukaryotic Translation Initiation (3), Cap-dependent Translation Initiation (4)
DPG 1 „Constant Low"	(plot: iPSC, HLC, PHH)	Liver (1)	Macrophage activation (1), Lipid oxidation (2), Cellular Response to biotic stimulus(3), Fatty acid oxidation (4)	MBD4 (1) PMEPA1 (2) SUMO2 (3) KMT2A (4)	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell (1),
DPG 2 „Insufficient Upregulation"	(plot: iPSC, HLC, PHH)	Liver (1), Kidney (2), Gallbladder (3), Small Intestine (4)	Acute inflammatory response (1) Steroid metabolic process (4), Cellular response to xenobiotic stimulus (6), Drug catabolic process (18),	MBD4 (1) SUMO2 (2) SMC1A (3) NR1H4 (4)	Biological oxidations (2), Phase I - Functionalization of compounds (4), Cytochrome P450 - arranged by substrate type (5), Bile acid and bile salt metabolism (6)
DPG 3 „Favorable Upregulation"	(plot: iPSC, HLC, PHH)	Liver (1), Duodenum (2), Small Intestine (3)	Lipid transport (1), Steroid metabolic process (5), Organophosphate ester transport (17), hepaticobiliary system development (25)	MBD4 (1) SMC1A (2) SUMO2 (3) TFAP2A (4)	Fatty acid metabolism (4), Metabolism of vitamins and cofactors (6), Phase II conjugation of compounds (7), Retinoid metabolism and transport (9)
DPG 4 „Excessive Upregulation"	(plot: iPSC, HLC, PHH)	Small Intestine (1), Duodenum (2), Gallbladder (3), Colon (4)	Extracellular structure organization (1), Collagen catabolic process (5) Intestinal absorption (6), Digestive system process (10)	HNF1B (1) SMC1A (2) NR1H3 (3) SUMO2 (4)	----
DPG 5 „Adverse Upregulation"	(plot: iPSC, HLC, PHH)	Small Intestine (1), Duodenum (2), Colon (3), Rectum (4)	Skin development (1), Epidermis development (2), Digestion (3), Epidermal cell differentiation (5)	SMC1A (1) SUMO2 (2) GRHL2 (3) MBD4 (4), HIF1A (8) KLF5 (13) CDX2 (21)	Cell junction organization (1), O-linked glycosylation of mucins (2), SLC-mediated transmembrane transport (5), Digestion and absorption (15)
DPG 6 „Constant High"	(plot: iPSC, HLC, PHH)	Testis (1), Fallopian Tube (2), Heart Muscle (3), Skin (4)	Sensory perception of smell (1), Neuropeptide signaling pathway (6), Neuron fate commitment (9), Epidermal cell differentiation (12)	FOXK2 (1) COPS2 (2) REST (3) ZBTB40 (4)	Olfactory Signaling Pathway (1), Neuronal System (4), Transmission across Chemical Synapses (8), Muscle contraction (12)
DPG 7 „Insufficient Downregulation"	(plot: iPSC, HLC, PHH)	Cerebral Cortex (1), Testis (2), Adrenal Gland (3), Seminal Vesicle (4)	Nuclear division (1), Modulation of chemical synaptic transmission (3), Meiotic cell cycle (6), Mitotic nuclear division (10)	JARID2 (1) CNOT3 (2) OTX2 (3) CTBP2 (4)	Cell Cycle Checkpoints (1), Neuronal System (2), Nucleosome assembly (13), DNA methylation (17)
DPG 8 „Favorable Downregulation"	(plot: iPSC, HLC, PHH)	Spleen (1), Appendix (2)	Nuclear DNA replication (1), Cell cycle DNA replication (2), DNA replication (4) DNA conformation change (5)	CNOT (1) KDM2B (2) KMT2A (3) KDM4A (4)	DNA strand elongation (1), Chromosome maintenance (2), Telomere maintenance (3), Extension of telomeres (4)
DPG 9 „Excessive Downregulation"	(plot: iPSC, HLC, PHH)	----	----	KDM2B (1) PMEPA1 (2) CNOT3 (3) TFAP2A (4)	Integrin surface interactions (1)
DPG 10 „Adverse Downregulation"	(plot: iPSC, HLC, PHH)	----	myeloid leukocyte migration (1), Cell chemotaxis (2), Leukocyte chemotaxis (3),	KMT2A (1) CNOT3 (2) SUMO2 (3) KDM2B (4)	G-protein beta:gamma signaling (1), G beta:gamma signalling through PI3Kgamma (3), G beta:gamma signalling through PLC beta (4), Activation of kainate receptors upon glutamate binding (7)

Among the genes with constant high expression from iPSC to HLC differentiation compared to PHH in DPG6 (14,430 genes), enriched genes represented several tissue identities, including reproductive tissues, muscle and skin. The distribution of genes projected onto the DiPa plot further illustrated that DPG6 contained the highest density of genes, apart from DPG0 (**Figure 18 A**), indicating that the protocol does not succeed in downregulating substantial amounts of genes that are expressed in iPSC and unrelated to hepatocytes, possibly reflecting differentiation potential into other tissues. Genes from DPG7 (1,960 genes) are not sufficiently downregulated as well, but their expression decreases in the course of differentiation. It contains genes representative of cell division activity as evidenced by GO and Reactome pathway enrichment, in addition to genes related to neural cell fate and processes. Interestingly, target genes of transcriptional regulators with roles in epigenetic regulation of stem cells, cell division and early development, such as JARID2 [189] and OTX2 [190], were enriched in DPG7, indicating that expression of developmental genes decreased during differentiation, but were not sufficiently downregulated in HLC. This is supported by overrepresented GO terms and Reactome pathways indicating processes related to neuronal cell fate and cell division activity. In contrast, DPG8 (1,539 genes) shows that other genes previously expressed at high levels in iPSC could be downregulated successfully, e.g. those regulated by CNOT, which has been shown to be required for maintenance of pluripotency in mouse ESC [191]. Furthermore, GO term enrichment and Reactome pathway enrichment indicated that DPG8 genes were mostly associated with cell cycle activity. While comparably few genes showed excessive downregulation in DPG9 (69 genes) which did not lead to significant results in the enrichment analysis, 563 genes showed adverse downregulation in DPG10 that led to enrichment of GO terms related to cell chemotaxis, as well as overrepresentation of G-protein signaling pathways. The complete results tables of the tissue enrichment analysis, the gene ontology enrichment analysis, the transcription factor overrepresentation analysis and the pathway enrichment analysis of each DPG can be found in **Supplementary table 3, 4, 5** and **6** of the digital appendix, respectively. Illustrations of tissue enrichment, gene ontology overrepresentation, transcriptional regulator overrepresentation and Reactome pathway enrichment can be found in **Supplementary figure 4, Supplementary file 2, Supplementary figure 5** and **Supplementary file 3**, respectively.
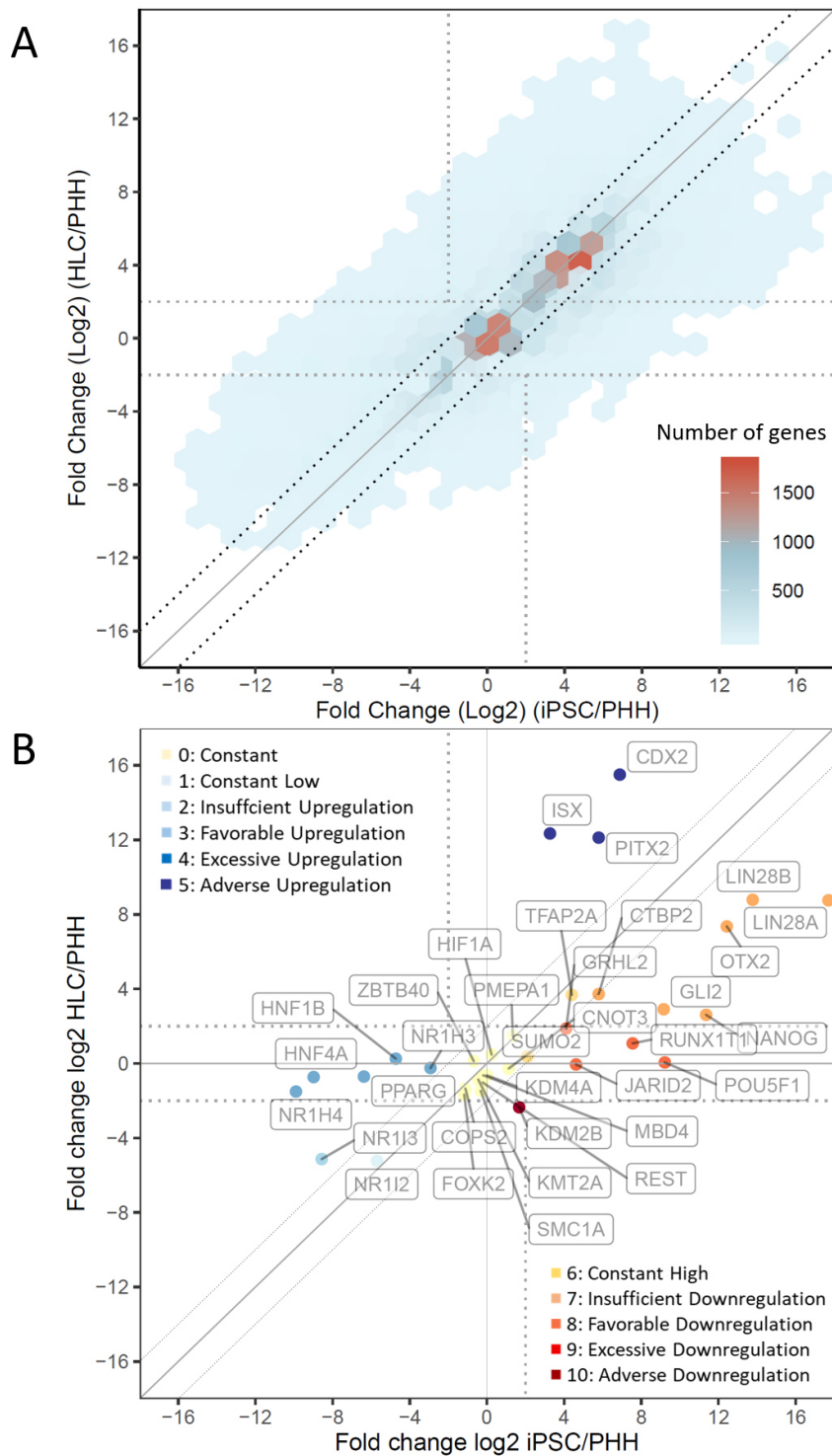
Figure 18: Gene distribution and regulator expression explored with the DiPa plot. A) Distribution of genes in the DiPa plot. The color scale represents the number of genes in the respective regions (hexagons). The x-axis represents $\log_2$ fold changes of iPSC over PHH, the y-axis indicates $\log_2$ fold changes of HLC over PHH. Dotted lines represent cutoffs of the clustering approach. **B)** Expression of selected transcriptional regulators in iPSC to HLC differentiation.

The DiPa plot was used to explore the expression of transcriptional regulators that were implicated in **Table 5** to get an overview of their expression in iPSC and HLC, compared to PHH. Additional regulators of hepatic (HNF4A, NR1I2, NR1I3), intestinal enterocytes (ISX, PITX2) and stem cell identity (LIN28, POU5F1, NANOG) were included (**Figure 18 B**). This showed that indeed intestinal regulators like CDX2, ISX and PITX2 showed high expression in DPG5, while stem cell regulators were insufficiently downregulated in DPG7 (NANOG, LIN28, OTX2) and some hepatic regulators known for their roles in xenobiotic metabolism were insufficiently upregulated (NR1I2, NR1I3). The metabolism-controlling transcription factor FXR (NR1H4) in DPG2 was particularly intriguing, since DPG2 contains genes that show upregulation in HLC compared to iPSC, but not sufficient to reach the levels observed in PHH. Furthermore, FXR is implicated in hepatic as well as intestinal gene regulation [192]. Interestingly, despite FXR being expressed (**Figure 3**), many liver-associated target genes were not upregulated in HLC (**Figure 19 A**), while intestinal genes regulated by FXR showed too high expression compared to PHH (**Figure 19 B**). One may hypothesize that a lack of FXR agonists is responsible for the low liver-specific FXR activity, representing a bottleneck to HLC maturation.
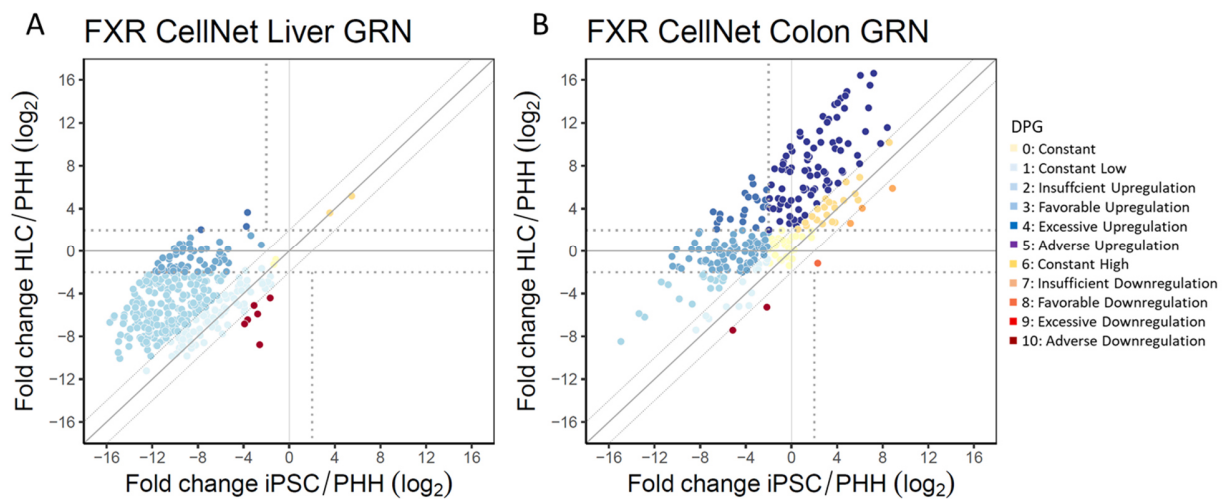
Figure 19: Liver associated genes regulated by FXR show insufficient upregulation during iPSC to HLC differentiation, while colon associated genes regulated by FXR show excessive and adverse upregulation. A) DiPa plot showing the FXR liver GRN extracted from CellNet and their association with differentiation pattern groups (DPGs). Log$_2$ fold changes of iPSC over PHH on the x-axis are compared to log$_2$ fold changes of HLC over PHH on the y-axis. **B)** DiPa plot showing the FXR colon GRN extracted from CellNet and their association with DPGs.

3.3 Integrative OMICS analysis identifies potentially FXR-responsive chromatin among insufficiently upregulated genes

The observed differences between HLC and PHH may be explained by an epigenetic landscape that does not allow expression of its dependent genes despite the presence of transcription factors such as FXR. Indeed, the enrichment of epigenetic regulators MBD4 and KMT2A may indicate altered chromatin in DPG1 and 2 (**Table 5**). Alternatively, adverse upregulation (DPG5) and insufficient downregulation (DPG7) may be a result of high chromatin accessibility.

To gain deeper insight into the interplay between gene expression and the epigenetic background in HLC, Reduced Representation Bisulfite Sequencing (RRBS) and the Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq) of iPSC, DE, HLC and PHH were performed. Promoter-focused analysis was provided by Kathrin Kattler from the group of Prof. Jörn Walter and visualized in a heatmap that shows key features of differential gene expression, promoter methylation, and chromatin accessibility in relation to the DPGs (**Figure 20)**. A large set of genes in iPSC, DE and HLC showed hypermethylation compared to PHH (**Figure 20**, mid panel). However, promoter hypermethylation did not necessarily correspond to decreased RNA expression, as observed in the region with open chromatin indicated as 'c1', including a cluster of adversely upregulated genes belonging to DPG5. Thus, despite promoter hypermethylation, upregulation of DPG5 genes, including intestine-associated genes, such as *CDX2*, *ISX*, and *GATA6*, occurred during iPSC differentiation to PHH.

Like 'c1', region 'c2' was also characterized by promoter hypermethylation, but had less open chromatin and was consistently associated with too low RNA expression in HLC. This region mostly contained genes from DPG1 and 2, of which those, especially in DPG2 were found enriched in FXR-binding motifs (**Table 5**). The complete list of z-scores for the promoter-focused analysis of RNA-seq, RRBS and ATAC-seq data for iPSC, DE, HLC and PHH can be found in **Supplementary table 7** of the digital appendix.
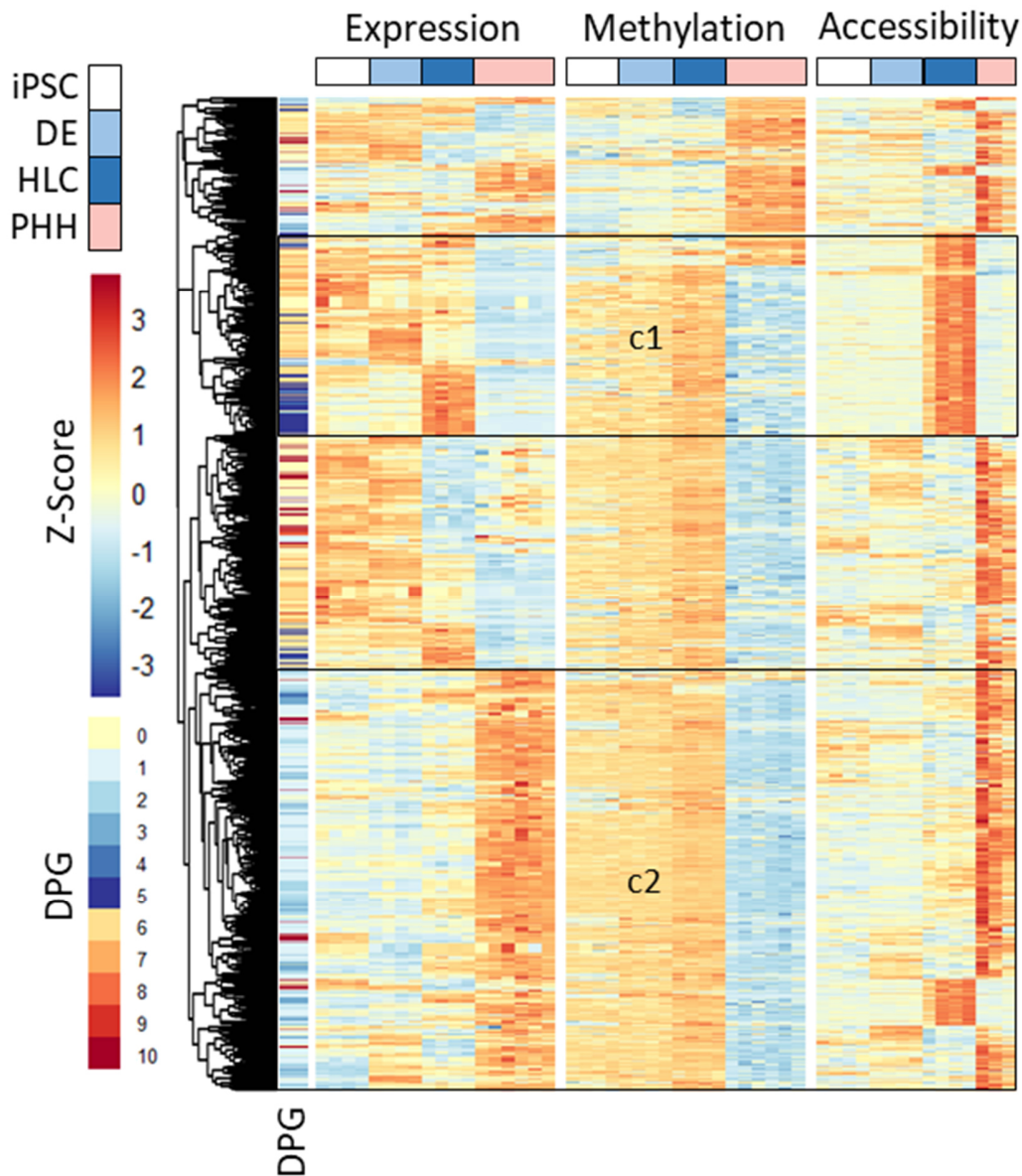
Figure 20: Integrative analysis of bulk RNA-seq, RRBS and chromatin accessibility data. Integrated heatmap of transcriptome, methylation and chromatin accessibility data obtained by RNA-sequencing, RRBS and ATAC-sequencing for iPSC, DE, HLC and PHH (columns). Data is represented as z-scores of expression values (log CPM+1) calculated for each assay. The identity of each gene (rows) was mapped to DPGs as obtained from supervised clustering. c1: gene cluster with increased chromatin accessibility and gene expression in HLC compared to PHH; c2: decreased chromatin accessibility and gene expression in HLC compared to PHH. The heatmap was kindly provided by Kathrin Kattler and was edited for illustration of results.

FXR-regulated genes with closed chromatin sites were most notably enriched with high significance in DPG2 compared to open chromatin sites (**Figure 21 A, B**). Also, DNA methylation status at promoters showed significant associations with FXR-dependent genes, particularly in DPG2 (**Figure 21 C, D**).
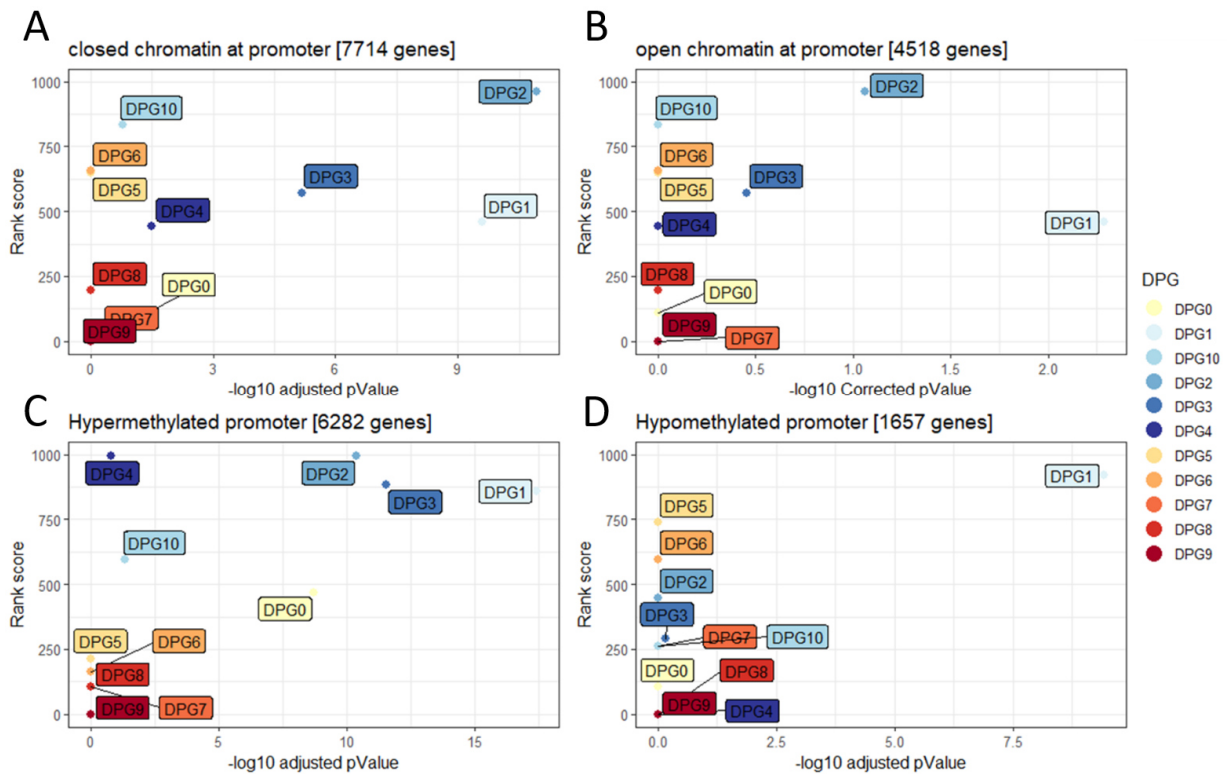
Figure 21: DPGs associated with low expression of hepatic genes show enrichment for closed chromatin and DNA methylation. A) FXR DNA target gene overrepresentation in closed chromatin promoter regions. The rank score was calculated as 1000-Ranking of FXR in the Regulatortrail overrepresentation analysis (y-axis) and plotted against the corrected p-value (-log$_{10}$, x-axis). **B)** FXR target gene overrepresentation in open chromatin promoter regions. **C)** FXR target gene overrepresentation in hypermethylated promoter regions. **D)** FXR DNA binding site overrepresentation in hypomethylated promoter. The figure was kindly provided by Kathrin Kattler.

The wide-spread promoter hypermethylation of HLC, similar to that of iPSC and DE, distinguishes these cells from PHH. However, it did not block gene expression in general and therefore does not preclude further interventions with transcription factor activation. Moreover, FXR-dependent genes were enriched in closed chromatin, which suggests that FXR activation is a promising intervention, in line with the conclusion of the gene expression analysis.

3.4 Single cell sequencing reveals a liver-intestine hybrid state in in-vitro-derived HLC

In the previous paragraphs, HLC were characterized and FXR was identified as a possible key factor required for HLC maturation. However, manipulation of FXR signaling would not be a promising strategy to improve differentiation if HLC consisted of distinct cell populations, e.g. with hepatocyte and intestinal phenotypes. In this case, cell sorting or other selection strategies would instead be adequate. Activation of FXR signaling could in this case favor maturation of the hepatic subpopulation, but simultaneously also of the intestinal cell population. Alternatively, the expression of intestinal genes may be a result of incomplete cell fate decision-making and reshaping of the epigenetic landscape, leaving HLC in a hybrid state. In such a scenario, manipulation of FXR signaling, which is involved in hepatic as well as intestinal gene regulation [192], may contribute to an important shift in the direction of either the hepatic or intestinal cell state. To address this, sc-RNA-seq was used to first study if HLC consisted of distinct subpopulations (data is available from the EGA repository under accession EGAS00001004201). Analysis of two biological HLC replicates (HLC1, n=83 cells; HLC2, n=65 cells) and PHH from three donors (PHH1, n=90 cells; PHH2 n=74 cells; PHH3, n=81 cells) showed that HLC and PHH populations clearly separated in the principal component analysis (PCA) (**Figure 22**). All three PHH donor populations clustered together, as did the biological replicates of HLC, although showed a wider distribution along PC1. No distinct HLC subpopulations were apparent in the PCA plot. Expression data (log$_2$ CPM+1) can be found in **Supplementary table 8** of the digital appendix.
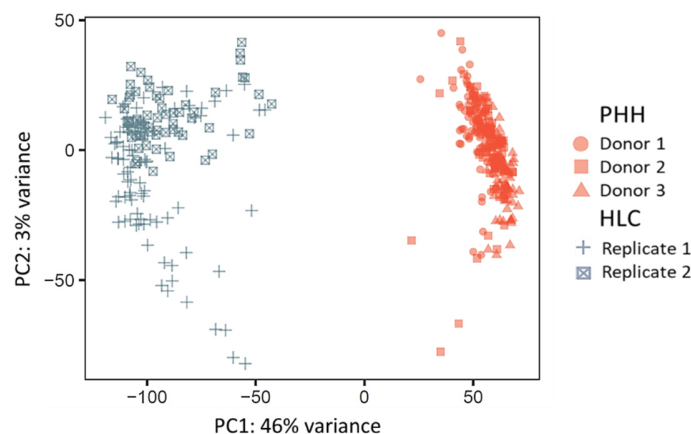
Figure 22: PCA does not show subpopulations among HLC. PCA of the top 1000 most variable genes obtained in single cell sequencing of two replicates of day 25 HLC and cryopreserved PHH of three different donors.

Addressing the hypothesis that hepatocyte-associated and non-hepatocyte genes are expressed in the same HLC, expression of hepatocyte- and intestine-associated genes from CellNet were plotted pairwise in single HLCs (**Figure 23**). As expected, PHH expressed high levels of adult hepatocyte-associated genes (y-axis), while intestine-associated genes were below the level of detection. In contrast, HLC not only expressed hepatocyte-associated genes, but variable levels of intestinal genes (x-axis) were also detectable in the same individual cells.
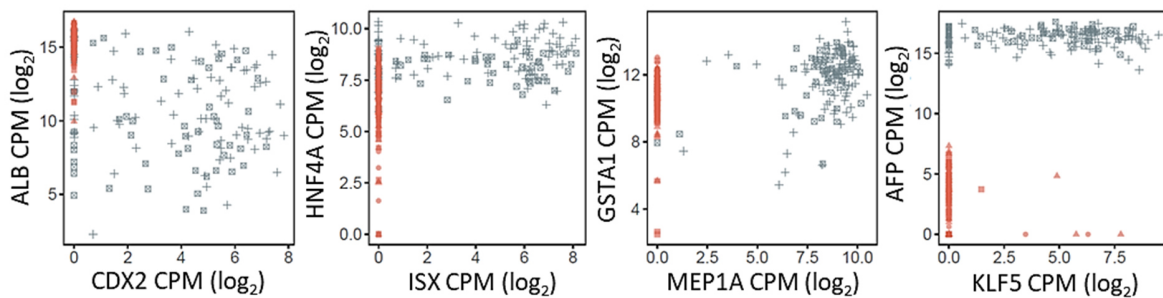
Figure 23: Co-expression of selected gene-pairs representative of liver and intestine phenotypes. Presented as normalized $\log_2$ counts per million (CPM) for each gene HLC (blue) and PHH (red).

On the protein level, this was confirmed by immunofluorescence co-staining of *AFP* and *CDX2*, as well as *AGR2* and *HNF4α* in HLC (**Figure 24**). The cytoplasmic signals of AFP and AGR2 clearly coincided with the nuclear signals of CDX2 and HNF4α in the same cells. Moreover, heatmaps of liver and intestine-associated genes from CellNet liver- and intestine-gene sets demonstrated the mixed identity of HLCs (**Figure 25**).
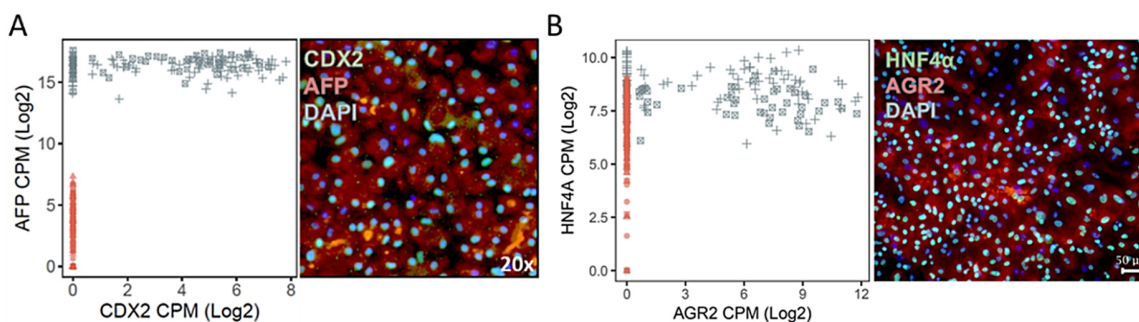
Figure 24: Selected hepatic and intestinal markers are expressed in the same cells on RNA and protein level. Co-expression of **A)** AFP and CDX2 and **B)** HNF4α and AGR2 in single HLC (blue) and PHH (red) and co-immunostaining; AFP/AGR2: red, cytoplasmic; CDX2/HNF4α: green, DAPI: blue.
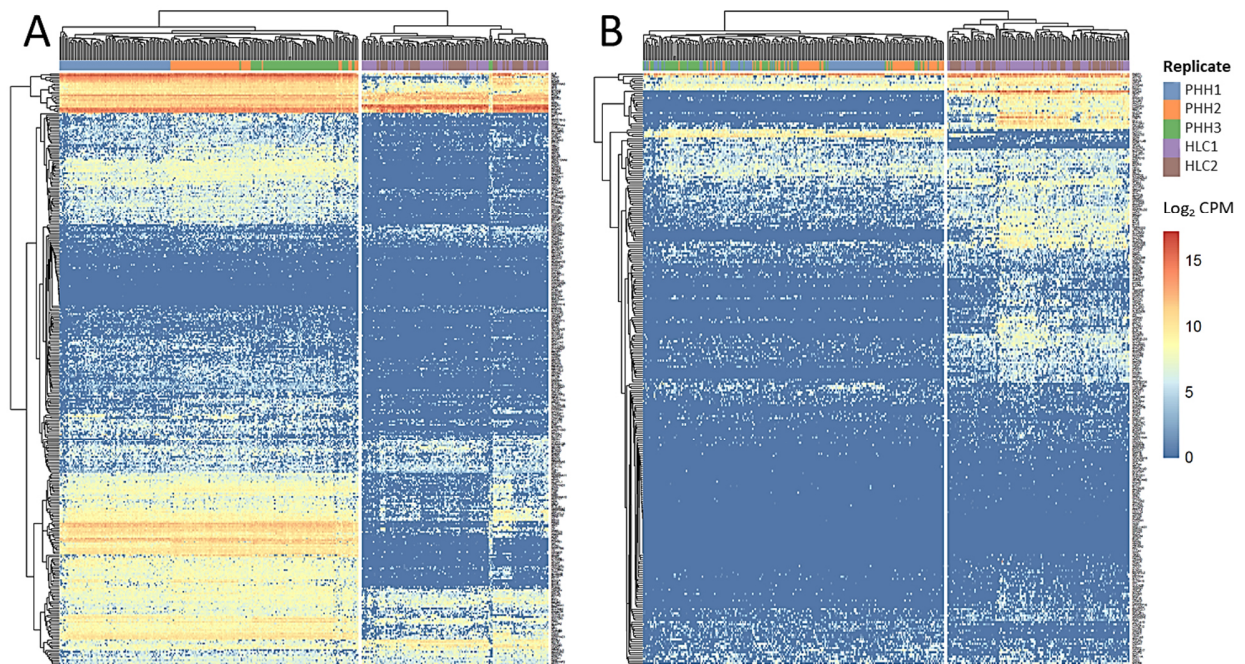
Figure 25: Single cell expression of the CellNet FXR liver and colon GRNs demonstrates hybrid cell state in HLC, compared to PHH. A) Clustered heatmap representation of $\log_2$ CPM expression of genes from the CellNet liver GRN in HLC replicate 1 and 2 and PHH from three donors, indicated by color annotation of columns. **B)** Clustered heatmap representation of $\log_2$ CPM expression of genes from the CellNet colon GRN in HLC replicate 1 and 2 and PHH from three donors, indicated by color annotation of columns.

Having shown co-expression of hepatocyte- and non-hepatocyte-related genes in HLC, the goal was to describe this protocol-induced 'hybrid state' on a genome-wide scale. Therefore, a publicly available sc-RNA-seq dataset generated by Camp et al. (2017) using a similar sequencing approach was included and allowed for definition of 'adult liver genes' as those showing at least a $\log_2$ fold change of five comparing iPSC to PHH based. Non-hepatocyte-associated genes, further referred to as 'hybrid state genes', were derived from the present sc-RNA-seq data using a modified definition of DPG4 and 5, which, according to the DiPa procedure, isolates genes that are excessively or adversely upregulated after differentiation: $\log_2$ FC[HLC/iPSC] > 1 and $\log_2$ FC[HLC/PHH] > 1. The mean scaled expression (z-scores) of all hybrid state genes was then analyzed in relation to the mean of the adult liver genes in individual HLCs (**Figure 25**). The complete list of hybrid genes and associated z-scores is available as **Supplementary table 9** of the digital appendix.
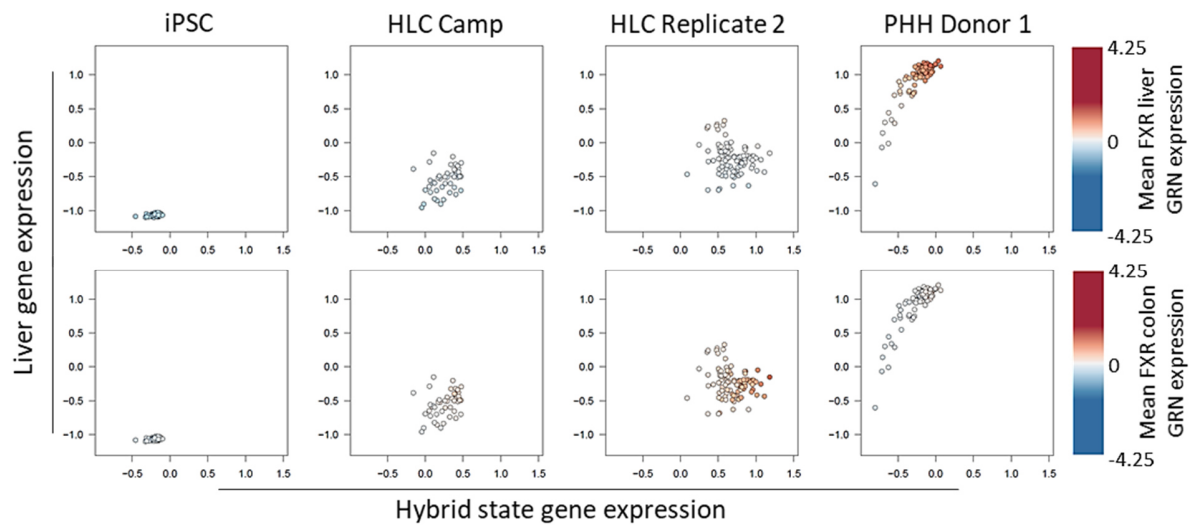
Figure 26: Liver- and hybrid state gene expression in representative replicates of iPSC, HLC and PHH. The color scale indicates expression of FXR-dependent genes of the CellNet liver and colon GRNs. The figure was kindly provided by Birte Hellwig and edited for illustration of results.

As per definition, adult liver genes displayed a strong increase in expression from iPSC to adult PHH, while this increase was less pronounced in HLC. In contrast to PHH, HLC showed an increase in hybrid state gene expression that coincided with the upregulation of adult liver genes in the same cells, which provides support for a wide-spread co-expression of adult liver and hybrid state genes. Interestingly, the more pronounced upregulation of liver genes in HLC replicate 2 also coincides with upregulation of hybrid genes, compared with Camp HLC. Highlighting the single cells according to the mean FXR target gene expression for the CellNet FXR liver and colon GRN showed that PHH activate the liver-associated, and HLC mostly the colon-associated FXR GRN (**Figure 26**), indicating that activation of FXR signaling may improve the liver character of HLC differentiation. The HLC hybrid state genes identified in the present study and by Camp et al (2017) overlapped 7.12 times more than randomly expected and showed enrichment of intestinal tissue identity in the overlap (**Figure 27**), which is remarkable considering that these cells were generated by different laboratories using distinct iPSC lines.
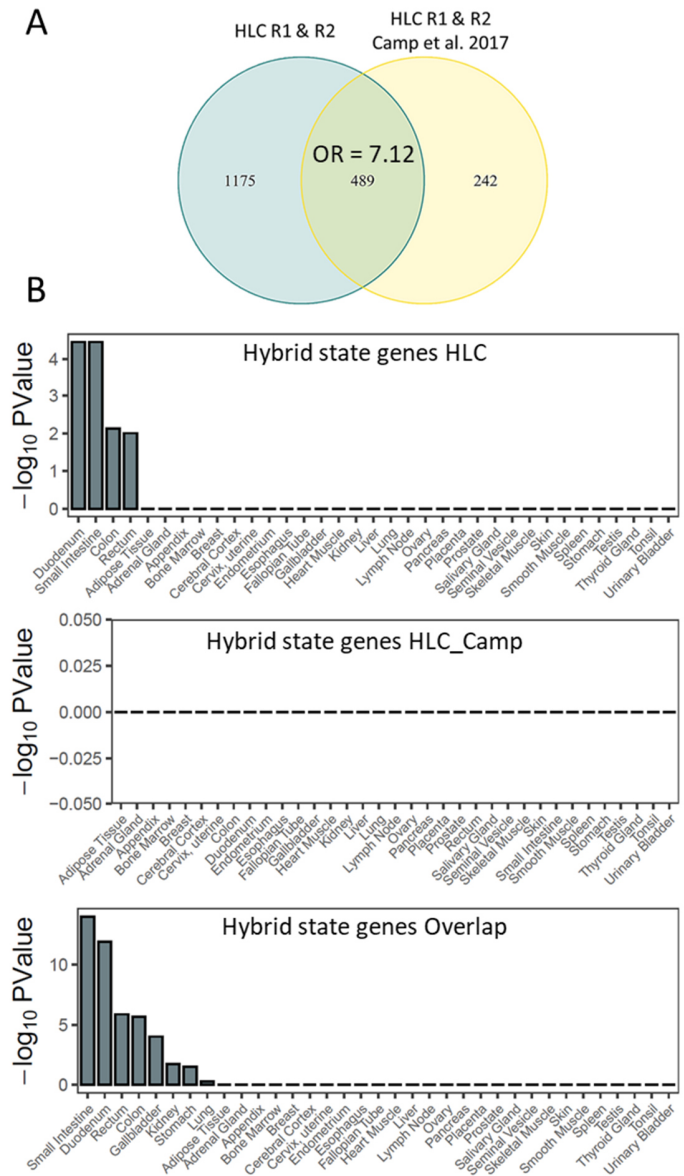
Figure 27: Hybrid state genes determined for HLC from the present study and Camp et al. 2017 show an overlap 7.12 times larger than expected by chance and enrichment for intestinal tissue identities. A) Venn diagram of HLC determined separately for HLC and HLC from Camp et al. 2017. The overlap ratio (OR) was calculated for these gene sets and displayed in the overlap region. **B)** TissueEnrich group-enrichment analysis of hybrid state genes in the unique fraction of HLC, HLC from Camp et al.2017 and the overlapping fraction (top to bottom).

Finally, further support of the HLC hybrid state concept was provided by a high throughput screen of immunofluorescence stainings against albumin (ALB) and two intestinal markers (CDX2 and AGR2) (**Figure 28**) during a wide range of time-points during the differentiation. ALB was detected

69

in 87% of HLC on day 25, but the protein abundance of the intestinal genes CDX2 and AGR2 also increased during HLC differentiation up to day 25, leading to positive cell fractions of 94% and 78%, respectively and thus demonstrating that liver and intestinal protein expression occur in the same cells. Furthermore, the time-course data indicates that intestinal hybrid genes become upregulated after DE differentiation from iPSC, which together with the qPCR data suggests that interventions could be applied in the early steps of HLC differentiation.
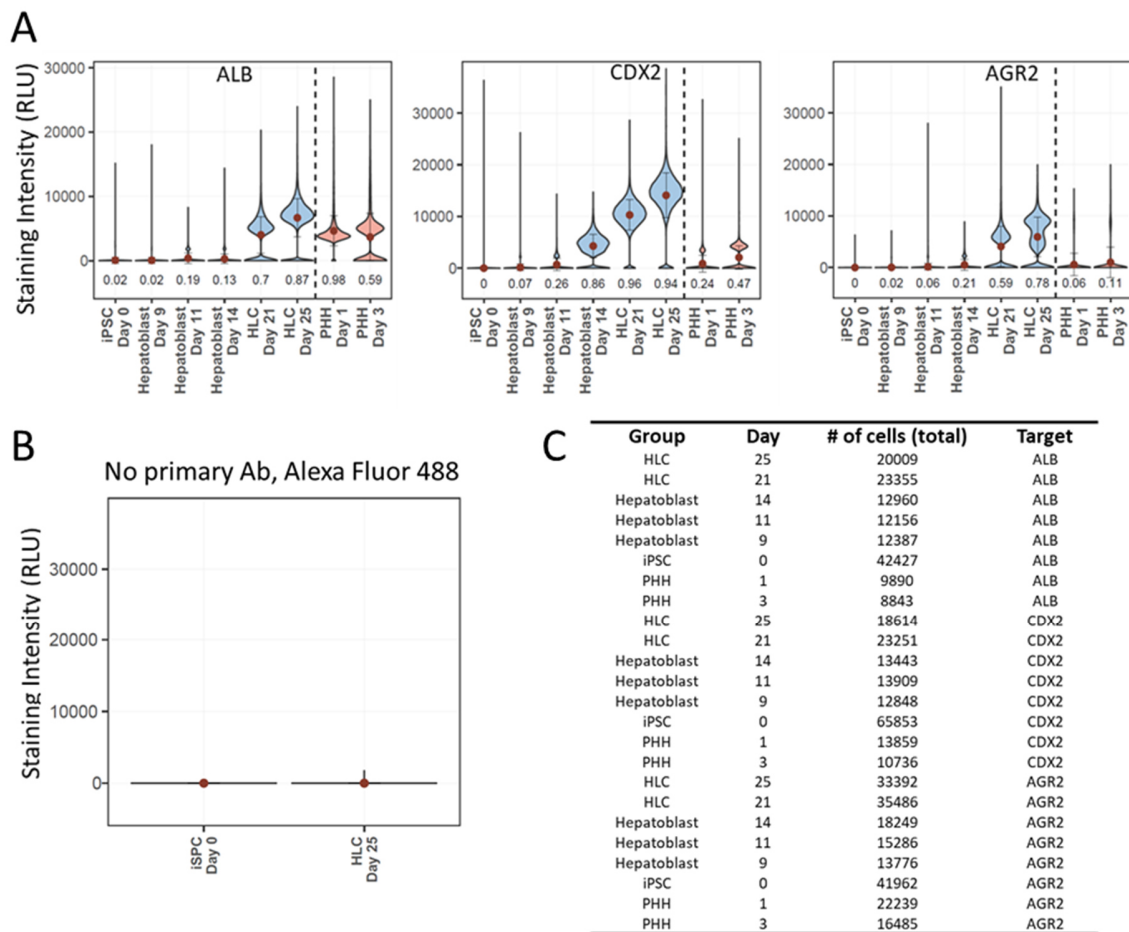
Figure 28: High-throughput immunofluorescence screening and quantification of HLC and PHH hepatic and intestinal protein expression in single cells. A) Single cell protein levels of ALB, CDX2 and AGR2 determined by high throughput screening of immunostainings of at least 10,000 cells per time point. Numbers shown below the violins indicate the fraction of positive cells. **B)** Immunocytochemistry stainings without primary antibody against ALB or CDX2, but with secondary antibody (Alexa Fluor 488 anti-rabbit). **C)** Cell counts determined by the ImageExpress Micro XLS system for the listed cell populations, time points and targets.

3.5 The liver-intestine hybrid state is not a feature of fetal hepatocytes ex vivo

In order to clarify whether the hybrid state of HLC observed in vitro also occurs in vivo during fetal gestation, fetal hepatocytes (FH) isolated at weeks 10 and 17 of gestation from a publicly available data set [82] were included into the analysis. A t-SNE representation generated from the top 1000 most variable genes in the integrated dataset used in the present study showed iPSC, HLC and adult PHH in separate clusters (**Figure 29**). The t-SNE analysis was performed using perplexity values of 1-100 to ensure representative results (available as animated **Supplementary file 4** in the digital appendix). In line with the PCA analysis (**Figure 22**), the t-SNE plot did not indicate the presence of subpopulations of HLC. In contrast, FH consisted of two subpopulations at week 10, and of several at week 17.
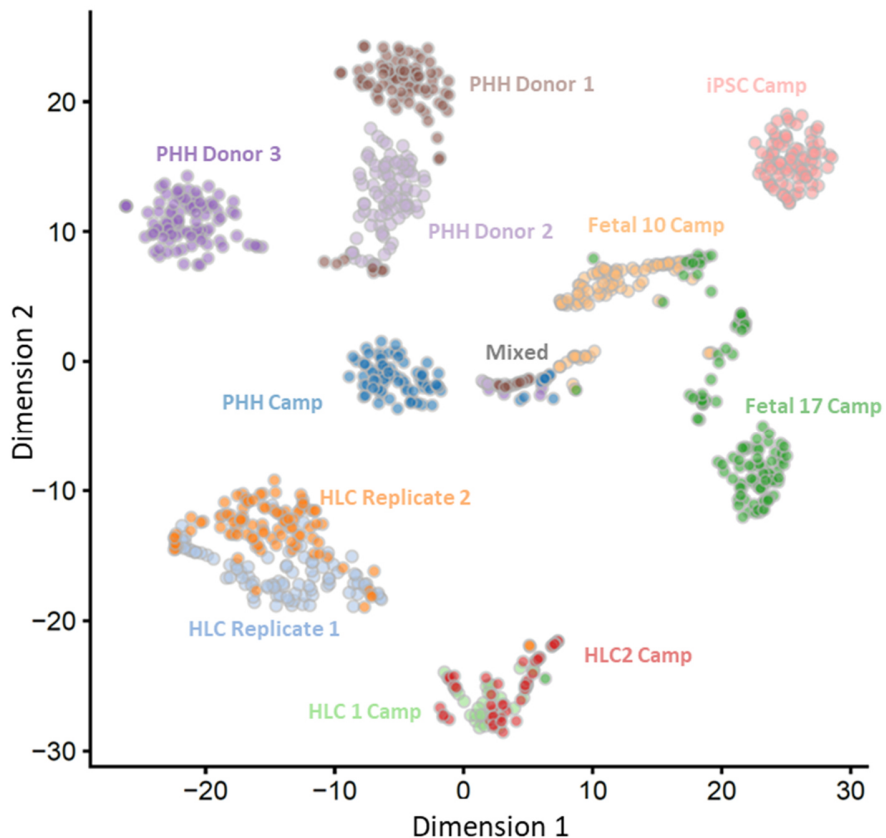
Figure 29: T-SNE representation of the integrated single cell dataset does not reveal subpopulations of HLC. HLC Replicates 1 and 2 are shown together with PHH from donors 1, 2 and 3 together with single cell sequencing data of fetal hepatocytes week 10 and 17 from Camp et al. (2017). A perplexity value of 28 was used and the top 1000 most variable genes were analyzed.

Visualization of the expression of selected hepatocyte-, and intestine-associated genes in the same t-SNE plot (**Figure 30**) provided further support that HLC do not consist of notable subpopulations. Since liver isolates from fetal stages of hepatogenesis commonly include other cell types than hepatocytes or hepatic progenitors, such as blood forming cells, potentially explaining the observed subpopulations, hematopoiesis-associated genes were included. FH consisted of two types of cells: (i) a 'hematopoietic cluster' that contained FH from week 10 and to a minor degree from week 17 with high expression of hematopoietic genes (*KLF1*, *TAL1*, and *GATA1*) but low expression of liver and intestinal genes, and (ii) 'hepatocyte precursor clusters' that expressed hepatocyte-associated genes (*ALB*, *HNF4Aα*, and *AFP*) and no hematopoietic genes. Importantly, no expression of the intestine-associated genes was observed in the hepatocyte precursor clusters of FH.
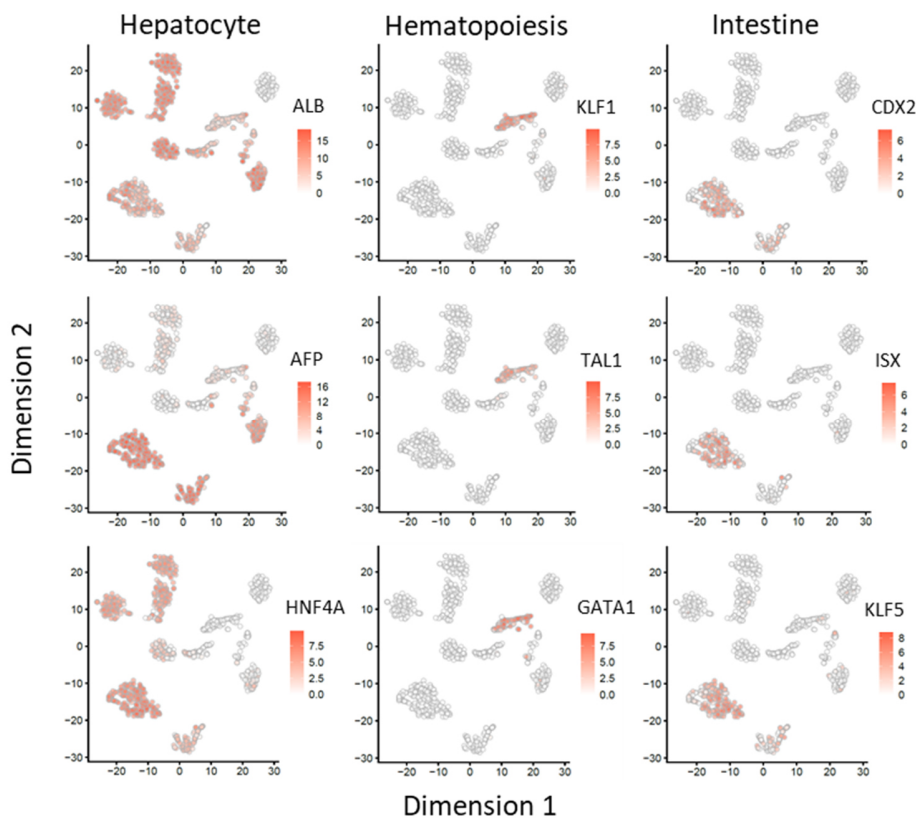
Figure 30: Expression (log$_2$ CPM) of marker genes representative of hepatocytes, hematopoietic and intestinal cells projected onto the t-SNE plot from Fig. 4 A).

Further profiling of FH on a genome-wide scale (**Figure 31 A**) showed an increase in adult liver genes in the hepatocyte precursor cell population, while the increase in hybrid state genes was lower compared to HLC shown in **Figure 26**. The slight increase in hybrid state gene expression in week 17 FH should be interpreted with caution, since these cells were transiently taken into culture for purification (Camp et al., 2017). Therefore, it cannot be excluded that the slight increase in hybrid gene expression was induced by in vitro subculture. The color scale in **Figure 31 A** illustrates the distinct hepatocyte precursor and hematopoietic cell clusters of FH week 17, expressing either *ALB* or *KLF1*, respectively. Compared to PHH, FH from week 10 or 17 only showed weak FXR liver GRN expression; colon FXR GRN expression was low in PHH and FH.
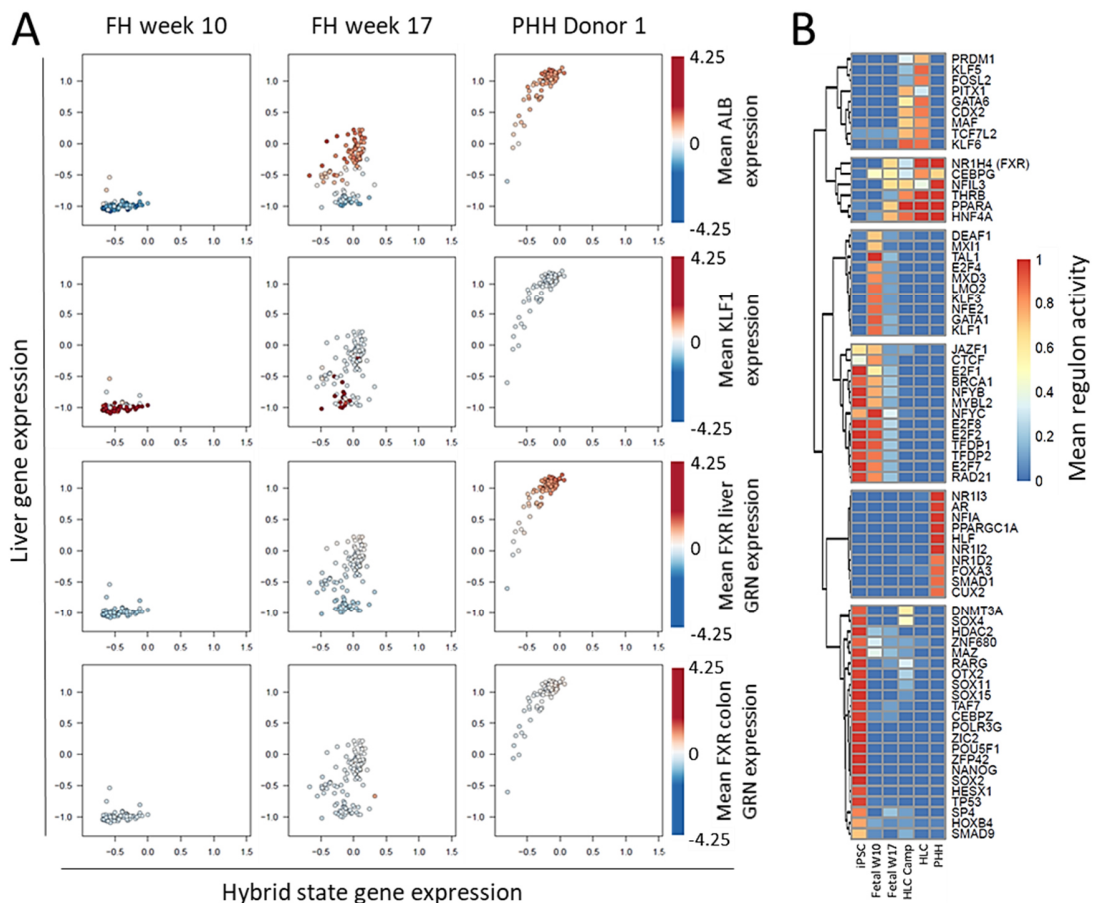
Figure 31: Expression of intestinal genes is exclusive to HLC and discriminates them from FH and PHH. A) Liver- and hybrid state gene expression in FH week 10, FH week 17 and PHH. The color scale indicates expression of the hepatocyte marker ALB, the erythropoiesis marker KLF1 and FXR-dependent genes of the CellNet liver and colon GRNs. The figure was kindly provided by Birte Hellwig and edited for illustration

of results. **B)** Heatmap of regulon activity discriminating iPSC, day 25 HLC, PHH and fetal hepatocyte populations.

The specific features of the individual cell types were further characterized by regulon activity analysis using SCENIC, a tool specifically designed to analyze single cell expression data by network inference and provide insight into the drivers of cell heterogeneity (**Figure 31 B**) [185]. Thereby, factors were identified that are involved in driving the cell states of iPSC, HLC, PHH and fetal hepatocytes. Strikingly, HLC clearly exhibited a liver-intestine hybrid state also on the level of transcriptional regulators, characterized by the simultaneous activity of liver-associated (FXR, CEBPG, NFIL3, etc.) and intestine-associated genes and regulators (PRDM1, KLF5, FOSL2, CDX2, etc.) in the same cell. In contrast, these intestine-associated genes and regulators were not observed in FH ex vivo, neither at week 10 nor 17, indicating that ex vivo FH are not in a hybrid state, which holds important implications for the apparently impaired lineage decision making of HLC, as it shows that activation of the described intestine-related GRNs is not occurring after DE patterning in vivo and thus the result of misguided differentiation that may be manipulated by intervention.

Furthermore, while some factors were exclusively observed in PHH, such as the metabolism controlling nuclear receptors NR1I2 (PXR) and NR1I3 (CAR), other factors, including NR1H4 (FXR) and PPARA were already contributing to the cellular state in HLC as well as PHH, but much less in FH. Taken together with the previous observation that HLC express some FXR, but seem to activate only the colon-associated part of the FXR GRN, this indicates that HLC may be already able to respond to activation of FXR signaling. This observation prompted the design of an intervention strategy for FXR activation, to elucidate if HLC respond to FXR stimulation by activating liver-associated FXR target genes, potentially representing a tunable switch to improve HLC maturation. Other interesting possibilities were that the response would be exclusively intestine-like, or whether it is a combination of both, functionally reflecting the hybrid state of HLC which would be in support of the hybrid state theory in that it would demonstrate the activation of functional features of intestine in liver in HLC. The complete results table for the regulon analysis are available as **Supplementary table 10** of the digital appendix.

3.6 Manipulation of FXR expression and activity enhances HLC maturation

Since FXR was identified as a critical gene regulatory node for HLC maturation, it represents a candidate for targeted intervention designed to improve the differentiation of HLC. Gene regulation through FXR seems to depend not only on its expression, because mRNA levels of FXR (*NR1H4*) in HLC are relatively similar to those of PHH (**Figure 15**). Nevertheless, many liver-associated FXR target genes remained at too low expression levels in HLC (**Figure 19**). Based on these observations it was an interesting question if, besides exogenous FXR expression, the additional stimulation by agonists may cause an increase in liver-associated gene expression, while simultaneously decreasing intestinal FXR target gene expression. To address this, a strategy combining lentiviral expression and exposure to FXR agonists, chenodeoxycholic acid (CDCA) and/or GW4064 (**Figure 32)**, was employed.
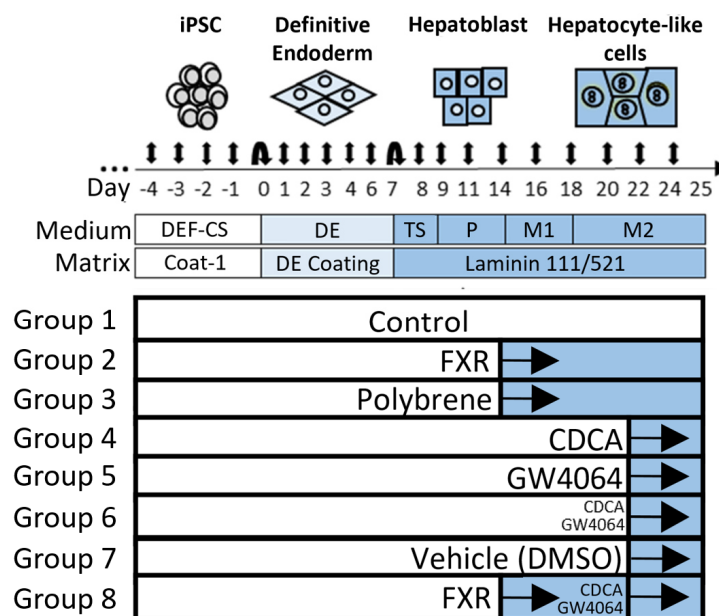
Figure 32: Intervention schedule for FXR transduction and agonist treatment (1.5 µM GW4064; 100 µM CDCA). At day 14, HLC were treated with polybrene and transduced with FXR expressing lentivirus at an MOI of 10 or treated with polybrene alone (Group 2 and 3). At day 22, treatment with CDCA and GW4064 or both was initiated, as well as a vehicle control (Groups 4-7). For Group 8, treatments were combined.
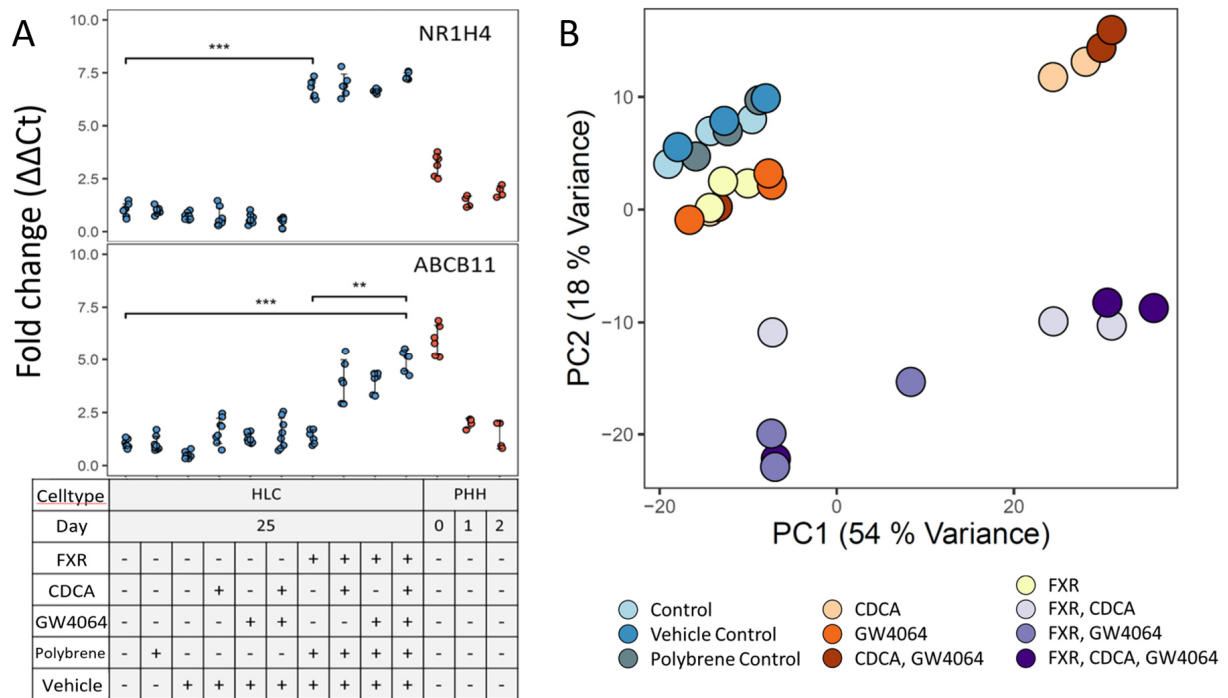
Figure 33: Combined intervention with FXR and agonists elicits a stronger response than individual interventions. A) Influence of the interventions on FXR (NR1H4) and BSEP (ABCB11) expression analyzed by qRT-PCR. Treatments and transduction (polybrene) control and vehicle control (DMSO) are indicated for HLC day 25 and PHH on day 0 (fresh PHH), 1 and 2 of culture. Error bars indicate standard deviation. Stars above comparisons indicate statistical significance of t-test results (* = 0-0.1, ** = 0-0-01, *** = 0-0-001). **B)** Principal component analysis of the top 500 variable genes in HLC at day 25 and HLC at day 25 that were treated either with CDCA, GW4064, an FXR expressing lentivirus or a combination thereof.

Lentiviral expression of FXR caused the expected increase in FXR mRNA levels, but did not induce expression of the FXR-dependent gene *ABCB11* (BSEP) (**Figure 33 A**). FXR expression was not significantly influenced by addition of the FXR agonists CDCA and GW4064. Interestingly, when FXR was overexpressed, addition of the FXR agonists caused an induction of BSEP to levels higher than observed in cultured PHH on day 1 and 2, but still lower than in freshly isolated and cryopreserved PHH. The addition of agonists alone did not induce BSEP expression, indicated that a combined intervention was needed to induce this liver-characteristic marker. Dimensionality reduction using the PCA method revealed that indeed combined intervention with FXR overexpression and agonist treatment elicited the strongest response among the top 500 variable genes (**Figure 33 B**). On the first principal component, accounting for 54% of the observed variance, the effect of CDCA treatment alone and the combination of FXR CDCA treatments

showed a high impact. In contrast, treatment with GW4064 alone did not show a pronounced effect. Also, FXR transduction alone did not lead to a major response among HLC. Only in combination with FXR overexpression, GW4064 treated HLC exhibited a response that was most noticeable along principle component two, which accounted for 18% of the observed variance. However, the strongest response was achieved when FXR transduction was combined with CDCA and GW4064 treatment. This supports the observation from qRT-PCR data that showed an upregulation of BSEP only when FXR transduction is combined with agonist treatment (**Figure 33 A**). Replicate one out of three showed a divergent pattern in response to the agonist and lentivirus treatments but was not excluded from the analysis, since Pearson correlation of gene expression values for the replicates were high (0.96 for replicate 1 and 2; 0.97 for replicate 1 and 3 and 0.98 for replicate 2 and 3 for FXR plus CDCA and GW6046 intervention, **Supplementary figure 6**). The observed effect may reflect a replicate-specific tendency to either respond more to CDCA (replicate 2 and 3) or GW4064 (replicate 1) (illustrated in **Supplementary figure 7**). Furthermore, absolute expression of selected markers, including ABCB11 and FXR that were previously analyzed by qPCR data (**Figure 33 A**), indicated that a similar overexpression of FXR and regulation of hepatic and intestinal markers among the replicates (**Supplementary figure 8**). Despite the observed variance, which is subject to future investigation, the combined FXR plus CDCA and GW4064 intervention clearly showed the most pronounced effect on global gene expression in HLC.

Differential expression analysis likewise suggested that FXR expression plus treatment with both agonists (CDCA and GW4064) induced a higher number of differentially expressed genes compared to FXR plus each agonist alone (**Figure 34**). A relatively large overlap of differential genes among the three treatments was obtained for FXR transduction in combination with either CDCA, GW4064 or both. At a cutoff of $|\log_2$ fold change$| > 1.5$ compared to vehicle control no differential genes were detected in control HLC or polybrene treated HLC at day 25. Only lowering the cutoff to $|\log_2$ fold change $> 0.3|$ allowed for detection of a difference of about 50 genes, thus confirming a high comparability between controls (**Figure 34 A**). At the cutoff of $|\log_2$ fold change$| > 1.5$, only a small number of differentially expressed genes were detected in HLC that were treated with the FXR agonists CDCA or GW, or both (**Figure 34 B**). In contrast, transduction

of HLC with FXR and subsequent treatment with the agonists drastically increased the number of differential genes and showed a relatively large overlap between FXR transduced HLC treated with either CDCA or GW4064 separately and the combined treatment, where the combined treatment clearly induced the highest number of differential genes (**Figure 34 C**). The complete list of differentially expressed genes can be found in **Supplementary table 11**.
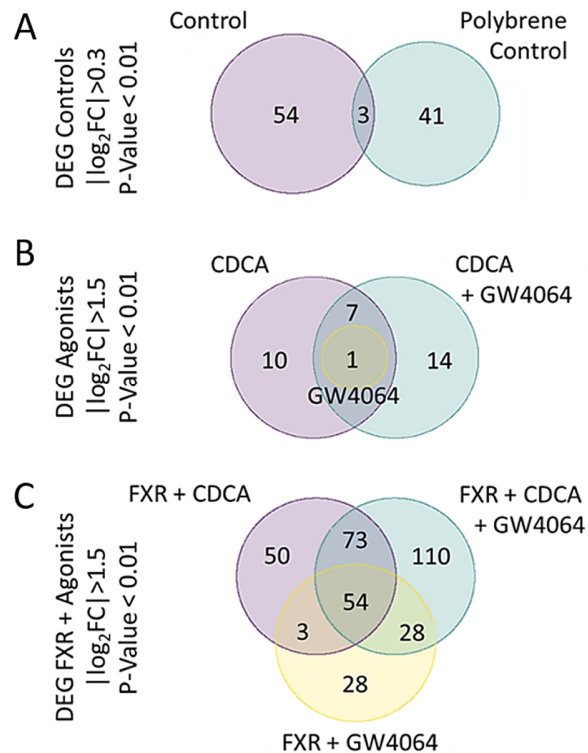
Figure 34: Combined FXR transduction and agonist stimulation elicits the strongest response in differentially expressed genes. A) Venn diagram of differentially expressed genes (DEG) with an absolute $\log_2$ fold change of at least 0.3, p-Value < 0.01 in control and polybrene treated hepatocyte-like cells at day 25 of differentiation, compared to vehicle control. At an absolute $\log_2$ fold change of at least 1.5, no DEG were found. **B)** Venn diagram of DEG with an absolute $\log_2$ fold change of at least 1.5, p-Value < 0.01 in CDCA, GW4064 or CDCA and GW4064 treated hepatocyte-like cells at day 25 of differentiation, compared to vehicle control. **C)** Venn diagram of DEG with an absolute $\log_2$ fold change of at least 1.5, p-Value < 0.01 in FXR+CDCA, FXR+GW4064 or FXR+CDCA and GW4064 treated hepatocyte-like cells at day 25 of differentiation, compared to vehicle control.

The DiPa plot was used to visualize how the interventions influenced gene expression within the different DPGs (**Figure 35 A**). Red and blue dots in the DiPa-plot indicate genes that were significantly up- and down-regulated by the combined intervention, respectively. The background

color of the DiPa plot (red to blue) represents the ratio of up- and down-regulated genes, which was 6.3, 1.3, 0.7, 0.2, and 0.4 in DPGs 1-5, respectively. DPG1 and 2 contained significantly more up-regulated genes than the average of all DPGs (**Figure 35 B**). Conversely, DPG4 and 5 included more than average down-regulated genes, while DGP3 contained both more up and down-regulated genes. Corresponding analyses in DPG 6-10 did not amount to statistical significance. Overall, the intervention enhanced the too low gene expression in DPG1 and 2, while reducing levels in DPG4 and 5, in which expression was too high in HLC compared to PHH.

DPG	P-Value upregulation	P-Value downregulation
DPG 0	1	1
DPG 1	0	0.9506
DPG 2	0	0
DPG 3	0.0022	0
DPG 4	0.5329	0.0001
DPG 5	0.0756	0
DPG 6	1	1
DPG 7	0.8263	0.9694
DPG 8	0.4378	0.9723
DPG 9	0.309	0.2139
DPG 10	0.0823	0.1337
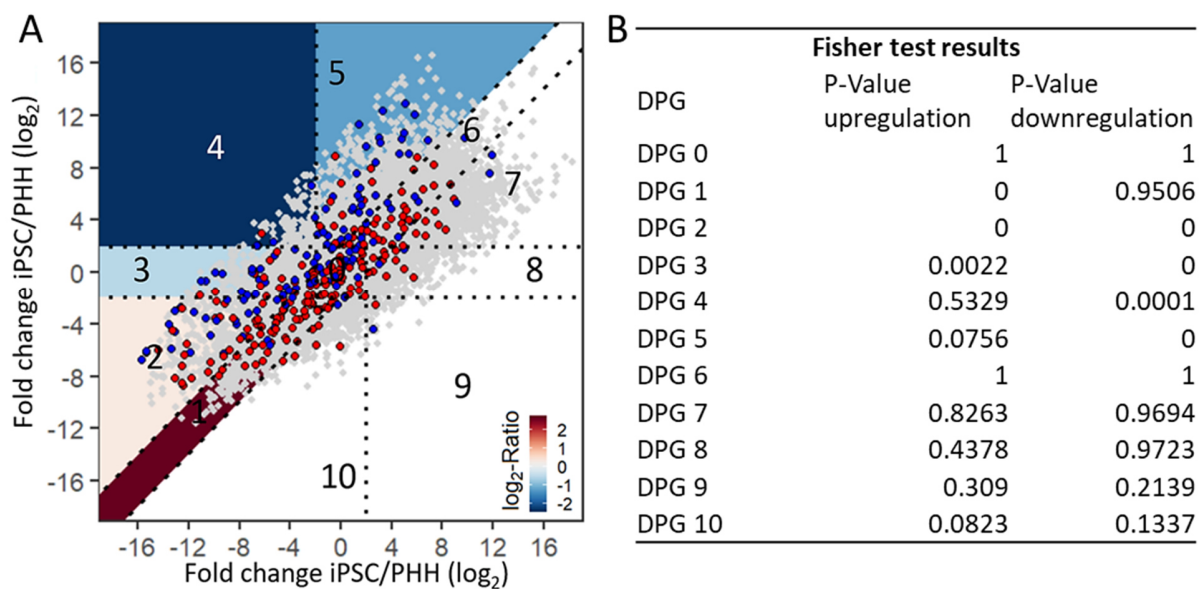
Figure 35: Influence of FXR transduction and agonist treatment visualized by the DiPa plot. A) Expression changes (given as $\log_2$ ratio) after FXR + agonist intervention projected onto the DiPa plot. B) Fisher-test results testing the null hypothesis that there are less up- or downregulated genes in the overlap than expected by chance.

Among the top 15 up- and downregulated DEGs for the combined FXR and agonist treatment intervention several genes associated to liver and intestinal tissues were identified (**Figure 36 A**). Interestingly, FGF15, which is released in the epithelial cells of the small intestine after stimulation with bile acids, showed the most pronounced upregulation (>150 fold). Nevertheless, several liver associated genes, including the liver-specific bile salt export pump *ABCB11* (BSEP), *ITIH3* and *KNG1*. Among the most downregulated genes, *FMO1*, normally expressed in fetal hepatocytes, showed the strongest downregulation. In addition, intestinal markers were among the top 15

downregulated genes, including sucrose isomaltase (*SI*) by the combined intervention. Finally, FXR target genes in the liver- (**Figure 36 B**) and intestine (**Figure 36 C**)-associated FXR networks from CellNet were extracted and their expression changes visualized for each condition of the intervention (FXR transductions plus agonists), considering fold changes of at least 1.5 over the vehicle control. Interestingly, most of the liver-associated genes were upregulated by the combined FXR transduction plus agonist treatment, while less liver-associated genes decreased (**Figure 36 B**). In contrast, most of the intestine-associated genes were downregulated, while only three increased (**Figure 36 C**). Thus, increased FXR activity predominantly enhanced expression of genes from the liver-associated network, while the FXR colon network was repressed, although a definite response with e.g. induction of all liver-associated and repression of all intestinal genes could not be achieved.
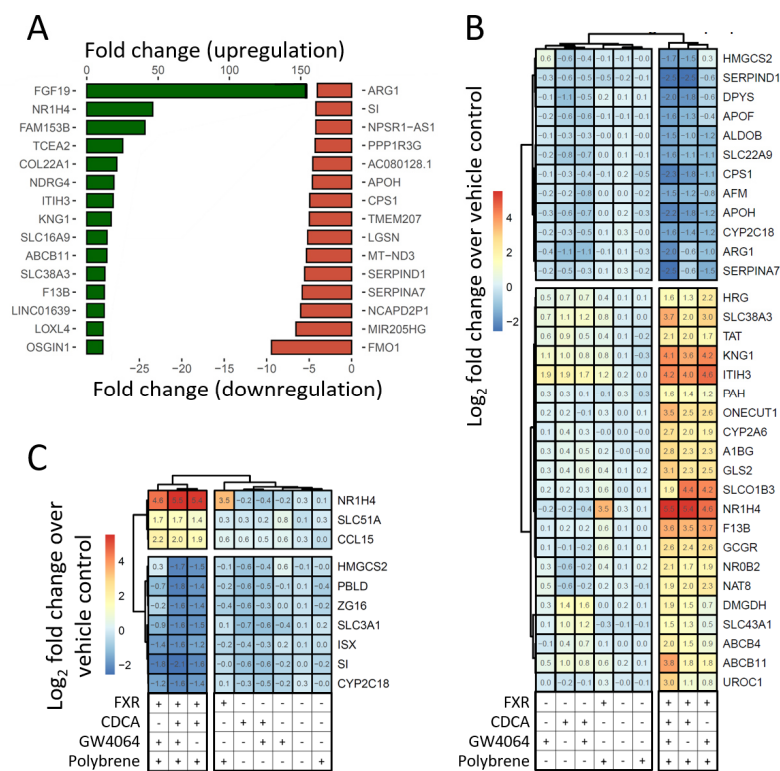
Figure 36: Intervention with FXR and agonist activation elicits the most pronounced up- and downregulation of hepatic and intestinal FXR target genes. HLC. **A)** Differential genes with the highest absolute fold changes and FDR-adjusted p-value lower than 0.001 for FXR transduced and agonist (1.5 µM GW4064; 100 µM CDCA) stimulated HLC. **B)** and **C)** Heatmaps visualizing expression changes larger than a log$_2$-fold change of |1.5| for liver- and intestine-associated genes for all interventions.

4. Discussion

Hepatocytes are an important tool in drug development [193–196] and preclinical research [197–199], yet their limited availability and high cost encourage the development of stem cell-based differentiation protocols. Despite remarkable achievements in the field, the use of stem cell-derived hepatocytes is hampered by their incomplete differentiation [78,88,200]. Genome-wide studies have demonstrated that human embryonic- and iPSC-derived hepatocytes show too low expression of a large cluster of metabolism-associated genes expressed in PHH but also too high levels of genes not expressed in PHH, such as genes normally expressed in the gastrointestinal tract [80,88,139]. These findings indicate that HLC require proper (biochemical) clues in order to complete differentiation to become mature hepatocytes that feature the expression of important metabolic genes and determination of their cellular identity. To characterize the stem cell-derived HLC and to design a strategy to improve their differentiation, RNA-seq analysis of bulk preparations and single cells, characterization of the epigenetic landscape and targeted interventions were employed to optimize the activity of gene regulatory networks involved in hepatic and intestinal signaling.

Initial analysis of HLC by immunofluorescence imaging and qRT-PCR showed that indeed HLC express markers of liver and intestine and seem to have fetal characteristics as well, seen at the example of AFP expression. Bulk sequencing of HLC followed by CellNet analysis demonstrated 82% liver and 56% intestine identity, confirming previous reports [80]. Single cell RNA-seq did not provide evidence of subpopulations that show a more mature hepatic, intestinal or other phenotype. In contrast, the results clearly demonstrate that the transcriptome of HLC represents a hybrid state, where genes representative of at least two adult cell types are expressed within the same cell. This hybrid state could be confirmed in an independent HLC dataset [82]. Demonstrating that the most dominant phenotypes - liver and intestine - occurred in single HLC instead of subpopulations with separate differentiation trajectories reveals major shortcomings of the current differentiation protocol in recapitulating developmental lineage decision making on the transcriptome and epigenome scale. While intestine and liver were the dominant signatures identified in HLC the presence of genes such as HNF1β and GRHL2 in DPG4 and 5, respectively, also indicated that HLC failed to fully commit to the hepatocyte or cholangiocyte

lineage. Moreover, DPG6, which contained genes that were expressed way above hepatocyte levels but of which the expression remains largely unaffected by the differentiation protocol, was shown to include genes representative of muscle tissue as well as neuronal and skin differentiation-associated processes and stem cell related gene expression. Taken together with the results from the integrated epigenetic analysis, where wide-spread hypermethylation in iPSC, DE and HLC in comparison to PHH and relatively high accessibility of promoter regions associated with genes from DPG5 and DPG6 were observed, the hepatocyte phenotype of HLC was shown to lack an unambiguous determination to the hepatocyte lineage.

This PhD thesis has also shown that fetal liver cell populations did not display an intestinal gene expression signature, thus not exhibiting a liver-intestine hybrid state (**Figure 38**). In contrast, FH isolated at week 10 and 17 of gestation by Camp et al. 2017 contained a subpopulation of cells with hematopoietic signature, presumably hematopoietic stem cells, which is in line with the role of the liver as a major site of hematopoiesis in fetal development. Thus, it was concluded that proper differentiation into the hepatocyte lineage requires suppression of the hybrid state and activation of hepatocyte-specific gene signatures. Profiling of HLC from additional laboratories and derived by varying protocols will provide further evidence on the question whether the hybrid state represents a universal feature of stem cell-derived HLC.

Another important question addressed in this PhD thesis was if HLC differentiation can be improved by inducing the expression of lowly expressed liver genes or suppressing hybrid state genes. To identify promising candidates and to assess the efficacy of interventions, the DiPa supervised logical clustering was established. It projects each gene onto a coordinate system in which the ratio of HLC/PHH is plotted against iPSC/PHH. Consequently, groups of genes with similar development of gene expression during the differentiation process cluster together within DPGs. The DiPa plot in combination with enrichment analysis pointed towards an influential role of the nuclear receptor FXR (*NR1H4*) among DPG1 and 2 that represented groups of genes that need to be upregulated to hepatocyte level.
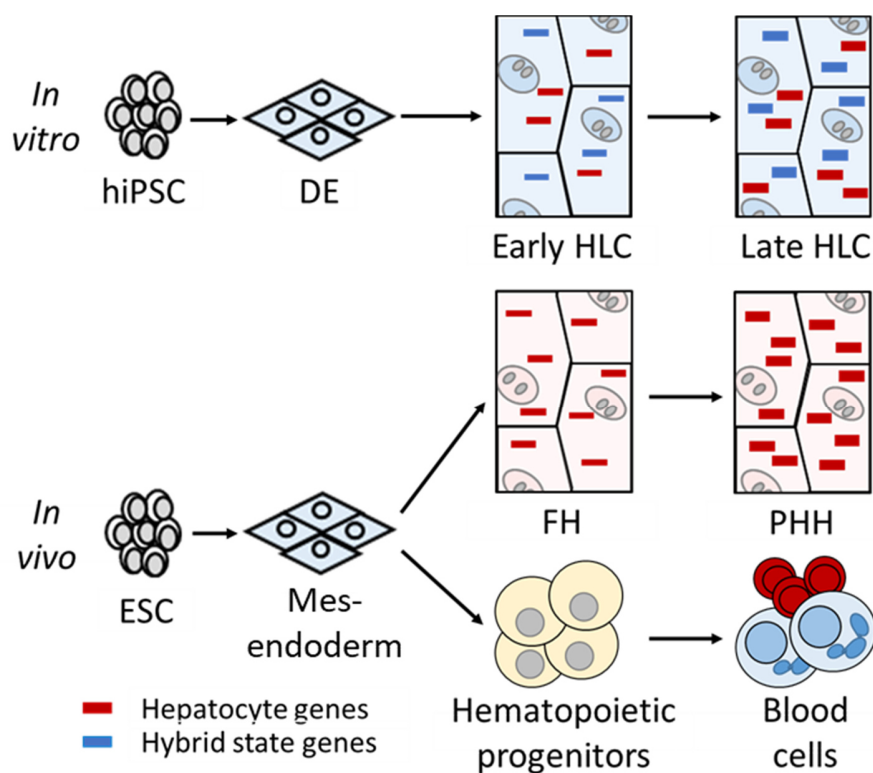
Figure 37: The hybrid state concept. Schematic of intestine-liver hybrid state gene expression as a consequence of in vitro HLC differentiation compared to in vivo differentiation.

An interesting feature of this transcription factor is that FXR is a key regulator of hepatic as well as intestinal gene expression (Lee et al., 2006; Pereira-Fantini et al., 2014; Wagner et al., 2003). This led to the question of whether experimentally increasing FXR activity in HLC would lead to an increase in DPG2 and DPG3 genes, or if genes belonging to DPG4 and DPG5 would increase, thus promoting the undesirable intestinal phenotype. A further argument speaking for an intervention to activate FXR was that regions with closed chromatin were enriched in FXR-dependent genes, since FXR was reported to remodel chromatin to an *open* configuration [204]. The fact that FXR-dependent genes showed too low expression although FXR itself was present was encouraging stimulation of cells with FXR agonists, in addition to exogenously expressing FXR. An intervention strategy with lentiviral FXR expression was applied to guarantee that its levels are not limiting, together with stimulation by the strong endogenous FXR agonist CDCA [205], the

potent, non-steroidal, synthetic FXR agonist GW4064 (Zhang et al., 2015), and a combination of both.

Indeed, the results showed that a major influence on global gene expression was only obtained by combined FXR expression and agonist exposure. FXR activity not only enhanced FXR-dependent liver genes in DPG1 and DPG2, but also suppressed intestinal genes in DPG4 and DPG5. This is conceptually important since it demonstrates that a single transcription factor alone can enhance expression of tissue specific genes as well as suppress unspecific or hybrid genes. Whether FXR enhances or suppresses gene expression depends on the presence of co-activators and interacting regulators of transcription. Numerous co-activators, e.g. PCG1, bind FXR and enhance its transactivation activity [204,207]. Moreover, FXR is known to inhibit the activity of other transcription factors, e.g. by binding to the small heterodimer partner (NR0B2) that reduces the transcriptional activity of liver receptor homologue 1 (LRH-1). The observation that expression of FXR plus agonist exposure enhanced the expression of liver- and suppressed intestine-associated genes suggests that HLC already have a hepatocyte-specific pattern of co-regulators and FXR interaction partners. Indeed, HLC strongly upregulated the transcription of transcriptional FXR the co-factor gene*NR5A2* (LRH1) compared to iPSC, while also expressing other co-regulator genes, such as *CARM1*, *GPS2* and *TRRAP*, but a functional analysis of the FXR-interactome has not yet been performed.

Since major differences between HLC and PHH remained after activation of FXR - particularly in DPG6-10 - that were not significantly influenced by the intervention, it will be important to combine the targeting of FXR with further interventions, e.g. to induce expression of other metabolism-controlling nuclear receptors such as NR1I2 and NR1I3 to find out if hepatic (xenobiotic) metabolism can be activated in HLC, increasing their potential for toxicology and pharmacology studies. The question remains if there is a tipping point beyond which the entire expression profile and chromatin landscape is brought closer to PHH. For instance, reducing the activity of the transcription factors CDX2, GATA6, and KLF5 may further decrease the expression of hybrid state genes. Such a strategy is experimentally feasible, because the application of viruses (if required due to low transcription of intervention targets) and agonists to the culture medium can be routinely established and could enhance specifically required features in certain model

system scenarios, where e.g. higher metabolic activity of hepatocyte like cells is required. However, the most promising strategy for proper directing of HLC differentiation is to thoroughly recapitulate embryonic development. To achieve this, detailed studies of highly time-resolved human embryonic endoderm patterning and liver development, the critical interaction of involved cell types and signaling mechanisms will be required. Current procedures utilize comparably short time periods and a limited selection of comparably well studied signaling pathways. In previous years, identification of master regulators for differentiation into certain cell types, such has HNF4α in case of hepatocytes or MyoD in muscle cells, allowed for impressive conversions of cells into other cell types, but in some cases the limited control over the underlying interplay of epigenome, transcriptome, proteome and metabolome interplay still represents a major challenge to cell type and tissue engineering. Thus, fine tuning of the differentiation processes may represent the only option to obtain cells that truly resemble their in vivo counterparts and that will be effective and safe in clinical scenarios. For other applications such as in vitro models of certain functions, cells may not need to fully recapitulate the target cell identity, but will require extensive validation before their routine application. In their current state, HLC may not yet be ready to substitute PHH in these scenarios, but they may already present a more suitable alternative compared to immortalized cell lines, which due to the shortage in primary cells are an often used resource for toxicological studies. This needs to be demonstrated in future studies.

In conclusion, here it was demonstrated that hiPSC derived HLC generated by a commonly-used in vitro protocol co-express both liver and undesired intestinal genes within the same cell. The nuclear factor FXR represents a critical control factor that can be targeted to shift the balance from a liver-intestine hybrid cell towards a hepatocyte.

5. Appendix

5.1 Supplementary figures

Due to space and format limitations, some of the material described in this appendix is provided in the digital version of the appendix.
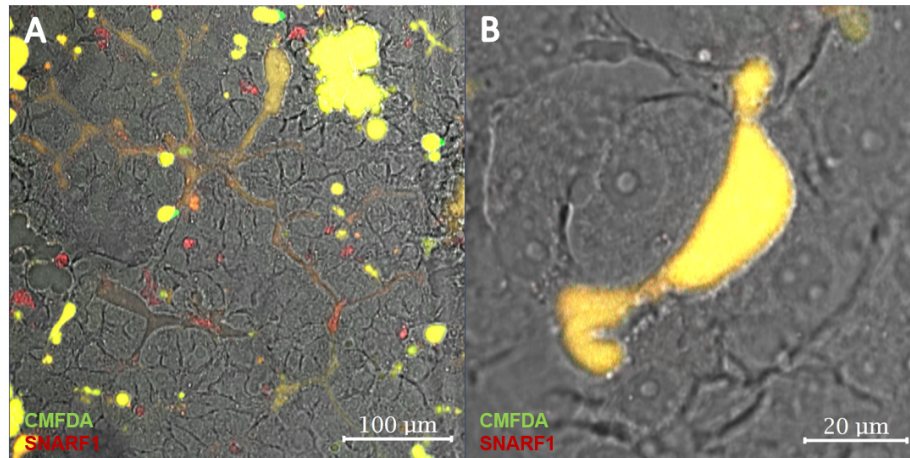
Supplementary figure 1: CMFDA (green) and SNARF-1 are excreted from day 25 HLC into bile canalicular structures of different size and shape.
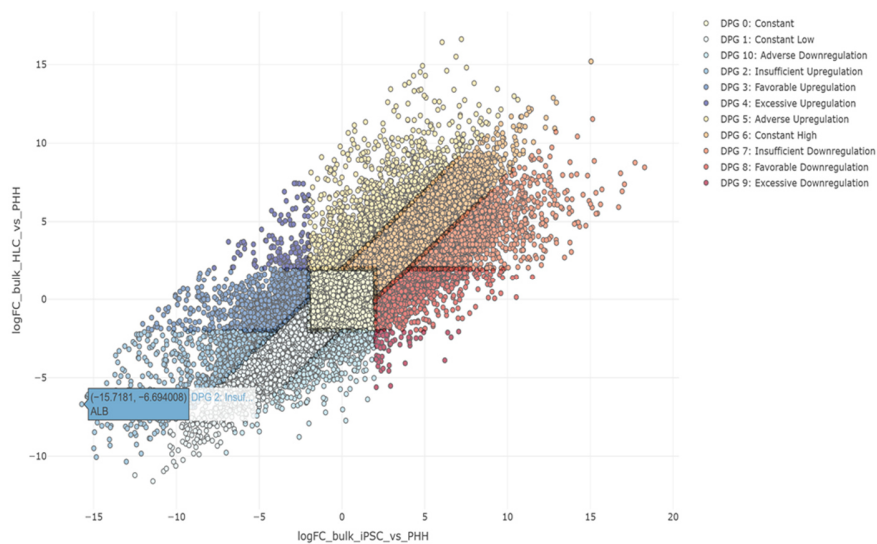
Supplementary figure 2: Interactive version of the DiPa plot provided as .html file of the digital appendix. The x-axis represents $\log_2$ fold changes of iPSC over PHH, the y-axis indicates $\log_2$ fold changes of HLC over PHH. Dotted lines represent cutoffs of the clustering approach. Gene expression can be explored by hovering over a gene with the computer mouse. The figure was generated using the R implementation of plot.ly.
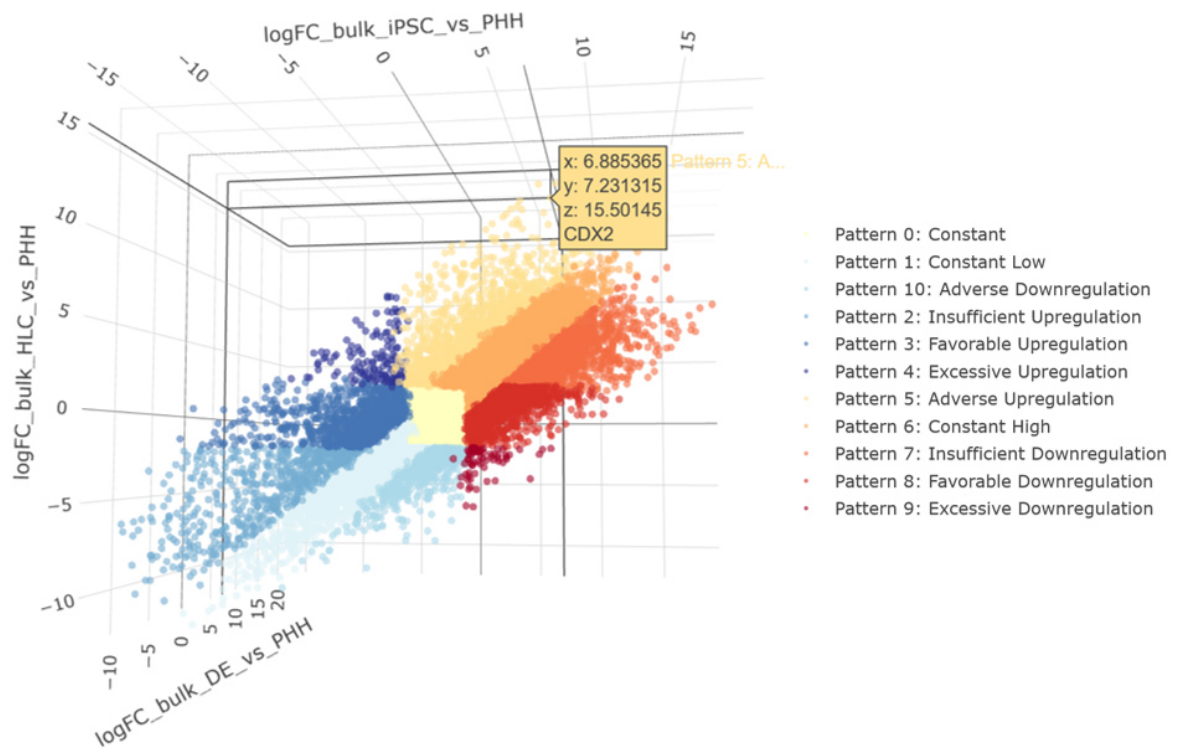
Supplementary figure 3: Interactive 3D version of the DiPa plot provided as .html file of the digital appendix. Definitive Endoderm expression was included as log2 fold change over PHH. The x-axis represents log_2 fold changes of iPSC over PHH, the y-axis indicates log_2 fold changes of DE over PHH. The z-axis indicates log_2 fold changes of HLC over PHH. Dotted lines represent cutoffs of the clustering approach. Gene expression can be explored by hovering over a gene with the computer mouse. The figure was generated using the R implementation of plot.ly.
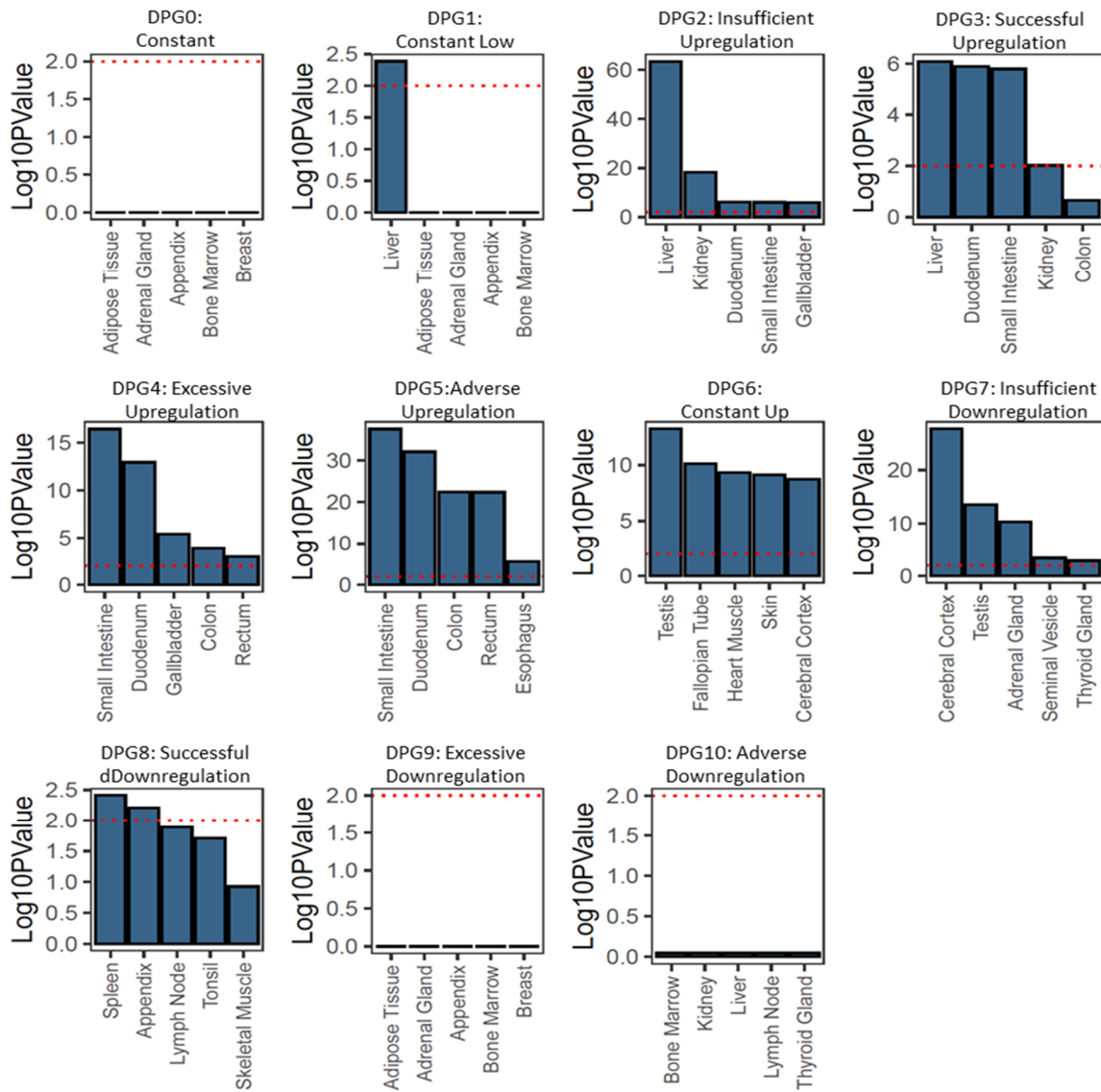
Supplementary figure 4: Tissue group enrichment according to the human protein atlas in DPGs established by Differentiation Pattern Clustering (DiPaC). The top 5 most significant results are displayed per DPG. The dotted red line indicates a p-Value cutoff of 0.01. The analysis was performed using the R TissueEnrich package version 1.8 by Jain and Tuteja 2019.
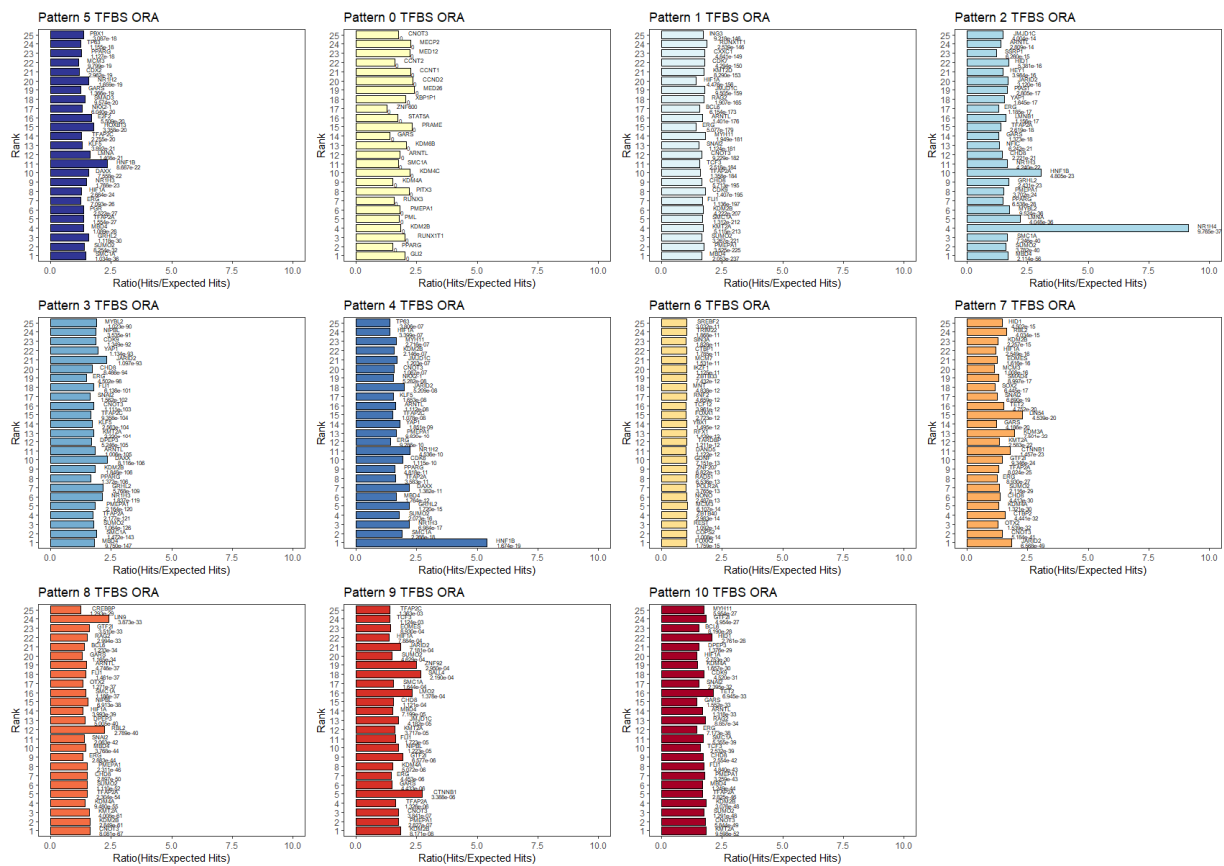
Supplementary figure 5: Transcriptional regulator overrepresentation in each DPG according to Regulatortrail overrepresentation analysis (Kehl et al. 2017). Names of regulators and the P-value associated with their enrichment are annotated for every bar. The x-axis represents the ratio of hits (genes associated with a specific regulator in the DPG) and expected hits. The y-axis represents the rank of overrepresentation of the transcriptional regulator in the corresponding DPG.
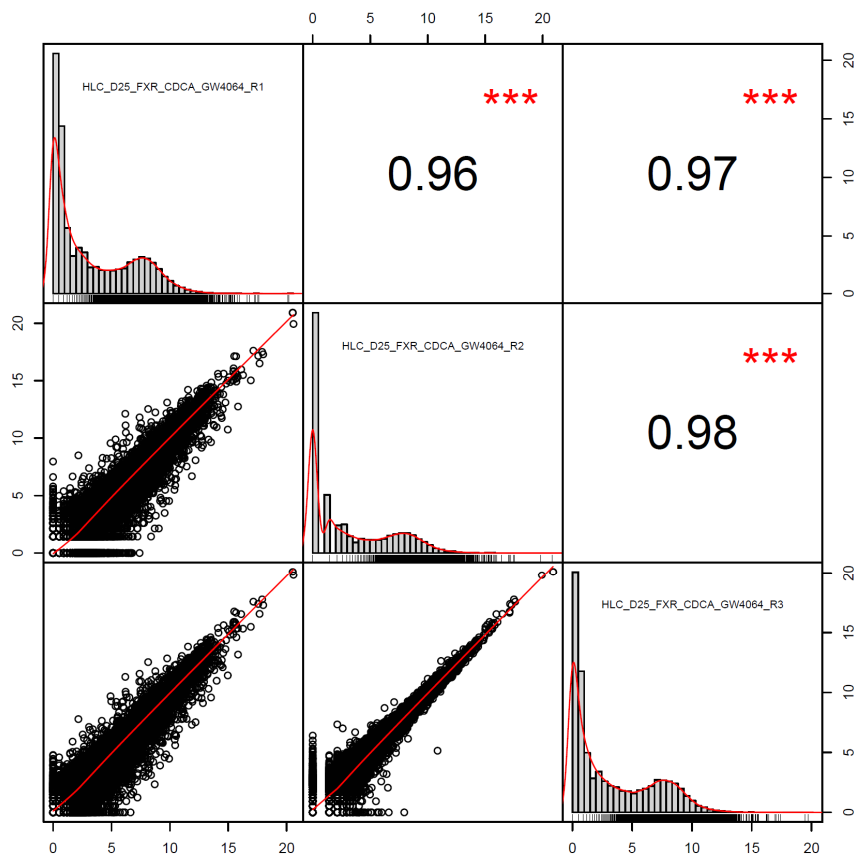
Supplementary figure 6: Gene expression Pearson correlation and gene count distribution of $\log_1$ normalized counts determined by DEseq2 for RNA-seq data for all three replicates of HLC transduced with FXR expression lentivirus and treated with the FXR agonists chenodeoxycholic acid (CDCA) and GW4064. Red stars indicate significance of the P-value obtained for the correlation analysis (*** = 0-0.001).
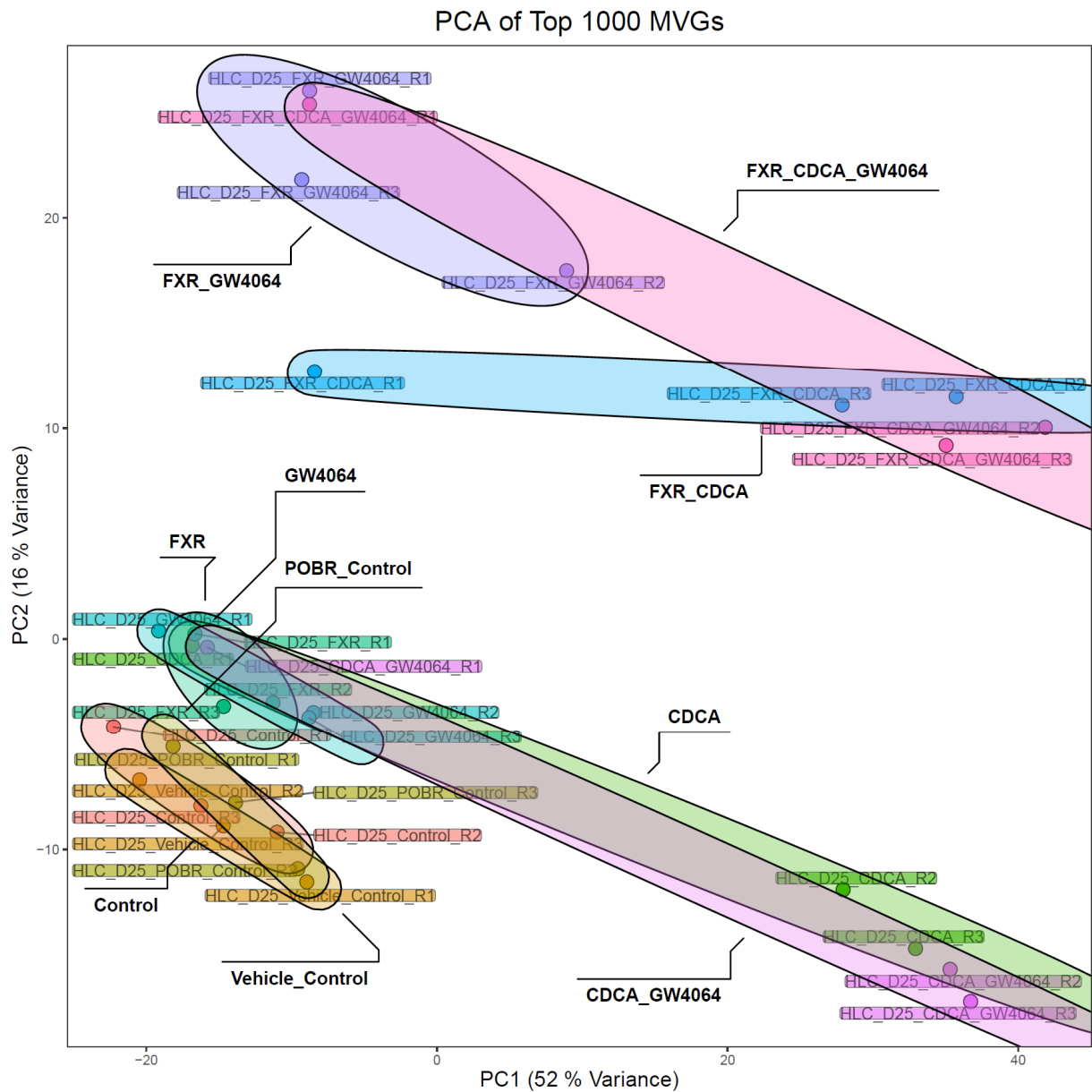
Supplementary figure 7: Principal component analysis (PCA) of the top 1000 variable genes in HLC at day 25 and HLC at day 25 that were treated either with CDCA, GW4064, an FXR expressing lentivirus or a combination thereof. Colors of ellipses and dots indicate the treatment, labels indicate the sample description and replicate number (R).
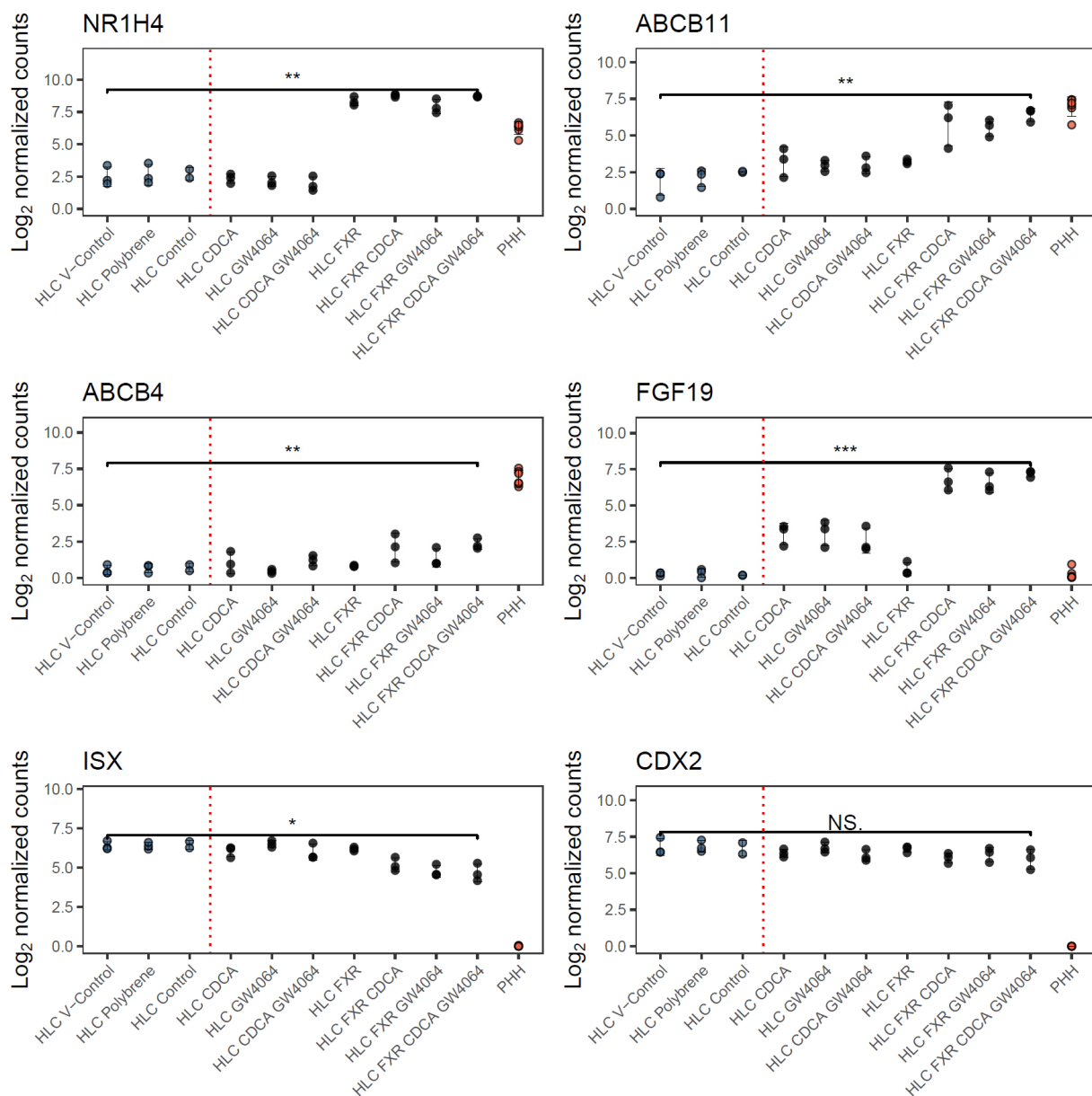
Supplementary figure 8: Absolute expression given as log$_2$ normalized counts determined with DEseq2 of selected genes representative of liver (NR1H4, ABCB11, ABCB4) and intestine (FGF19, ISX, SI) responses to FXR and agonist intervention treatments. The red dotted line separates control (left) and treatment samples (right). Error bars indicate standard deviation (n=3).

5.2 Supplementary tables

Supplementary tables are available in the digital version of the appendix.

Supplementary table 1: Differentially expressed genes in human induced plutripotent stem cells (iPSC) that were differentiated to definitive endoderm (DE) and subsequently to hepatocyte-like cells (HLC), compared to primary hepatocytes (PHH). Fold change differences (Log2) are given with corresponding p-values and FDR adjusted p-values.

Supplementary table 2: Annotation of differentially expressed genes to corresponding differentiation pattern groups (DPGs).

Supplementary table 3: Tissue group enrichment analysis of differentiation pattern groups (DPG0-10) defined by Differentiation Pattern Clustering (DiPaC) of differentially expressed genes (RNA-seq) in iPSC to HLC differentiation compared to PHH, according to Jain, A., & Tuteja, G. (2018).

Supplementary table 4: GO enrichment results for DPG0-10 as obtained by enrichGO (Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012).

Supplementary table 5: Regulatortrail overrepresentation analysis of DPG0-10 according to Kehl et al. (2017).

Supplementary table 6: Reactome pathway enrichment analysis of DPG0-10 according to Yu, G., & He, Q. Y. (2016).

Supplementary table 7: Z-Scores calculated separately for RNA-seq, RRBS and ATAC-seq data of iPSC, DE, HLC and PHH.

Supplementary table 8: Single cell gene expression ($\log_2$CPM + 1) matrix for all cells analyzed and corresponding metadata.

Supplementary table 9: Mean scaled expression (z-scores) for hybrid state genes found in HLC from the present study and HLC from Camp et al. 2017, in human induced pluripotent stem cells (iPSC), fetal human hepatocytes week 10 (FH week 10), fetal human hepatocytes week 17 (FH week 17) and hepatocyte-like cells (HLC).

Supplementary table 10: SCENIC Regulon activity scores of human induced pluripotent stem cells (iPSC), fetal human hepatocytes week 10 (FH week 10), fetal human hepatocytes week 17 (FH week 17), hepatocyte-like cells (HLC) from Camp et al. 2017 and HLC and primary human hepatocytes (PHH) of the present study.

Supplementary table 11: Differentially expressed genes of hepatocyte-like cells (HLC; Control), HLC treated with polybrene (POBR Control), HLC subjected to lentiviral transduction with FXR and/or treatment with synthetic FXR agonist GW4064 or the endogenous FXR agonist chenodeoxycholic acid (CDCA) or both, compared to vehicle control (VC). Fold change differences (Log2) are given with corresponding p-values and fdr adjusted p-values.

Supplementary table 12: P-values of Fisher tests testing the null hypothesis, that there are less genes in the overlap than expected by chance.

5.3 Supplementary files

Supplementary file 1.1: Time-series of CMFDA (green) and SNRAF-1 (red) excretion in HLC day 25 captured with the LSM 880 at 20x magnification (available as .avi and .czi with preserved imaging data).

Supplementary file 1.2: Time-series of CMFDA (green) and SNRAF-1 (red) excretion in PHH at day 3 of culture captured with the LSM 880 at 20x magnification (available as .avi and .czi with preserved imaging data).

Supplementary file 1.3: Time-series of CMFDA (green) and SNRAF-1 (red) excretion in HLC day 25 of culture, showing bile canaliculi and other duct-like structures of various sizes. Captured with the LSM 880 at 20x magnification (available as .avi and .czi with preserved imaging data).

Supplementary file 1.4: 3D stack visualization of CMFDA (green) and SNRAF-1 (red) excretion in HLC day 25 of culture into a bile-canaliculi-network structure that surrounds multiple HLC. Captured with the LSM 880 at 20x magnification (available as .avi and .czi with preserved imaging data).

Supplementary file 2: GO term overrepresentation barplots can be found in the digital appendix. For each DPG, top 100 GO terms are displayed, sorted by the adjusted P-value associated with their enrichment. The Adjusted P-value cutoff of 0.001 is indicated by a red dotted line. The x-axis represents the adjusted P-value given as -log10(Adjusted P-value). The color scale indicated the ratio of hits (number of genes associated with a specific GO term) and background genes (number of genes in the DPG).

Supplementary file 3: Reactome pathway enrichment bar plots are provided in the digital appendix. For each DPG, the top 100 Reactome pathways are displayed, sorted by the adjusted P-value associated with their enrichment. The Adjusted P-value cutoff of 0.001 is indicated by a red dotted line. The x-axis represents the adjusted P-value given as -log10(Adjusted P-value). The color scale indicated the ratio of hits (number of genes associated with a specific Reactome pathway) and background genes (number of genes in the DPG).

Supplementary file 4: T-SNE representations of the integrated single cell dataset with perplexity parameter 1-50 and top 100-2000 variable genes. HLC Replicates 1 and 2 are shown together with PHH

from donors 1, 2 and 3 together with single cell sequencing data of fetal hepatocytes week 10 and 17 from Camp et al. (2017).

6. Availability of data and materials

The RNA-seq, sc-RNA-seq, RRBS, and ATAC-seq datasets are available in the EGA repository under accession EGAS00001004201.

7. List of figures

8. List of tables

9. References

[1] Owen RD. Immunogenetic consequences of vascular anastomoses between bovine twins. Science (80-) 1945;102:400–1. https://doi.org/10.1126/science.102.2651.400.

[2] TILL JE, McCULLOCH EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res 1961;14:213–22. https://doi.org/10.2307/3570892.

[3] Cheng H, Merzel J, Leblond CP. Renewal of Paneth cells in the small intestine of the mouse. Am J Anat 1969;126:507–25. https://doi.org/10.1002/aja.1001260409.

[4] Cheng H, Leblond CP. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine V. Unitarian theory of the origin of the four epithelial cell types. Am J Anat 1974;141:537–61. https://doi.org/10.1002/aja.1001410407.

[5] Lewis PD. Mitotic activity in the primate subependymal layer and the genesis of gliomas. Nature 1968;217:974–5. https://doi.org/10.1038/217974a0.

[6] Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci 1981;78:7634–8. https://doi.org/10.1073/pnas.78.12.7634.

[7] Thomson JA. Embryonic stem cell lines derived from human blastocysts. Science (80-) 1998;282:1145–7. https://doi.org/10.1126/science.282.5391.1145.

[8] Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 2006;126:663–76. https://doi.org/10.1016/j.cell.2006.07.024.

[9] Takahashi K, Okita K, Nakagawa M, Yamanaka S. Induction of pluripotent stem cells from fibroblast cultures. Nat Protoc 2007;2:3081–9. https://doi.org/10.1038/nprot.2007.418.

[10] Thomas ED, Buckner CD, Clift RA, Fefer A, Johnson FL, Neiman PE, et al. Marrow Transplantation for Acute Nonlymphoblastic Leukemia in First Remission. N Engl J Med 1979;301:597–9. https://doi.org/10.1056/NEJM197909133011109.

[11]   Statistics | Eurostat n.d. https://ec.europa.eu/eurostat/databrowser/view/HLTH_CO_PROC2__custom_1044243/settings_1/table?lang=en (accessed June 9, 2021).

[12]   Scott JG, Dhawan A, Hjelmeland A, Lathia J, Chumakova A, Hitomi M, et al. Recasting the Cancer Stem Cell Hypothesis: Unification Using a Continuum Model of Microenvironmental Forces. Curr Stem Cell Reports 2019;5:22–30. https://doi.org/10.1007/s40778-019-0153-0.

[13]   Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105–11. https://doi.org/10.1038/35102167.

[14]   Magner LN. A history of the life sciences, revised and expanded. CRC Press; 2002.

[15]   Eguizabal C, Aran B, Chuva de Sousa Lopes SM, Geens M, Heindryckx B, Panula S, et al. Two decades of embryonic stem cells: a historical overview. Hum Reprod Open 2019;2019:1–17. https://doi.org/10.1093/hropen/hoy024.

[16]   Aly RM. Current state of stem cell-based therapies: An overview. Stem Cell Investig 2020;7. https://doi.org/10.21037/sci-2020-001.

[17]   Chari S, Nguyen A, Saxe J. Stem Cells in the Clinic. Cell Stem Cell 2018;22:781–2. https://doi.org/10.1016/j.stem.2018.05.017.

[18]   Waddington CH. The strategy of the genes: a discussion of some aspects of theoretical biology. 1957.

[19]   Hemberger M, Dean W, Reik W. Epigenetic dynamics of stem cells and cell lineage commitment: Digging Waddington's canal. Nat Rev Mol Cell Biol 2009;10:526–37. https://doi.org/10.1038/nrm2727.

[20]   Takahashi K. Cellular reprogramming - lowering gravity on Waddington's epigenetic landscape. J Cell Sci 2012;125:2553–60. https://doi.org/10.1242/jcs.084822.

[21]   Noble D. Conrad Waddington and the origin of epigenetics. J Exp Biol 2015;218:816–8. https://doi.org/10.1242/jeb.120071.

[22] Gurdon JB, Elsdale TR, Fischberg M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature 1958;182:64–5. https://doi.org/10.1038/182064a0.

[23] Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KHS. Viable offspring derived from fetal and adult mammalian cells. Nature 1997;385:810–3. https://doi.org/10.1038/385810a0.

[24] Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 1987;51:987–1000. https://doi.org/10.1016/0092-8674(87)90585-X.

[25] Schaub JR, Huppert KA, Kurial SNT, Hsu BY, Cast AE, Donnelly B, et al. De novo formation of the biliary system by TGFβ-mediated hepatocyte transdifferentiation. Nature 2018;557:247–51. https://doi.org/10.1038/s41586-018-0075-5.

[26] Vugler A, Carr AJ, Lawrence J, Chen LL, Burrell K, Wright A, et al. Elucidating the phenomenon of ESC-derived RPE: Anatomy of cell genesis, expansion and retinal transplantation. Exp Neurol 2008;214:347–61. https://doi.org/10.1016/j.expneurol.2008.09.007.

[27] Shroff G. Human Embryonic Stem Cell Therapy in Chronic Spinal Cord Injury: A Retrospective Study. Clin Transl Sci 2016;9:168–75. https://doi.org/10.1111/cts.12394.

[28] Gardner RL. Mouse chimaeras obtained by the injection of cells into the blastocyst. Nature 1968;220:596–7. https://doi.org/10.1038/220596a0.

[29] Ilic D, Ogilvie C. Concise Review: Human Embryonic Stem Cells-What Have We Done? What Are We Doing? Where Are We Going? Stem Cells 2017;35:17–25. https://doi.org/10.1002/stem.2450.

[30] Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, et al. Modelling the long QT syndrome with induced pluripotent stem cells. Nature 2011;471:225–30. https://doi.org/10.1038/nature09747.

[31] Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and

familial Alzheimer's disease using induced pluripotent stem cells. Nature 2012;482:216–20. https://doi.org/10.1038/nature10821.

[32] M M, Y K, M T. Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration. N Engl J Med 2017;377:792–3. https://doi.org/10.1056/nejmc1706274.

[33] Takahashi J. iPS cell-based therapy for Parkinson's disease: A Kyoto trial. Regen Ther 2020;13:18–22. https://doi.org/10.1016/j.reth.2020.06.002.

[34] Takebe T, Wells JM, Helmrath MA, Zorn AM. Organoid Center Strategies for Accelerating Clinical Translation. Cell Stem Cell 2018;22:806–9. https://doi.org/10.1016/j.stem.2018.05.008.

[35] Qiao Y, Agboola OS, Hu X, Wu Y, Lei L. Tumorigenic and Immunogenic Properties of Induced Pluripotent Stem Cells: a Promising Cancer Vaccine. Stem Cell Rev Reports 2020;16:1049–61. https://doi.org/10.1007/s12015-020-10042-5.

[36] Feng W, Dai Y, Mou L, Cooper D, Shi D, Cai Z. The Potential of the Combination of CRISPR/Cas9 and Pluripotent Stem Cells to Provide Human Organs from Chimaeric Pigs. Int J Mol Sci 2015;16:6545–56. https://doi.org/10.3390/ijms16036545.

[37] Ma H, Marti-Gutierrez N, Park SW, Wu J, Lee Y, Suzuki K, et al. Correction of a pathogenic gene mutation in human embryos. Nature 2017;548:413–9. https://doi.org/10.1038/nature23305.

[38] Edwards RG, Beard HK. Oocyte polarity and cell determination in early mammalian embryos. Mol Hum Reprod 1997;3:863–905. https://doi.org/10.1093/molehr/3.10.863.

[39] de Paepe C, Krivega M, Cauffman G, Geens M, van de Velde H. Totipotency and lineage segregation in the human embryo. Mol Hum Reprod 2014;20:599–618. https://doi.org/10.1093/molehr/gau027.

[40] Mann MRW, Bartolomei MS. Epigenetic reprogramming in the mammalian embryo: Struggle of the clones. Genome Biol 2002;3:1–4. https://doi.org/10.1186/gb-2002-3-2-reviews1003.

[41] Curradi M, Izzo A, Badaracco G, Landsberger N. Molecular mechanisms of gene silencing mediated by DNA methylation. Mol Cell Biol 2002;22:3157–73. https://doi.org/10.1128/mcb.22.9.3157-3173.2002.

[42] Wu F, Wu Q, Li D, Zhang Y, Wang R, Liu Y, et al. Oct4 regulates DNA methyltransferase 1 transcription by direct binding of the regulatory element. Cell Mol Biol Lett 2018;23:39. https://doi.org/10.1186/s11658-018-0104-2.

[43] Raab S, Klingenstein M, Liebau S, Linta L. A Comparative View on Human Somatic Cell Sources for iPSC Generation. Stem Cells Int 2014;2014. https://doi.org/10.1155/2014/768391.

[44] Yoshioka N, Gros E, Li HR, Kumar S, Deacon DC, Maron C, et al. Efficient generation of human iPSCs by a synthetic self-replicative RNA. Cell Stem Cell 2013;13:246–54. https://doi.org/10.1016/j.stem.2013.06.001.

[45] Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, et al. A more efficient method to generate integration-free human iPS cells. Nat Methods 2011;8:409–12. https://doi.org/10.1038/nmeth.1591.

[46] Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. Nature 2010;467:285–90. https://doi.org/10.1038/nature09342.

[47] Noguchi H, Miyagi-Shiohira C, Nakashima Y. Induced tissue-specific stem cells and epigenetic memory in induced pluripotent stem cells. Int J Mol Sci 2018;19. https://doi.org/10.3390/ijms19040930.

[48] Daikeler T, Hügle T, Farge D, Andolina M, Gualandi F, Baldomero H, et al. Allogeneic hematopoietic SCT for patients with autoimmune diseases. Bone Marrow Transplant 2009;44:27–33. https://doi.org/10.1038/bmt.2008.424.

[49] Hernigou P, Poignard A, Manicom O, Mathieu G, Rourd H. The use of percutaneous autologous bone marrow transplantation in nonunion and avascular necrosis of bone. J Bone Jt Surg - Ser B 2005;87:896–902. https://doi.org/10.1302/0301-620X.87B7.16289.

[50] Miyahara Y, Nagaya N, Kataoka M, Yanagawa B, Tanaka K, Hao H, et al. Monolayered

mesenchymal stem cells repair scarred myocardium after myocardial infarction. Nat Med 2006;12:459–65. https://doi.org/10.1038/nm1391.

[51]   Steinhoff G, Nesteruk J, Wolfien M, Kundt G, Börgermann J, David R, et al. Cardiac Function Improvement and Bone Marrow Response –: Outcome Analysis of the Randomized PERFECT Phase III Clinical Trial of Intramyocardial CD133+ Application After Myocardial Infarction. EBioMedicine 2017;22:208–24. https://doi.org/10.1016/j.ebiom.2017.07.022.

[52]   Lehmann R. GermLine stem cells: Origin and destiny. Cell Stem Cell 2012;10:729–39. https://doi.org/10.1016/j.stem.2012.05.016.

[53]   Blanpain C, Fuchs E. Epidermal stem cells of the skin. Annu Rev Cell Dev Biol 2006;22:339–73. https://doi.org/10.1146/annurev.cellbio.22.010305.104357.

[54]   Lautt WW. Overview 2009.

[55]   Li F, Tian Z. The liver works as a school to educate regulatory immune cells. Cell Mol Immunol 2013;10:292–302. https://doi.org/10.1038/cmi.2013.7.

[56]   MATSUMURA T, THURMAN RG. Predominance of glycolysis in pericentral regions of the liver lobule. Eur J Biochem 1984;140:229–34. https://doi.org/10.1111/j.1432-1033.1984.tb08091.x.

[57]   Chew SA, Moscato S, George S, Azimi B, Danti S. Liver cancer: Current and future trends using biomaterials. Cancers (Basel) 2019;11:2026. https://doi.org/10.3390/cancers11122026.

[58]   Adams DH, Eksteen B. Aberrant homing of mucosal T cells and extra-intestinal manifestations of inflammatory bowel disease. Nat Rev Immunol 2006;6:244–51. https://doi.org/10.1038/nri1784.

[59]   Kusminski CM, Scherer PE. New zoning laws enforced by glucagon. Proc Natl Acad Sci U S A 2018;115:4308–10. https://doi.org/10.1073/pnas.1804203115.

[60]   Bilir BM, Gong TWL, Kwasiborski V, Shen CS, Fillmore CS, Berkowitz CM, et al. Novel

control of the position-dependent expression of genes in hepatocytes. The GLUT-1 transporter. vol. 268. 1993. https://doi.org/10.1016/S0021-9258(19)36581-0.

[61] Kmiec Z. Cooperation of liver cells in health and disease. Adv Anat Embryol Cell Biol 2001;161. https://doi.org/10.1007/978-3-642-56553-3.

[62] Stanca C, Jung D, Meier PJ, Kullak-Ublick GA. Hepatocellular transport proteins and their role in liver disease. World J Gastroenterol 2001;7:157–69. https://doi.org/10.3748/wjg.v7.i2.157.

[63] Nagarajan SR, Paul-Heng M, Krycer JR, Fazakerley DJ, Sharland AF, Hoy AJ. Lipid and glucose metabolism in hepatocyte cell lines and primary mouse hepatocytes: a comprehensive resource for in vitro studies of hepatic metabolism. Am J Physiol Metab 2019;316:E578–89. https://doi.org/10.1152/ajpendo.00365.2018.

[64] Zhou Z, Xu MJ, Gao B. Hepatocytes: A key cell type for innate immunity. Cell Mol Immunol 2016;13:301–15. https://doi.org/10.1038/cmi.2015.97.

[65] Thomson AW, Knolle PA. Antigen-presenting cell function in the tolerogenic liver environment. Nat Rev Immunol 2010;10:753–66. https://doi.org/10.1038/nri2858.

[66] Hundt M, John S. Physiology, Bile Secretion. StatPearls Publishing; 2018.

[67] Tabibian JH, Masyuk AI, Masyuk T V., O'Hara SP, LaRusso NF. Physiology of cholangiocytes. Compr Physiol 2013;3:541–65. https://doi.org/10.1002/cphy.c120019.

[68] Roberts MS, Magnusson BM, Burczynski FJ, Weiss M. Enterohepatic circulation: Physiological, pharmacokinetic and clinical implications. Clin Pharmacokinet 2002;41:751–90. https://doi.org/10.2165/00003088-200241100-00005.

[69] Dixon LJ, Barnes M, Tang H, Pritchard MT, Nagy LE. Kupffer cells in the liver. Compr Physiol 2013;3:785–97. https://doi.org/10.1002/cphy.c120026.

[70] Crispe IN. Liver antigen-presenting cells. J Hepatol 2011;54:357–65. https://doi.org/10.1016/j.jhep.2010.10.005.

[71] Wisse E, Luo D, Vermijlen D, Kanellopoulou C, De Zanger R, Braet F. On the function of pit

cells, the liver-specific natural killer cells. Semin Liver Dis 1997;17:265–86. https://doi.org/10.1055/s-2007-1007204.

[72]   Geerts A. History, heterogeneity, developmental biology, and functions of quiescent hepatic stellate cells. Semin Liver Dis 2001;21:311–35. https://doi.org/10.1055/s-2001-17550.

[73]   Rockey DC. Hepatic blood flow regulation by stellate cells in normal and injured liver. Semin Liver Dis 2001;21:337–49. https://doi.org/10.1055/s-2001-17551.

[74]   Li W, Li L, Hui L. Cell Plasticity in Liver Regeneration. Trends Cell Biol 2020;30:329–38. https://doi.org/10.1016/j.tcb.2020.01.007.

[75]   Michalopoulos GK. Liver regeneration. J Cell Physiol 2007;213:286–300. https://doi.org/10.1002/jcp.21172.

[76]   Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. J Hepatol 2019;70:151–71. https://doi.org/10.1016/j.jhep.2018.09.014.

[77]   Iansante V, Mitry RR, Filippi C, Fitzpatrick E, Dhawan A. Human hepatocyte transplantation for liver disease: Current status and future perspectives. Pediatr Res 2018;83:232–40. https://doi.org/10.1038/pr.2017.284.

[78]   Sachinidis A, Albrecht W, Nell P, Cherianidou A, Hewitt NJ, Edlund K, et al. Road Map for Development of Stem Cell-Based Alternative Test Methods. Trends Mol Med 2019;25:470–81. https://doi.org/10.1016/j.molmed.2019.04.003.

[79]   Wang Y, Alhaque S, Cameron K, Meseguer-Ripolles J, Lucendo-Villarin B, Rashidi H, et al. Defined and Scalable Generation of Hepatocyte-like Cells from Human Pluripotent Stem Cells. J Vis Exp 2017:1–8. https://doi.org/10.3791/55355.

[80]   Godoy P, Schmidt-Heck W, Natarajan K, Lucendo-Villarin B, Szkolnicka D, Asplund A, et al. Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. J Hepatol 2015;63:934–42. https://doi.org/10.1016/j.jhep.2015.05.013.

[81]   Mitani S, Takayama K, Nagamoto Y, Imagawa K, Sakurai F, Tachibana M, et al. Human ESC/iPSC-Derived Hepatocyte-like Cells Achieve Zone-Specific Hepatic Properties by Modulation of WNT Signaling. Mol Ther 2017;25:1420–33. https://doi.org/10.1016/j.ymthe.2017.04.006.

[82]   Camp JG, Sekine K, Gerber T, Loeffler-Wirth H, Binder H, Gac M, et al. Multilineage communication regulates human liver bud development from pluripotency. Nature 2017;546:533–8. https://doi.org/10.1038/nature22796.

[83]   Pettinato G, Lehoux S, Ramanathan R, Salem MM, He L-X, Muse O, et al. Generation of fully functional hepatocyte-like organoids from human induced pluripotent stem cells mixed with Endothelial Cells. Sci Rep 2019;9:8920. https://doi.org/10.1038/s41598-019-45514-3.

[84]   Thompson WL, Takebe T. Generation of multi-cellular human liver organoids from pluripotent stem cells. Methods Cell Biol., vol. 159, Academic Press Inc.; 2020, p. 47–68. https://doi.org/10.1016/bs.mcb.2020.03.009.

[85]   Velazquez JJ, LeGraw R, Moghadam F, Tan Y, Kilbourne J, Maggiore JC, et al. Gene Regulatory Network Analysis and Engineering Directs Development and Vascularization of Multilineage Human Liver Organoids. Cell Syst 2020. https://doi.org/10.1016/j.cels.2020.11.002.

[86]   Gijbels E, Vanhaecke T, Vinken M. Establishment of sandwich cultures of primary human hepatocytes. Methods Mol. Biol., vol. 1981, Humana Press Inc.; 2019, p. 325–33. https://doi.org/10.1007/978-1-4939-9420-5_21.

[87]   Godoy P, Hewitt NJ, Albrecht U, Andersen ME, Ansari N, Bhattacharya S, et al. Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. Arch Toxicol 2013;87:1315–530. https://doi.org/10.1007/s00204-013-1078-5.

[88]   Godoy P, Schmidt-Heck W, Hellwig B, Nell P, Feuerborn D, Rahnenführer J, et al.

Assessment of stem cell differentiation based on genome-wide expression profiles. Philos Trans R Soc B Biol Sci 2018;373. https://doi.org/10.1098/rstb.2017.0221.

[89] Bell CC, Dankers ACA, Lauschke VM, Sison-Young R, Jenkins R, Rowe C, et al. Comparison of Hepatic 2D Sandwich Cultures and 3D Spheroids for Long-term Toxicity Applications: A Multicenter Study. Toxicol Sci 2018;162:655–66. https://doi.org/10.1093/toxsci/kfx289.

[90] Shinde V, Hoelting L, Srinivasan SP, Meisig J, Meganathan K, Jagtap S, et al. Definition of transcriptome-based indices for quantitative characterization of chemically disturbed stem cell development: introduction of the STOP-Toxuknand STOP-Toxukktests. Arch Toxicol 2017;91:839–64. https://doi.org/10.1007/s00204-016-1741-8.

[91] Tam PPL, Loebel DAF. Gene function in mouse embryogenesis: Get set for gastrulation. Nat Rev Genet 2007;8:368–81. https://doi.org/10.1038/nrg2084.

[92] Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol 2009;25:221–51. https://doi.org/10.1146/annurev.cellbio.042308.113344.

[93] Chazaud C, Yamanaka Y. Lineage specification in the mouse preimplantation embryo. Dev 2016;143:1063–74. https://doi.org/10.1242/dev.128314.

[94] Drukker M, Tang C, Ardehali R, Rinkevich Y, Seita J, Lee AS, et al. Isolation of primitive endoderm, mesoderm, vascular endothelial and trophoblast progenitors from human pluripotent stem cells. Nat Biotechnol 2012;30:531–42. https://doi.org/10.1038/nbt.2239.

[95] Perea-Gomez A, Meilhac SM. Formation of the Anterior-Posterior Axis in Mammals. Princ. Dev. Genet. Second Ed., Elsevier Inc.; 2015, p. 171–88. https://doi.org/10.1016/B978-0-12-405945-0.00010-7.

[96] Schöck F, Perrimon N. Molecular Mechanisms of Epithelial Morphogenesis. Annu Rev Cell Dev Biol 2002;18:463–93. https://doi.org/10.1146/annurev.cellbio.18.022602.131838.

[97] Alexanian M, Maric D, Jenkinson SP, Mina M, Friedman CE, Ting CC, et al. A transcribed enhancer dictates mesendoderm specification in pluripotency. Nat Commun 2017;8:1–19. https://doi.org/10.1038/s41467-017-01804-w.

[98] Nowotschin S, Setty M, Kuo YY, Liu V, Garg V, Sharma R, et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. Nature 2019;569:361–7. https://doi.org/10.1038/s41586-019-1127-1.

[99] Morgani SM, Hadjantonakis AK. Signaling regulation during gastrulation: Insights from mouse embryos and in vitro systems. Curr. Top. Dev. Biol., vol. 137, Academic Press Inc.; 2020, p. 391–431. https://doi.org/10.1016/bs.ctdb.2019.11.011.

[100] Pauklin S, Vallier L. Activin/nodal signalling in stem cells. Dev 2015;142:607–19. https://doi.org/10.1242/dev.091769.

[101] McCauley HA, Wells JM. Pluripotent stem cell-derived organoids: Using principles of developmental biology to grow human tissues in a dish. Dev 2017;144:958–62. https://doi.org/10.1242/dev.140731.

[102] Burtscher I, Lickert H. Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. Development 2009;136:1029–38. https://doi.org/10.1242/dev.028415.

[103] Davenport C, Diekmann U, Budde I, Detering N, Naujok O. Anterior-Posterior Patterning of Definitive Endoderm Generated from Human Embryonic Stem Cells Depends on the Differential Signaling of Retinoic Acid, WNT-, and BMP-Signaling. Stem Cells 2016;34:2635–47. https://doi.org/10.1002/stem.2428.

[104] Zorn AM. Liver development 2008. https://doi.org/10.3824/STEMBOOK.1.25.1.

[105] Grapin-Botton A. Antero-posterior patterning of the vertebrate digestive tract: 40 Years after Nicole Le Douarin's PhD thesis. Int J Dev Biol 2005;49:335–47. https://doi.org/10.1387/ijdb.041946ag.

[106] Woo J, Miletich I, Kim B-M, Sharpe PT, Shivdasani RA. Barx1-Mediated Inhibition of WNT Signaling in the Mouse Thoracic Foregut Controls Tracheo-Esophageal Septation and Epithelial Differentiation. PLoS One 2011;6:e22493. https://doi.org/10.1371/journal.pone.0022493.

[107] Kim BM, Buchner G, Miletich I, Sharpe PT, Shivdasani RA. The stomach mesenchymal

transcription factor barx1 specifies gastric epithelial identity through inhibition of transient WNT signaling. Dev Cell 2005;8:611–22. https://doi.org/10.1016/j.devcel.2005.01.015.

[108] Zhang Z, Rankin SA, Zorn AM. Different thresholds of WNT-Frizzled 7 signaling coordinate proliferation, morphogenesis and fate of endoderm progenitor cells. Dev Biol 2013;378:1–12. https://doi.org/10.1016/j.ydbio.2013.02.024.

[109] Duncan SA, Watt AJ. BMPs on the road to hepatogenesis. Genes Dev 2001;15:1879–84. https://doi.org/10.1101/gad.920601.

[110] Rossi JM, Dunn NR, Hogan BLM, Zaret KS. Distinct mesodermal signals, including BMPs from the septum, transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. Genes Dev 2001;15:1998–2009. https://doi.org/10.1101/gad.904601.

[111] Jung J. Initiation of Mammalian Liver Development from Endoderm by Fibroblast Growth Factors. Science (80-) 1999;284:1998–2003. https://doi.org/10.1126/science.284.5422.1998.

[112] Zorn. Liver Development. StemBook 2008. https://doi.org/10.3824/stembook.1.25.1.

[113] Li J, Ning G, Duncan SA. Mammalian hepatocyte differentiation requires the transcription factor HNF-4α. Genes Dev 2000;14:464–74. https://doi.org/10.1101/gad.14.4.464.

[114] Lemaigre FP. Development of the biliary tract. Mech Dev 2003;120:81–7. https://doi.org/10.1016/S0925-4773(02)00334-9.

[115] Sergi C, Kahl P, Otto HF. Contribution of apoptosis and apoptosis-related proteins to the malformation of the primitive intrahepatic biliary system in Meckel syndrome. Am J Pathol 2000;156:1589–98. https://doi.org/10.1016/S0002-9440(10)65031-6.

[116] Miyajima A, Kinoshita T, Tanaka M, Kamiya A, Mukouyama Y, Hara T. Role of oncostatin M in hematopoiesis and liver development. Cytokine Growth Factor Rev 2000;11:177–83. https://doi.org/10.1016/S1359-6101(00)00003-4.

[117]   Kamiya A, Kinoshita T, Miyajima A. Oncostatin M and hepatocyte growth factor induce hepatic maturation via distinct signaling pathways. FEBS Lett 2001;492:90–4. https://doi.org/10.1016/S0014-5793(01)02140-8.

[118]   Nakamura T, Mizuno S. The discovery of Hepatocyte Growth Factor (HGF) and its significance for cell biology, life sciences and clinical medicine. Proc Japan Acad Ser B Phys Biol Sci 2010;86:588–610. https://doi.org/10.2183/pjab.86.588.

[119]   Schmidt C, Bladt F, Goedecke S, Brinkmann V, Zschiesche W, Sharpe M, et al. Scatter factor/hepatocyte growth factor is essential for liver development. Nature 1995;373:699–702. https://doi.org/10.1038/373699a0.

[120]   Han L, Chaturvedi P, Kishimoto K, Koike H, Nasr T, Iwasawa K, et al. Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. Nat Commun 2020;11:1–16. https://doi.org/10.1038/s41467-020-17968-x.

[121]   Mikkola HKA, Orkin SH. The journey of developing hematopoietic stem cells. Development 2006;133:3733–44. https://doi.org/10.1242/dev.02568.

[122]   Wang Y, Hay DC. Mass production of stem cell derived human hepatocytes for experimental medicine. Expert Rev Gastroenterol Hepatol 2016;10:769–71. https://doi.org/10.1080/17474124.2016.1182862.

[123]   Lucendo-Villarin B, Nell P, Hellwig B, Filis P, Feuerborn D, O'shaughnessy PJ, et al. GENOME-WIDE EXPRESSION CHANGES INDUCED BY BISPHENOL A, F AND S IN HUMAN STEM CELL DERIVED HEPATOCYTE-LIKE CELLS. EXCLI J 2020;19:1459–76. https://doi.org/10.17179/excli2020-2934.

[124]   Takayama K, Mizuguchi H. Generation of human pluripotent stem cell-derived hepatocyte-like cells for drug toxicity screening. Drug Metab Pharmacokinet 2017;32:12–20. https://doi.org/10.1016/j.dmpk.2016.10.408.

[125]   Toivonen S, Lundin K, Balboa D, Ustinov J, Tamminen K, Palgi J, et al. Activin A and WNT-dependent specification of human definitive endoderm cells. Exp Cell Res

2013;319:2535–44. https://doi.org/10.1016/j.yexcr.2013.07.007.

[126] Baxter M, Withey S, Harrison S, Segeritz CP, Zhang F, Atkinson-Dell R, et al. Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. J Hepatol 2015;62:581–9. https://doi.org/10.1016/j.jhep.2014.10.016.

[127] Li Q, Hutchins AP, Chen Y, Li S, Shan Y, Liao B, et al. A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes. Nat Commun 2017;8:1–12. https://doi.org/10.1038/ncomms15166.

[128] Yoney A, Etoc F, Ruzo A, Carroll T, Metzger JJ, Martyn I, et al. WNT signaling memory is required for ACTIVIN to function as a morphogen in human gastruloids. Elife 2018;7. https://doi.org/10.7554/eLife.38279.

[129] Naujok O, Diekmann U, Lenzen S. The Generation of Definitive Endoderm from Human Embryonic Stem Cells is Initially Independent from Activin A but Requires Canonical WNT-Signaling. Stem Cell Rev Reports 2014;10:480–93. https://doi.org/10.1007/s12015-014-9509-0.

[130] Bone HK, Nelson AS, Goldring CE, Tosh D, Welham MJ. A novel chemically directed route for the generation of definitive endoderm from human embryonic stem cells based on inhibition of GSK-3. Development 2011;138:1992–2000. https://doi.org/10.1242/jcs.081679.

[131] Gao X, Li R, Cahan P, Zhao Y, Yourick JJ, Sprando RL. Hepatocyte-like cells derived from human induced pluripotent stem cells using small molecules: Implications of a transcriptomic study. Stem Cell Res Ther 2020;11:1–21. https://doi.org/10.1186/s13287-020-01914-1.

[132] Snykers S, De Kock J, Rogiers V, Vanhaecke T. In Vitro Differentiation of Embryonic and Adult Stem Cells into Hepatocytes: State of the Art. Stem Cells 2009;27:577–605. https://doi.org/10.1634/stemcells.2008-0963.

[133] Cai J, Zhao Y, Liu Y, Ye F, Song Z, Qin H, et al. Directed differentiation of human embryonic

stem cells into functional hepatic cells. Hepatology 2007;45:1229–39. https://doi.org/10.1002/hep.21582.

[134] Song Z, Cai J, Liu Y, Zhao D, Yong J, Duo S, et al. Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. Cell Res 2009;19:1233–42. https://doi.org/10.1038/cr.2009.107.

[135] Ogawa S, Surapisitchat J, Virtanen C, Ogawa M, Niapour M, Sugamori KS, et al. Three-dimensional culture and cAMP signaling promote the maturation of human pluripotent stem cell-derived hepatocytes. Development 2013;140:3285–96. https://doi.org/10.1242/dev.090266.

[136] Brolén G, Sivertsson L, Björquist P, Eriksson G, Ek M, Semb H, et al. Hepatocyte-like cells derived from human embryonic stem cells specifically via definitive endoderm and a progenitor stage. J Biotechnol 2010;145:284–94. https://doi.org/10.1016/j.jbiotec.2009.11.007.

[137] Ameri J, Ståhlberg A, Pedersen J, Johansson JK, Johannesson MM, Artner I, et al. FGF2 specifies ESC-derived definitive endoderm into foregut/midgut cell lineages in a concentration-dependent manner. Stem Cells 2010;28:45–56. https://doi.org/10.1002/stem.249.

[138] Sullivan GJ, Hay DC, Park IH, Fletcher J, Hannoun Z, Payne CM, et al. Generation of functional human hepatic endoderm from human induced pluripotent stem cells. Hepatology 2010;51:329–35. https://doi.org/10.1002/hep.23335.

[139] Gao X, Li R, Cahan P, Zhao Y, Yourick JJ, Sprando RL. Hepatocyte-like cells derived from human induced pluripotent stem cells using small molecules: Implications of a transcriptomic study. Stem Cell Res Ther 2020;11:393. https://doi.org/10.1186/s13287-020-01914-1.

[140] Alizadeh E, Zarghami N, Eslaminejad MB, Akbarzadeh A, Barzegar A, Mohammadi SA. The effect of dimethyl sulfoxide on hepatic differentiation of mesenchymal stem cells. Artif Cells, Nanomedicine, Biotechnol 2016;44:157–64.

https://doi.org/10.3109/21691401.2014.928778.

[141]   Czysz K, Minger S, Thomas N. Dmso efficiently down regulates pluripotency genes in human embryonic stem cells during definitive endoderm derivation and increases the proficiency of hepatic differentiation. PLoS One 2015;10:1–16. https://doi.org/10.1371/journal.pone.0117689.

[142]   Pettinato G, Ramanathan R, Fisher RA, Mangino MJ, Zhang N, Wen X. Scalable Differentiation of Human iPSCs in a Multicellular Spheroid-based 3D Culture into Hepatocyte-like Cells through Direct WNT/β-catenin Pathway Inhibition. Sci Rep 2016;6:1–17. https://doi.org/10.1038/srep32888.

[143]   Chaudhari P, Tian L, Deshmukh A, Jang YY. Expression kinetics of hepatic progenitor markers in cellular models of human liver development recapitulating hepatocyte and biliary cell fate commitment. Exp Biol Med 2016;241:1653–62. https://doi.org/10.1177/1535370216657901.

[144]   Cameron K, Tan R, Schmidt-Heck W, Campos G, Lyall MJ, Wang Y, et al. Recombinant Laminins Drive the Differentiation and Self-Organization of ESC-Derived Hepatocytes. Stem Cell Reports 2015;5:1250–62. https://doi.org/10.1016/j.stemcr.2015.10.016.

[145]   Chen C, Soto-Gutierrez A, Baptista PM, Spee B. Biotechnology Challenges to In Vitro Maturation of Hepatic Stem Cells. Gastroenterology 2018;154:1258–72. https://doi.org/10.1053/j.gastro.2018.01.066.

[146]   Agarwal S, Holton KL, Lanza R. Efficient Differentiation of Functional Hepatocytes from Human Embryonic Stem Cells. Stem Cells 2008;26:1117–27. https://doi.org/10.1634/stemcells.2007-1102.

[147]   Kamiya A, Kinoshita T, Ito Y, Matsui T, Morikawa Y, Senba E, et al. Fetal liver development requires a paracrine action of oncostatin M through the gp130 signal transducer. EMBO J 1999;18:2127–36. https://doi.org/10.1093/emboj/18.8.2127.

[148]   Cameron K, Tan R, Schmidt-Heck W, Campos G, Lyall MJ, Wang Y, et al. Recombinant Laminins Drive the Differentiation and Self-Organization of ESC-Derived Hepatocytes.

Stem Cell Reports 2015;5:1250–62. https://doi.org/10.1016/j.stemcr.2015.10.016.

[149] Kuo T-C, Cho CH-H, Chen K-T, Shen C-N. Hepatic microRNAs promote generation of metabolically functioning hepatocytes from human pluripotent stem cells. Mech Dev 2017;145:S166. https://doi.org/10.1016/j.mod.2017.04.478.

[150] Takayama K, Inamura M, Kawabata K, Sugawara M, Kikuchi K, Higuchi M, et al. Generation of metabolically functioning hepatocytes from human pluripotent stem cells by FOXA2 and HNF1α transduction. J Hepatol 2012;57:628–36. https://doi.org/10.1016/j.jhep.2012.04.038.

[151] Baharvand H, Hashemi SM, Ashtiani SK, Farrokhi A. Differentiation of human embryonic stem cells into hepatocytes in 2D and 3D culture systems in vitro. Int J Dev Biol 2006;50:645–52. https://doi.org/10.1387/ijdb.052072hb.

[152] Lucendo-Villarin B, Meseguer-Ripolles J, Drew J, Fischer L, Ma E, Flint O, et al. Development of a cost-effective automated platform to produce human liver spheroids for basic and applied research. Biofabrication 2020;13:015009. https://doi.org/10.1088/1758-5090/abbdb2.

[153] Rashidi H, Luu N-T, Alwahsh SM, Ginai M, Alhaque S, Dong H, et al. 3D human liver tissue from pluripotent stem cells displays stable phenotype in vitro and supports compromised liver function in vivo. Arch Toxicol 2018;92:3117–29. https://doi.org/10.1007/s00204-018-2280-2.

[154] Estève J, Blouin JM, Lalanne M, Azzi-Martin L, Dubus P, Bidet A, et al. Generation of induced pluripotent stem cells-derived hepatocyte-like cells for ex vivo gene therapy of primary hyperoxaluria type 1. Stem Cell Res 2019;38:101467. https://doi.org/10.1016/j.scr.2019.101467.

[155] Ghosheh N, Küppers-Munther B, Asplund A, Edsbagge J, Ulfenborg B, Andersson TB, et al. Comparative transcriptomics of hepatic differentiation of human pluripotent stem cells and adult human liver tissue. Physiol Genomics 2017;49:430–46. https://doi.org/10.1152/physiolgenomics.00007.2017.

[156] Sachinidis A, Albrecht W, Nell P, Cherianidou A, Hewitt NJ, Edlund K, et al. Road Map for Development of Stem Cell-Based Alternative Test Methods. Trends Mol Med 2019;25:470–81. https://doi.org/10.1016/j.molmed.2019.04.003.

[157] Baxter M, Withey S, Harrison S, Segeritz CP, Zhang F, Atkinson-Dell R, et al. Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. J Hepatol 2015;62:581–9. https://doi.org/10.1016/j.jhep.2014.10.016.

[158] Viiri LE, Rantapero T, Kiamehr M, Alexanova A, Oittinen M, Viiri K, et al. Extensive reprogramming of the nascent transcriptome during iPSC to hepatocyte differentiation. Sci Rep 2019;9:1–12. https://doi.org/10.1038/s41598-019-39215-0.

[159] Cahan P, Li H, Morris SA, Lummertz Da Rocha E, Daley GQ, Collins JJ. CellNet: Network biology applied to stem cell engineering. Cell 2014;158:903–15. https://doi.org/10.1016/j.cell.2014.07.020.

[160] Radley AH, Schwab RM, Tan Y, Kim J, Lo EKW, Cahan P. Assessment of engineered cells using CellNet and RNA-seq. Nat Protoc 2017;12:1089–102. https://doi.org/10.1038/nprot.2017.022.

[161] Kim DS, Ryu JW, Son MY, Oh JH, Chung KS, Lee S, et al. A liver-specific gene expression panel predicts the differentiation status of in vitro hepatocyte models. Hepatology 2017;66:1662–74. https://doi.org/10.1002/hep.29324.

[162] Godoy P, Schmidt-Heck W, Hellwig B, Nell P, Feuerborn D, Rahnenführer J, et al. Assessment of stem cell differentiation based on genome-wide expression profiles. Philos Trans R Soc B Biol Sci 2018;373. https://doi.org/10.1098/rstb.2017.0221.

[163] Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 2013;10:1096–100. https://doi.org/10.1038/nmeth.2639.

[164] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding

proteins and nucleosome position. Nat Methods 2013;10:1213–8. https://doi.org/10.1038/nmeth.2688.

[165] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.

[166] Hovestadt V, Jones DTW, Picelli S, Wang W, Kool M, Northcott PA, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature 2014;510:537–41. https://doi.org/10.1038/nature13268.

[167] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

[168] Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: Fast, accurate and versatile alignment by filtration. Nat Methods 2012;9:1185–8. https://doi.org/10.1038/nmeth.2221.

[169] Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc 2012;7:1728–40. https://doi.org/10.1038/nprot.2012.101.

[170] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:187–91. https://doi.org/10.1093/nar/gku365.

[171] Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656.

[172] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

[173] Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323. https://doi.org/10.1186/1471-2105-12-323.

[174] Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.

[175] Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis. Bioinformatics 2018. https://doi.org/10.1093/bioinformatics/bty890.

[176] Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi A J Integr Biol 2012;16:284–7. https://doi.org/10.1089/omi.2011.0118.

[177] Kehl T, Schneider L, Schmidt F, Stöckel D, Gerstner N, Backes C, et al. RegulatorTrail: A web service for the identification of key transcriptional regulators. Nucleic Acids Res 2017;45:W146–53. https://doi.org/10.1093/nar/gkx350.

[178] Yu G, He QY. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst 2016;12:477–9. https://doi.org/10.1039/c5mb00663e.

[179] Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 2012;13:R87. https://doi.org/10.1186/gb-2012-13-10-R87.

[180] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. Genome Res 2012;22:1760–74. https://doi.org/10.1101/gr.135350.111.

[181] Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 2013;9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

[182] Lun ATL, Smyth GK. Csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res 2015;44:e45. https://doi.org/10.1093/nar/gkv1191.

[183] Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009;26:139–40. https://doi.org/10.1093/bioinformatics/btp616.

[184] Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15. https://doi.org/10.1186/s13059-017-1382-0.

[185] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083–6. https://doi.org/10.1038/nmeth.4463.

[186] Vartak N, Guenther G, Joly F, DamLe-Vartak A, Wibbelt G, Fickel J, et al. Intravital dynamic and correlative imaging reveals diffusion-dominated canalicular and flow-augmented ductular bile flux. Hepatology 2020:hep.31422. https://doi.org/10.1002/hep.31422.

[187] Wang H, Quiroga AD, Lehner R. Analysis of lipid droplets in hepatocytes. Methods Cell Biol., vol. 116, Academic Press Inc.; 2013, p. 107–27. https://doi.org/10.1016/B978-0-12-408051-5.00007-3.

[188] Limaye PB, Alarcón G, Walls AL, Nalesnik MA, Michalopoulos GK, Demetris AJ, et al. Expression of specific hepatocyte and cholangiocyte transcription factors in human liver disease and embryonic development. Lab Investig 2008;88:865–72. https://doi.org/10.1038/labinvest.2008.56.

[189] Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jørgensen HF, et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol 2010;12:618–24. https://doi.org/10.1038/ncb2065.

[190] Vernay B, Koch M, Vaccarino F, Briscoe J, Simeone A, Kageyama R, et al. Otx2 regulates subtype specification and neurogenesis in the midbrain. J Neurosci 2005;25:4856–67. https://doi.org/10.1523/JNEUROSCI.5158-04.2005.

[191] Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, et al. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. Stem Cells 2012;30:910–22. https://doi.org/10.1002/stem.1070.

[192] Zhu Y, Li F, Guo GL. Tissue-specific function of farnesoid X receptor in liver and intestine. Pharmacol Res 2011;63:259–65. https://doi.org/10.1016/j.phrs.2010.12.018.

[193]   Hewitt NJ, Lechón MJG, Houston JB, Hallifax D, Brown HS, Maurel P, et al. Primary
        hepatocytes: Current understanding of the regulation of metabolic enzymes and
        transporter proteins, and pharmaceutical practice for the use of hepatocytes in
        metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. Drug
        Metab Rev 2007;39:159–234. https://doi.org/10.1080/03602530601093489.

[194]   Albrecht W, Kappenberg F, Brecklinghaus T, Stoeber R, Marchan R, Zhang M, et al.
        Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood
        concentrations. Arch Toxicol 2019;93:1609–37. https://doi.org/10.1007/s00204-019-
        02492-9.

[195]   Grinberg M, Stöber RM, Edlund K, Rempel E, Godoy P, Reif R, et al. Toxicogenomics
        directory of chemically exposed human hepatocytes. Arch Toxicol 2014;88:2261–87.
        https://doi.org/10.1007/s00204-014-1400-x.

[196]   Grinberg M, Stöber RM, Albrecht W, Edlund K, Schug M, Godoy P, et al. Toxicogenomics
        directory of rat hepatotoxicants in vivo and in cultivated hepatocytes. Arch Toxicol
        2018;92. https://doi.org/10.1007/s00204-018-2352-3.

[197]   Nussler A, Konig S, Ott M, Sokal E, Christ B, Thasler W, et al. Present status and
        perspectives of cell-based therapies for liver diseases. J Hepatol 2006;45:144–59.
        https://doi.org/10.1016/j.jhep.2006.04.002.

[198]   Aurich I, Mueller LP, Aurich H, Luetzkendorf J, Tisljar K, Dollinger MM, et al. Functional
        integration of hepatocytes derived from human mesenchymal stem cells into mouse
        livers. Gut 2007;56:405–15. https://doi.org/10.1136/gut.2005.090050.

[199]   Aurich H, Sgodda M, Kaltwaßer P, Vetter M, Weise A, Liehr T, et al. Hepatocyte
        differentiation of mesenchymal stem cells from human adipose tissue in vitro promotes
        hepatic integration in vivo. Gut 2009;58:570–81.
        https://doi.org/10.1136/gut.2008.154880.

[200]   Brulport M, Schormann W, Bauer A, Hermes M, Elsner C, Hammersen FJ, et al. Fate of
        extrahepatic human stem and precursor cells after transplantation into mouse livers.

Hepatology 2007;46:861–70. https://doi.org/10.1002/hep.21745.

[201] Lee FY, Lee H, Hubbert ML, Edwards PA, Zhang Y. FXR, a multipurpose nuclear receptor. Trends Biochem Sci 2006;31:572–80. https://doi.org/10.1016/j.tibs.2006.08.002.

[202] Wagner M, Fickert P, Zollner G, Fuchsbichler A, Silbert D, Tsybrovskyy O, et al. Role of farnesoid X receptor in determining hepatic ABC transporter expression and liver injury in bile duct-ligated mice. Gastroenterology 2003;125:825–38. https://doi.org/10.1016/S0016-5085(03)01068-0.

[203] Pereira-Fantini PM, Lapthorne S, Joyce SA, Dellios NL, Wilson G, Fouhy F, et al. Altered FXR signalling is associated with bile acid dysmetabolism in short bowel syndrome-associated liver disease. J Hepatol 2014;61:1115–25. https://doi.org/10.1016/j.jhep.2014.06.025.

[204] Kemper JK. Regulation of FXR transcriptional activity in health and disease: Emerging roles of FXR cofactors and post-translational modifications. Biochim Biophys Acta - Mol Basis Dis 2011;1812:842–50. https://doi.org/10.1016/j.bbadis.2010.11.011.

[205] Ali AH, Carey EJ, Lindor KD. Recent advances in the development of farnesoid X receptor agonists. Ann Transl Med 2015;3. https://doi.org/10.3978/j.issn.2305-5839.2014.12.06.

[206] Zhang S, Pan X, Jeong H. GW4064, an agonist of farnesoid x receptor, represses CYP3A4 expression in human hepatocytes by inducing small heterodimer partner expression. Drug Metab Dispos 2015;43:743–8. https://doi.org/10.1124/dmd.114.062836.

[207] Massafra V, van Mil SWC. Farnesoid X receptor: A "homeostat" for hepatic nutrient metabolism. Biochim Biophys Acta - Mol Basis Dis 2018;1864:45–59. https://doi.org/10.1016/j.bbadis.2017.10.003.

10.Acknowledgment