Epigenetic characterization of human hepatocyte subpopulations in context of complex metabolic diseases and during *in vitro* differentiation of hepatocyte-like cells

Dissertation

zur Erlangung des Grades des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Kathrin Kattler, M.Sc.

Saarbrücken

2019

Tag des Kolloquiums: 15.06.2020

Dekan: Prof. Dr. Guido Kickelbick

Berichterstatter: Prof. Dr. Jörn Walter

Prof. Dr. Hans-Peter Lenhof

Prof. Dr. Wolf Reik

Vorsitz: Prof. Dr. Alexandra Kiemer

Akad. Mitarbeiter: Dr. Jens Neunzig

"Yet what is any ocean, but a multitude of drops?"

— David Mitchell, Cloud Atlas

Acknowledgements

Summary

The comprehensive transcriptional and epigenetic characterization of human hepatocyte subpopulations is necessary to achieve a better understanding of regulatory processes in health and complex metabolic diseases as well as during *in vitro* differentiation. Based on integrative analysis of genome-wide sequencing data, this thesis aims to unravel hepatocyte heterogeneity in different biological contexts.

A deeper understanding of spatial organization of cells in human tissues is an important challenge. Using a unique experimental set-up based on laser capture microdissection coupled to next generation sequencing, which preserves spatial orientation and still provides genome-wide data of well defined subpopulations, the first combined spatial analysis of transcriptomes and methylomes across three micro-dissected zones of human liver provides a wealth of new positional insights, both in health and in context of fatty liver disease. In addition, these spatial maps serve as reference for projection of single cell data into hepatic pseudospace, which is still a major challenge. Hence, a novel pseudospace inference approach, which considerably improves spatial reconstruction of single cells into tissue context, is demonstrated for human liver. Finally, the identification of underlying regulatory networks by integrative epigenomic analysis of *in vitro* differentiated hepatocyte-like cells contributes to the development of reasonable cell culture interventions to improve differentiation.

Kurzfassung

Die umfassende transkriptionelle und epigenetische Charakterisierung humaner Leberzellsubpopulationen ist notwendig für die Aufklärung regulatorischer Prozesse in gesundem Gewebe, sowie im Zusammenhang mit komplexen metabolischen Erkrankungen und während der *in vitro* Differenzierung. Ziel dieser Arbeit ist es, basierend auf der integrativen Analyse genomweiter Sequenzierungsdaten, die Heterogenität von Leberzellen besser zu verstehen.

Die räumliche Organisation von Zellen in humanem Gewebe stellt eine große Herausforderung dar. Mit Hilfe von Lasermikrodissektion gekoppelt an Hochdurchsatzsequenzierung ist es möglich definierte Subpopulationen hinsichtlich ihres Gewebekontextes zu analysieren. Somit konnte die erste räumliche Analyse von Transkriptom und Methylom dreier Zonen der humanen Leber erstellt werden, die eine Vielzahl neuer Erkenntnisse sowohl in gesundem Lebergewebe als auch in Zusammenhang mit Fettlebererkrankungen liefert. Außerdem wurde auf Grundlage dieser räumlichen Karten ein neuer Ansatz zur Projektion von Einzelzelldaten in den räumlichen Gewebekontext etabliert. Schließlich konnte durch die integrative Analyse der ausschlaggebenden regulatorischen Netzwerke während der *in vitro* Differenzierung von Hepatozyten-ähnlichen Zellen neue Strategien zur Verbesserung der Differenzierung entwickelt werden.

Contents

1. Introduction

1.1 Epigenetics

Epigenetics is the study of changes in genomic regulation that can be mitotically and or meiotically heritable, but do not comprise alterations in the underlying nucleotide sequence (Feil and Fraga, 2012). Already in 1942, its eponym Conrad H. Waddington described this new field of research as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1968, 2012).

In mammals, key epigenetic mechanisms are DNA methylation and a set of posttranscriptional histone modifications, which, together with non-coding RNAs and regulatory proteins, shape the epigenetic landscape of each cell (Figure 1). The fact that multicellular organisms develop from a single omnipotent zygote is a striking example of the importance of epigenetic regulation. The zygote's progeny, which all share the same DNA sequence, have the potential to differentiate into a multitude of cell-types with defined gene expression profiles, resulting in highly specialized functions. The establishment of those specific programs, which is based on epigenetic memory, but also fast transcriptional changes as reaction to external stimuli, is only possible due to the intricate interplay of several layers of epigenetic regulation.
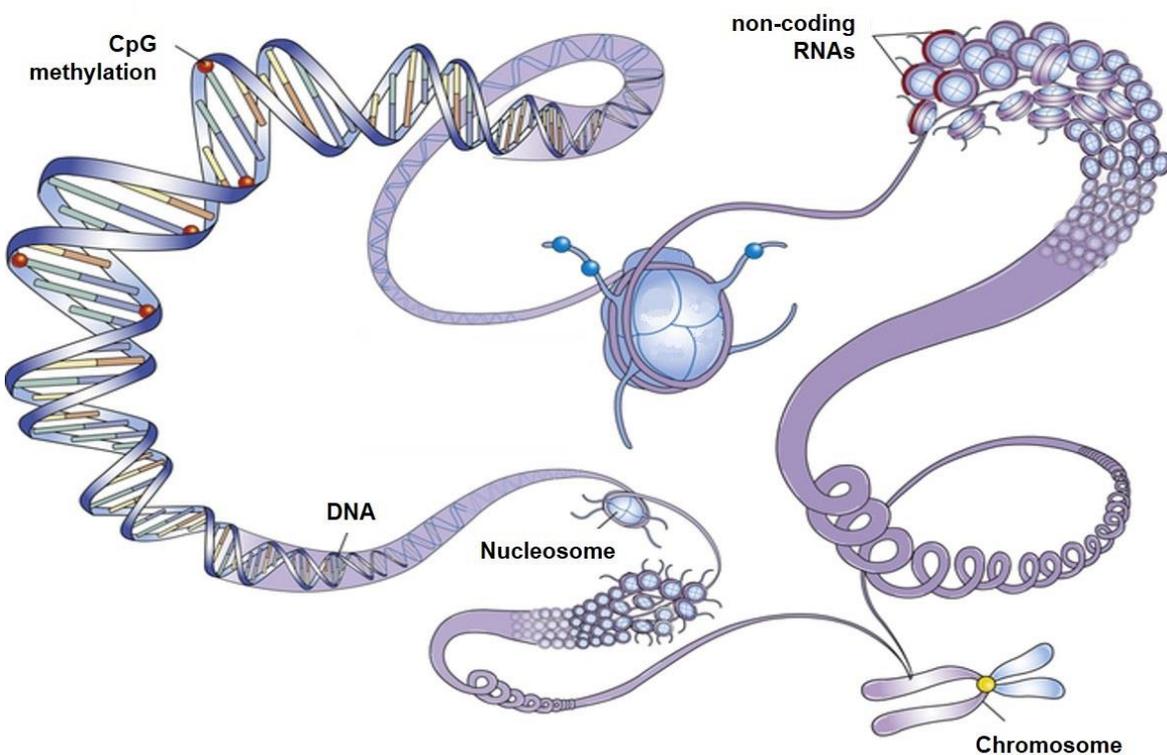
Figure 1: Epigenetic mechanisms and chromatin condensation (Lee et al., 2014, modified). In mammals, DNA methylation at CpG positions, histone modifications and non-coding RNAs are the major players of epigenetic regulation. For instance, they are involved in the intricate regulation of dynamic chromatin condensation, which is determined by the density of nucleosome complexes consisting out of DNA wrapped around histone proteins.

1.1.1 Chromatin structure and histone modifications

The DNA of a single diploid human cell consists of approximately 6 billion base pairs. In total, this would sum up to a length of 2 m, while the nucleus – the cellular compartment containing the DNA – is only 6 µm in diameter. Thus, in eukaryotic cells, the DNA forms a complex nucleoprotein structure called chromatin, which enables a strictly regulated dynamic DNA condensation.

The basic repeating unit of chromatin is the nucleosome, typically consisting out of 147 bp of DNA wrapped around a histone protein octamer (Kornberg and Thomas, 1974). There are five major histone families, which are highly conserved across species. The canonical core histones H2A, H2B, H3 and H4 form the nucleosome particles (Figure 2 A, Luger et al., 1997), while the linker histone H1 stabilizes naked DNA (Thoma and Koller, 1977) linking neighboring nucleosomes like beads on a string.

Histones serve not only as coils for DNA packaging, but – together with further epigenetic mechanisms – certain histone variants and histone modifications also exhibit regulatory potential by influencing the dynamic state of chromatin condensation (Henikoff et al., 2004; Jenuwein and Allis, 2001). The two major states of chromatin condensation are open euchromatin and closed heterochromatin. Euchromatic nucleosomes are loosely packaged, making the DNA accessible for regulatory proteins such as transcription factors or the transcriptional machinery. In contrast, in heterochromatin nucleosomes are tightly condensed, forming an inaccessible structure that for example impedes transcriptional activity or contributes to genomic stability by silencing of repetitive elements (Strålfors and Ekwall, 2011).

Canonical histone proteins are generally expressed during S-phase to complex newly synthesized DNA during replication (Marzluff et al., 2002). In contrast, the replacement of canonical histones by histone variants is a targeted process, supported by ATP-depending histone chaperones, which is replication independent (Talbert and Henikoff, 2017). Histone variants play a role in active transcription, heterochromatin formation as well as DNA repair. For example, H2A.Z in promoter regions is involved in transcriptional initiation (Fan et al., 2002; Guillemette et al., 2005; Hardy et al., 2009), while H2A.W promotes heterochromatin condensation (Yelagandula et al., 2014). Phosphorylated ɣ-H2A.X accumulates around DNA double strand breaks (Rogakou et al., 1998) and recruits members of the DNA repair machinery such as MDC1 (Stucki et al., 2005), a mediator of the DNA repair initiating MRN complex (Lamarche et al., 2010). Moreover, histone variants interact with posttranslational histone modifications (Li and Fang, 2015).

Presumably, the best described reversible histone modifications are lysine acetylation and successive mono-, di- or tri-methylation of lysines (K) and arginines (R) at distinct positions of the positively charged histone tails protruding the nucleosome complex (Figure 2 B). However, there are also histone core modifications such as H3K122ac that have regulatory potential (Pradeepa et al., 2016). Besides acetylation and methylation, further posttranslational histone modifications include ubiquitination of lysines, phosphorylation of serine, threonine and tyrosine (Jenuwein and Allis, 2001), and also more recently described serotonylation of glutamines (Farrelly et al., 2019). The regulatory potential of histone modifications is largely attributed to changing interaction affinities with chromatin associated regulatory proteins (Jenuwein and Allis, 2001). However, lysine acetylation may also be directly involved in the stabilization of euchromatin (Sterner and Berger, 2000), as acetyl groups contribute to the partial neutralization of positively charged histone tails. The resulting reduction of charge-based interactions

between neighboring nucleosomes could contribute to the formation and maintenance of a more accessible chromatin structure (Fletcher and Hansen, 1996).

The so called *histone code* can be used to segment the genome into regions by their potential function (Biesinger et al., 2013), as certain histone modification patterns correlate with genomic states. For example, H3K4me3, accompanied by several further active histone modifications such as H3K4me1 and H3K27ac, is a strong indicator for promoter regions (Koch et al., 2007; Santos-Rosa et al., 2002) . In absence of H3K4me3, enrichment of H3K27ac (Creyghton et al., 2010) and or H3K122ac indicates enhancer activity (Pradeepa et al., 2016). H3K36me3 covers the bodies of transcribed genes and is therefore closely linked to active transcription (Kolasinska-Zwierz et al., 2009). H3K27me3 is a repressive mark that can either be involved in local polycomb repressed gene silencing or span large heterochromatic domains together with H3K9me3 (Pauler et al., 2009).

Interestingly, there are genomic regions enriched for both activating H3K4me3 and repressive H3K27me3. These regions, which are referred to as bivalent domains (Bernstein et al., 2006; Kinkley et al., 2016; Young et al., 2011), are probably involved in the silencing of strictly regulated genes such as developmental genes that need to be transcriptionally activated at precise time-points. Thus, the bivalent character probably results in gene silencing, while still keeping the gene poised for transcriptional activation (Bernstein et al., 2006).
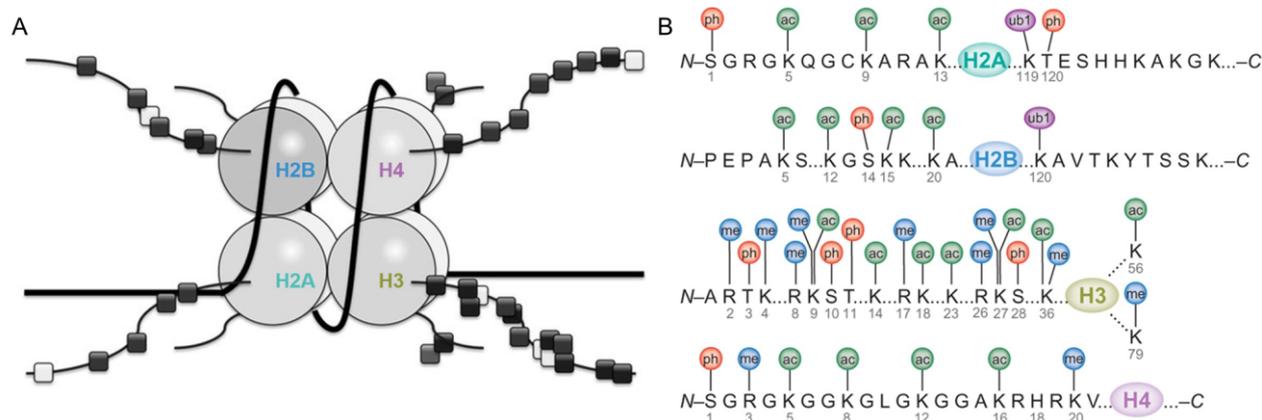
Figure 2: Histone modifications. A – Nucleosomes consist out of approximately 146 bp DNA wrapped around 4 histone protein dimers (Kato et al., 2015, modified). B – The most common modifications of histones H2A, H2B, H3 and H4 with regulatory implications are acetylation, methylation, phosphorylation and ubiquitination at specific amino acids of the nucleosome core protruding histone tails (Bhaumik et al., 2007).

Beside the nucleosome-based level of condensation, chromatin also forms three dimensional (3D) higher-order structures, with metaphasis chromosomes representing the highest degree of compaction. Although, to date, the organization of those structures is still not well characterized, it is evident that chromosomes do not fold randomly into the nucleus. During recent years it was shown that the overall nuclear organization, including long range interactions across and between chromosomes or the association of chromatin with the nuclear lamina, are of regulatory importance. For instance, long-range interactions between distant regulatory elements suggest an essential impact of spatial genome compartmentalization on the regulation of transcription and replication (Pombo and Dillon, 2015) . Moreover, lamina associated domains (LADs), which are composed of H3K9me3 enriched chromatin, typically spanning about 0.5 Mb, interact with the

nuclear lamina and correlate with low transcriptional activity and late replication timing (Guelen et al., 2008; Pope et al., 2014). More recently, the development of chromosome conformation capture techniques such as Hi-C (Lieberman-Aiden et al., 2009) entailed the detection of topologically associated domains (TADs, Dixon et al., 2012; Sexton et al., 2012). These are large genomic regions that exhibit preferential long range interactions within themselves. The disruption or formation of additional TAD boundaries can result in changes of gene expression patterns and the development of diseases such as cancer (Lupiáñez et al., 2016; Taberlay et al., 2016). Thus, TADs are considered as fundamental structural units, which, for example, regulate distant promoter enhancer interactions and represent stable replication domains (Pope et al., 2014). However, precise mechanisms are still under investigation.

1.1.2 DNA methylation

DNA methylation was first discovered in 1948 by Rollin Hotchkiss. Since then, it has become the most studied epigenetic modification in mammalian cells and many fundamental functions, such as in regulation of gene silencing (Holliday and Pugh, 1975), genomic imprinting (Li et al., 1993), X chromosome inactivation (Heard et al., 1997; Riggs, 1975) and genome stability (Yoder et al., 1997), have been described. Furthermore, DNA methylation, representing a kind of epigenetic memory, plays a crucial role during early embryogenesis and cellular differentiation, but is also implicated in the development and progression of many diseases including cancer (Portela and Esteller, 2010).

In mammals, DNA methylation occurs almost exclusively at the C5 position of cytosines (Hotchkiss, 1948), although more recently N6-methyladenine was detected in mouse embryonic stem cells (Wu et al., 2016). Cytosine methylation is catalyzed by different classes of DNA methyltransferases (DNMTs) that transfer a methyl-group from their cofactor S-adenyl methionine (SAM), resulting in covalently linked 5-methyl-cytosine (5mC, Figure 3 A). Maintenance methyltransferase DNMT1, which preferentially methylates hemi-methylated DNA, is responsible for re-establishment of DNA methylation patterns on the newly synthesized DNA strand after replication (Hermann et al., 2004). In contrast, the *de novo* methyltransferases DNTM3A and DNTM3B can methylate unmethylated as well as hemi-methylated DNA, allowing the introduction of newly methylated CpG sites (Jurkowska et al., 2011; Okano et al., 1999).

Regarding DNA demethylation two different mechanisms have been described. Passive loss of methylation takes place when DNA methylation patterns are not maintained by DNMT1 after replication, thus it requires cell divisions (Howlett and Reik, 1991). Active DNA demethylation, which can also occur in non-dividing cells (Mayer et al., 2000; Oswald et al., 2000), is a more complex process. Mediated by ten-eleven translocation family (TET) enzymes, 5mC can be oxidized to 5-hydroxy-cytosine (5hmC), 5-formyl-cytosine (5fC) and 5-carboxy-cytosine (5cC) in a iterative manner (Ito, 2011; Tahiliani et al., 2009). These modifications might themselves have regulatory functions (Bachman et al., 2015), but are also essential for active DNA demethylation (Figure 3 A). For instance, both 5fC and 5cC can be recognized and removed by the thymine DNA glycoxylase (TDG), which mediates base excision repair (BER), entailing the incorporation of unmodified cytosine (Dalton and Bellacosa, 2012).

In the genome cytosine methylation is primarily found in a CpG dinucleotide context (Doskočil and Šorm, 1962). The palindromic sequence facilitates the maintenance of symmetric DNA methylation patterns over cell divisions (Li and Zhang, 2014). Non-CpG methylation is mostly restricted to certain cell-types such as

pluripotent stem cells and neurons (Laurent et al., 2010; Lister et al., 2013). While CpG dinucleotides are underrepresented throughout the genome, they are enriched in so called CpG islands, mostly at highly regulatory promoter regions (Ioshikhes and Zhang, 2000). CpG islands upstream of transcriptionally active genes are generally unmethylated (Figure 3, A), as high DNA methylation in those regions results in gene silencing of the corresponding gene (Lister et al., 2009). Outside of CpG islands, most CpGs across the genome are methylated (Ehrlich et al., 1982), thereby contributing to genome stability (Meng et al., 2015), for instance by silencing of transposable and repetitive elements (Figure 3, A).
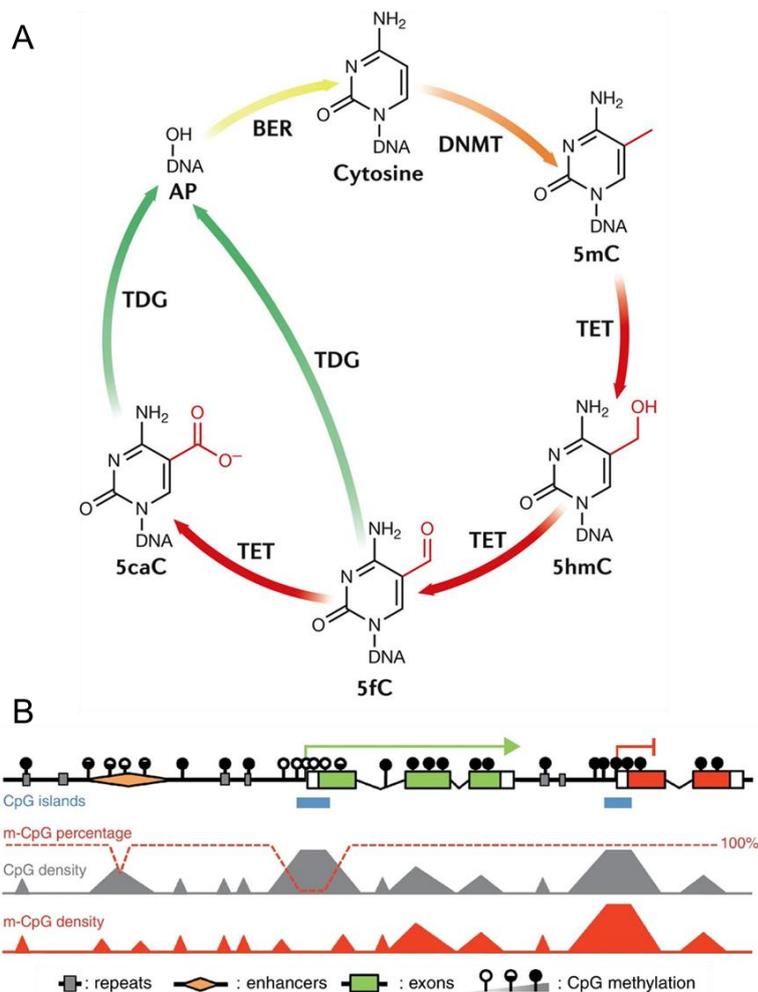
Figure 3: DNA methylation. A – Methylation of Cytosines at the 5' carbon is catalyzed by DNMTs. 5mC can be further oxidized to hydroxy-, formyl- or carboxy-Cytosine by TET enzymes. Active Demethylation can occur by TDG followed by base excision repair (Raiber et al., 2017) B – CpGs are underrepresented in the mammalian genome, but cluster at certain regulatory regions (grey). While most CpGs are methylated to retain genomic stability, promoter CpG islands are mostly unmethylated, and further regions with regulatory potential such as enhancers can have variable CpG methylation patterns (red) depending on their current regulatory state (Baubec and Schübeler, 2014).

For a long time DNA methylation and gene expression were considered as purely anti-correlated (Razin and Cedar, 1991), with hypomethylation promoting active transcription. Regarding promoter regions and in particular CpG islands, this dogma certainly holds true, but in other regulatory elements such as enhancers, insulators or along the gene body, the situation appears to be much more complex. For example, DNA methylation is often enriched in exons of transcribed genes and apparently has an effect on alternative

splicing (Shayevitch et al., 2018) or inhibits unwanted spurious initiation of transcription (Neri et al., 2017). DNA methylation in interaction with histone modifications influences chromatin condensation and can recruit regulatory proteins as further described in chapter 1.1.4.

1.1.3 Non-coding RNAs

The full transcriptome of a cell does not only contain protein coding messenger-RNA (mRNA), but also different classes of non-coding RNAs with diverse functions. The total amount of non-coding RNAs in the genome is still unknown and many newly described classes have still to be assessed regarding their functions. The most abundant RNAs are ribosomal RNA (rRNA) and transfer RNAs (tRNA), which are crucial factors for mRNA translation into proteins (Blanco and Blanco, 2017). Moreover, small non-coding RNAs of 20 to 30 nucleotides are involved in multiple regulatory pathways. In particular, micro-RNAs (miRNA), short interfering RNAs (siRNA) and PIWI interacting RNAs (piRNA) play a fundamental role in transcriptional and posttranscriptional RNA silencing, which basically relies on complementary binding to the targeted mRNA (O'Brien et al., 2018; Siomi et al., 2011; Tijsterman et al., 2002). There are further small non-coding RNAs such as small nucleolar RNAs (snoRNA) that guide modifying enzymes to other RNA classes (Dupuis-Sandoval et al., 2015) or small nuclear RNAs (snRNAs) involved in pre-processing of nuclear pre-mRNA (Guiro and Murphy, 2017). In most cases the functionality of long non-coding RNAs (lncRNA), which are typically longer than 200 nucleotides and frequently show features of processed mRNA like 5' cap or polyadenylation, is still not well understood (Uszczynska-Ratajczak et al., 2018). Functional studies imply a diverse set of regulatory processes including transcriptional and post-transcriptional regulation as well as chromatin remodeling. For example, the lncRNA Xist induces extensive changes of the chromatin structure in mammalian female cells, ultimately leading to X chromosome inactivation (Morey et al., 2004). However, the full regulatory potential of lncRNAs is to date only starting to become apparent.

1.1.4 Regulatory interplay of epigenetic modifications

Cellular mechanisms, such as active transcription, are orchestrated by a complex interplay of multiple layers of regulation including DNA methylation, histone modifications and non-coding RNAs, together with regulatory proteins such as chromatin remodeling factors, transcription factors and structural proteins. The patterns of different epigenetic marks, especially at certain putative regulatory elements, already suggest a crosstalk between the different layers of epigenetic control (Figure 4). Frequently, CpG island harbouring promoter regions of active genes are unmethylated and nucleosome-depleted (Figure 5), while flanking nucleosomes are marked by H3K4me3, H3K27ac and a more wide-spread H3K4me1 (Figure 4, green). This pattern typically corresponds to an accessible chromatin structure, which allows the binding of transcription factors as well as of the transcriptional machinery (Fuda et al., 2009). Gene bodies of actively transcribed genes are marked by H3K36me3 and CpG methylation, preventing aberrant transcription and apparently influencing alternative splicing (Jones, 2012; Teissandier and Bourc'his, 2017). Furthermore, potentially active enhancer regions (Figure 4, blue) are characterized by a strong enrichment of H3K27ac, H3K4me1 and H3K122ac, but a notable depletion of H3K4me3. While being rather CpG-poor, enhancers

display diverse DNA methylation patterns (Figure 5, Stadler et al., 2011), which might be due to high DNA methylation dynamics mediated by TET enzymes (Wang et al., 2018).
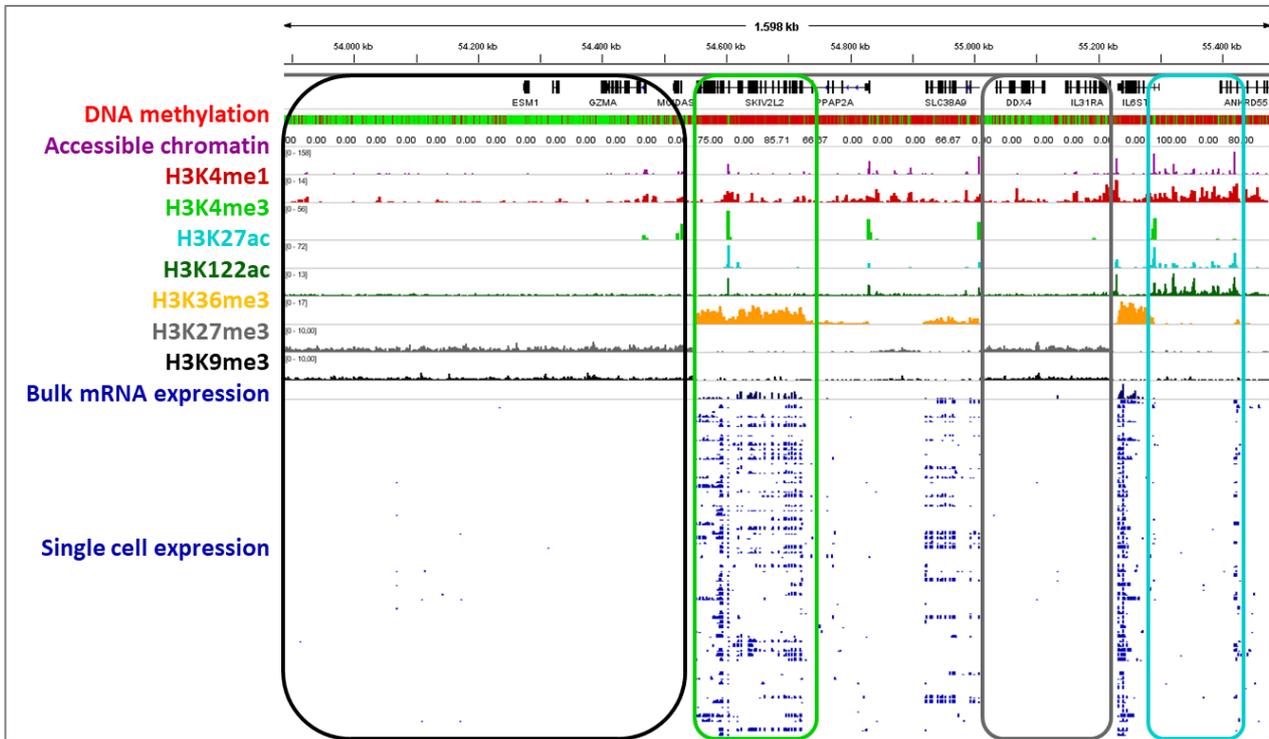
Figure 4: Multiple layers of NGS based epigenomic and transcriptional data in the hepatocyte-like cell line HepG2. Exemplary genome browser view of data generated within the DEEP project (Behjati Ardakani et al., 2018; Kattler, 2015; Kattler, 2013; Salhab et al., 2018). The first track displays whole genome bisulfite sequencing derives DNA methylation with unmethylated CpG positions in green and methylated CpGs in red. Accessible chromatin assessed by ATAC-seq is shown in purple. A set of 7 histone modifications (H3K4me1 in red, H3K4me3 in green, H3K27ac in light blue, H3K122ac in dark green, H3K36me3 in orange, H3K27me3 in grey and H3K9me3 in black) was analyzed by ChIP-seq. Bulk mRNA transcription data as well as single cell expression data of 64 HepG2 cells generated by SmartSeq2 are displayed in blue. The 1,598 kb window covers a large heterochromatic region (black), an active bidirectional promoter (green), silenced genes (grey) and a potential enhancer (blue).

In contrast, promoters of silenced genes are highly methylated or accumulate nucleosomes with repressive histone modifications, in particular H3K27me3, resulting in a facultative heterochromatin structure (Figure 4, grey). Interestingly, it was reported for several cell-types that CpG methylation and H3K27me3 deposition appear to be mutually exclusive, as high DNA methylation levels seem to prevent the binding of polycomb repressive complex 2, which is associated with histone methyltransferases like EzH that mediates H3K27 trimethylation (Reddington et al., 2013). Gene-poor constitutive heterochromatin is generally characterized by high CpG methylation, enrichment of H3K9me3 and absence of any activating histone marks. Figure 4 displays genome-wide data generated from the liver cancer derived cell line HepG2. Here, hypomethylation of large heterochromatic regions can be observed (Figure 4, black). This immortalization-related loss of DNA methylation is presumably compensated by enrichment of H3K9me3, together with polycomb-mediated H3K27me3 – a feasible alternative mechanism that has been reported for several specific circumstances,

such as the perturbation of the original epigenetic mechanism (Saksouk et al., 2015), and nicely illustrates the interoperability between epigenetic layers.
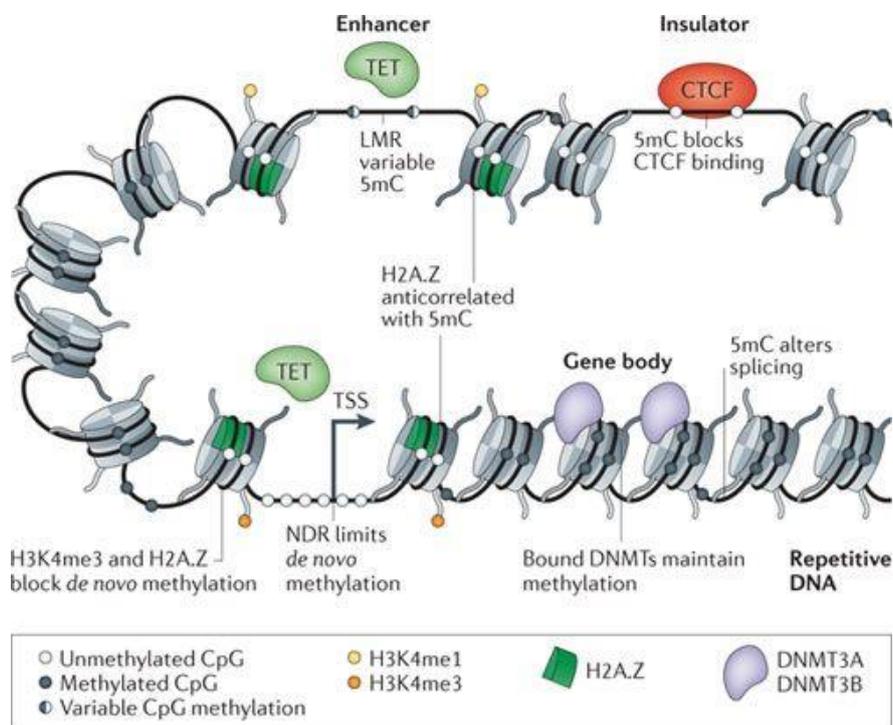
Figure 5: Transcription is regulated by a complex interplay of epigenetic mechanisms (Jones, 2012). Promoters of active genes are generally accessible nucleosome-depleted regions (NDR) with unmethylated CpG sites. Flanking nucleosomes are frequently marked by the activating histone modifications H3K4me1 (yellow) and H3K4me3 (orange), together with the histone variant H2A.Z (green), which is antagonistic to DNA methylation and DNMT binding. DNA methylation patterns of gene bodies can influence alternative splicing events. Enhancers, which can loop back to the promoter region over short or long distances, often show variable DNA methylation patterns. DNA methylation at insulators can prevent CTCF binding which may also impact gene expression. The figure provides a simplified overview of transcriptional control, as it does not include further factors such as non-coding RNAs, transcription factor binding and further regulatory proteins.

The establishment of these epigenetic patterns is a highly complex process. The binding of many regulatory proteins, including DNA sequence motive-specific transcription factors, frequently relies on the state of chromatin condensation. Thus, most transcription factors bind only in accessible chromatin regions (Grossman et al., 2018). However, epigenetic modifications can also themselves recruit regulatory proteins. For instance, there are transcription factors, such as homeodomain proteins, that selectively bind to methylated CpGs (Yin et al., 2017). Moreover, CpG methylation can recruit DNA methylation binding proteins such as MBD1, which is again associated with chromatin modifying complexes like Suv39H1, responsible for the deposition of H3K9me3 (Fujita et al., 2003; Ng et al., 2000). In turn, H3K9me3 binds URHF1, which is involved in the regulation of DNMT1 (Liu et al., 2013) and also interacts with H3K9 specific histone methyltransferase G9a (Kim et al., 2009) as well as with several histone deacetylases (Ghoshal et al., 2002), thereby constituting a kind of self-enhancing feedback loop, resulting in spreading of heterochromatin (Bronner et al., 2013). Moreover, chromo-domain containing HP1, which also interacts with

G9a (Ruan et al., 2012), binds to H3K9me2 and H3K9me3, further stabilizing heterochromatic structures (Eissenberg and Elgin, 2014).

On the other hand, bromo-domain containing trithorax group complexes are associated with histone acetyltransferases (HATs) and H3K4 specific histone methyltransferases (Tie et al., 2014), thus promoting accessible chromatin (Schuettengruber et al., 2007). HATs are also often associated with transcription activator complexes (Ogryzko et al., 1996). In this context, nucleosome depletion, H3K4me3 and histone variant H2A.Z are reported to prevent *de novo* methylation at active regulatory elements (Jones, 2012).

The state of chromatin condensation can also be directly altered by chromatin remodeling complexes, which are recruited to specific sites by small regulatory RNAs (Holoch and Moazed, 2015). For example, SWI/SNF family complexes can disrupt DNA-histone contacts, leading to nucleosome shifting (Bouazoune et al., 2009). Nucleosomes can also be exchanged directly by histone chaperones, for example to allow the replication-independent incorporation of histone variants (Venkatesh and Workman, 2015).

The spatial organization within the nucleus represents an additional layer of regulation. Higher order chromatin structures promote a nuclear compartmentalization. For instance, association of chromatin to the nuclear lamina results in heterochromatin formation (Finlan et al., 2008). TADs are suggested to contribute to the regulation of gene expression by guiding enhancers to their target genes (Andrey and Mundlos, 2017). The looping between enhancers and promoters, which can be disrupted by insulators, is a crucial element of transcriptional regulation (Figure 5). In turn, methylation of CpG sites in insulators prevents the binding of CTCF (Figure 5), which itself exhibits multiple functions including insulator activity and transcriptional control, but presumably also the regulation of long range chromatin interactions, as it is enriched at TAD and LAD boundaries (Phillips and Corces, 2009; Wutz et al., 2017).

Together, the mechanisms described above form an intricately balanced network, which is in its extent not yet fully understood. The examples outlined in this chapter, however, illustrate the vast complexity of epigenetic regulation and transcriptional control.

1.2 Methods for genome-wide transcriptional and epigenetic profiling

Recently, enormous progress has been made in the mapping of epigenomes, especially by large sequencing consortia such as the international human epigenome consortium (IHEC) including the German contribution DEEP. This was only feasible by the advent of next generation sequencing (NGS) and the plethora of NGS based methods for genome-wide transcriptional and epigenetic profiling that have been published since. Although long read sequencing techniques are advancing (Pollard et al., 2018), the currently most popular sequencing approach is still sequencing by synthesis of short sequences (Ambardar et al., 2016). It allows high-throughput sequencing of 50 to 200 bp reads with very high accuracy. Thus, it is possible to assess transcriptomes and several layers of epigenetic modifications for whole tissues or isolated cell populations, and some methods can even be scaled to single cell level. Several genome-wide methods relevant for this thesis are described in the following chapter.

1.2.1 Transcriptome analysis

Genome-wide transcriptome analysis relies on the reverse transcription of RNA and sequencing of the resulting cDNA, as direct sequencing of RNA remains technically challenging (Marinov, 2017). For many studies the main interest lies in differential expression analysis of protein coding genes. Thus, the most common approach is based on capturing polyadenylated RNA fragments prior to cDNA preparation. If sequencing of total RNA is desired, the high level of ribosomal RNA should be taken into account. To this end, so called ribo-depletion protocols can be employed which exclude non-informative ribosomal RNA from sequencing library preparation (Sooknanan et al., 2010; Zhulidov, 2004).
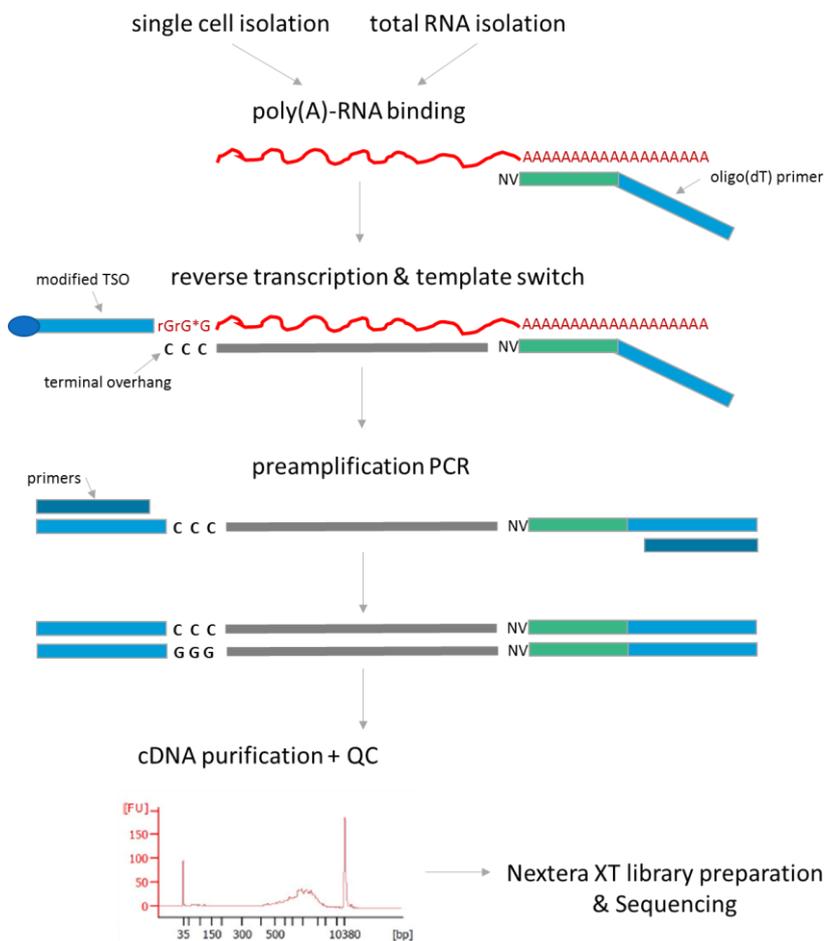
Figure 6: Smartseq2 workflow for RNA-seq of isolated RNA or single cells (modified after Picelli et al., 2013). After single cell isolation respective total RNA isolation, polyadenylated RNA is reverse transcribed using a reverse transcriptase which adds a CCC overhang to the first strand 3' end. This overhang allows the binding of a template switch oligonucleotide (TSO), ensuring full-length coverage during second strand synthesis. cDNA is preamplified by PCR, purified and controlled for degradation. A typical mammalian cDNA yields mainly fragments around 2 kb. Finally, sequencing libraries can be prepared from cDNA using Nextera kits (Illumina).

To date, many different RNA-seq protocols have been developed. One of the most cited protocols is the Smartseq2 protocol (Picelli et al., 2013), which was originally published for single cell mRNA-seq. As it yields high quality full length transcriptomes and library preparation is affordable and straightforward, it became increasingly popular for bulk mRNA-seq as well. The Smartseq2 workflow (Figure 6) starts either with

isolated single cells in a triton based lysis buffer or with purified total RNA. Polyadenylated RNA is reverse transcribed using poly-dT primers and a reverse transcriptase that produces a CCC overhang at the 3' end of the first strand. This allows a template switch, ensuring full length transcript coverage in the second strand synthesis. cDNA is preamplified, purified, and can be quantified and controlled for degradation on an Agilent Bioanalyzer. A typical high quality mammalian cDNA fragment size distribution (Figure 6) shows mainly fragments around 2 kb. Tagmentation based library preparation can then be easily achieved by commercially available kits such as Illumina's Nextera kits. After sequencing, reads are aligned to a gene annotated reference and transformed to count-based estimations of gene-wise transcription levels.

1.2.2 DNA methylation analysis

Whole genome bisulfite sequencing (WGBS) is currently considered as the gold standard for genome-wide DNA methylation analysis. Bisulfite treatment of DNA results in conversion of unmethylated cytosines into uracil, due to sulfonation, followed by hydrolytic deamination and final desulfonation (Frommer et al., 1992). In subsequent PCR, uracil is paired with adenine, ultimately leading to an exchange of unmethylated cytosine by thymine in the analyzed sequences. In contrast, methylated cytosines are protected from conversion because of their methyl-group. Thus, sequencing of bisulfite treated DNA and comparing the obtained sequence with the genomic reference sequence allow the efficient detection of methylated CpGs at base resolution.

However, covering the whole genome with sufficient sequencing depth requires high amounts of reads per sample, which is immensely increasing costs and limits the number of affordable replicates. Moreover, for many projects DNA methylation analysis focuses mainly on potential regulatory regions, representing only a small proportion of the genome. Thus, reduced representation bisulfite sequencing (RRBS) can be a reasonable alternative. For RRBS genomic DNA is digested using CpG methylation insensitive restriction enzymes such as MspI, HaeIII or AluI, that cut at distinct sequence motives accumulated in CpG-rich regions. In the subsequent adapter ligation based library preparation (Figure 7) those regions are enriched. Consequently, the number of necessary reads per sample is significantly decreased, allowing higher CpG coverage and more samples, respective replicates, for a given project.

The choice of the restriction enzyme decides about the predominantly covered regions. In the human genome the classical MspI restriction yields about 1 million CpGs, mostly located in CpG islands (Gu et al., 2011). If interested in further regulatory regions, HaeIII is also a valid choice. HaeIII RRBS yields roughly 4 to 5 million CpGs, covering promoter regions but also gene bodies. For even broader coverage a mixture of different restriction enzymes is advisable. For example, the combination of HaeIII and AluI yields around 10 million CpGs covering about 35 % of all CpGs in the human genome. Those CpGs are enriched in regions of interest such as promoters, gene bodies and distal regulatory elements. Moreover, the rather broad CpG coverage of HaeIII or HaeIII/AluI RRBS allows also more complex analysis such as the assessment of partially methylated domains (Salhab et al., 2018).

Initial RRBS protocols required up to 300 ng of genomic DNA for library preparation (Gu et al., 2011). More recent developments allow the downscaling to low amounts of DNA (Schillebeeckx et al., 2013) and even to single cell level (Guo et al., 2013). However, single cell RRBS provides rather sparse data, covering 200,000 to 1 million CpGs per single cell, and requires strong bioinformatical computation to develop

meaningful biological conclusions. Thus, RRBS of specific, carefully isolated cell populations, for example by fluorescence-activated cell sorting or laser capture microdissection, is to date a more reasonable way to answer many biological questions. In summary, RRBS represents a versatile method for the comprehensive study of genome-wide DNA methylation in many samples.
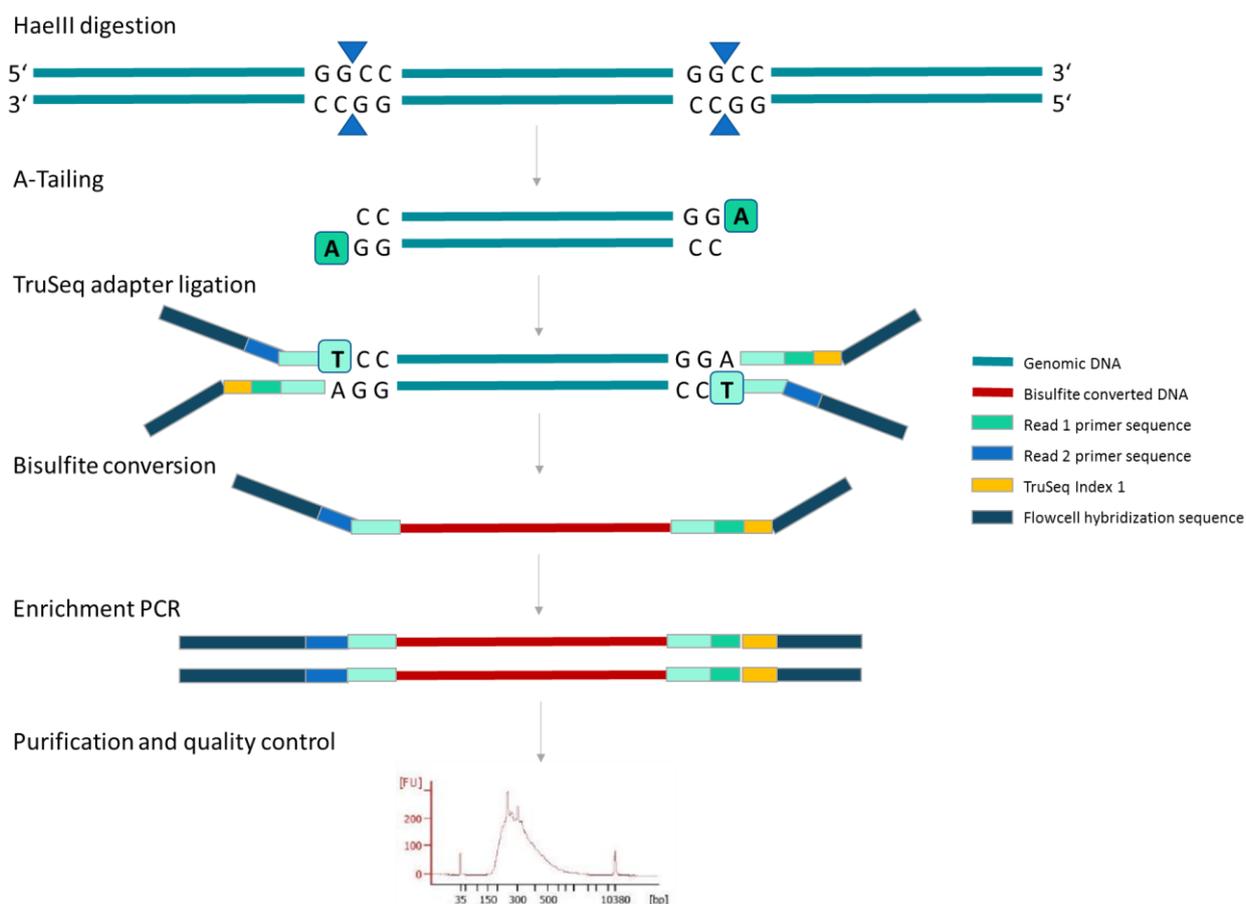
Figure 7: Reduced Representation Bisulfite Sequencing workflow. Genomic DNA is digested by a CpG methylation insensitive restriction enzyme such as HaeIII. After A-tailing, ligation of methylated TruSeq adapters and bisulfite conversion DNA fragments are enriched by PCR, resulting in a sequencing library with characteristic fragment size distribution.

1.2.3 Chromatin accessibility analysis

There are several methods for the genome-wide analysis of chromatin accessibility. DNaseI-seq could be considered as the classical method, as it was widely used and is also part of full epigenome analysis, as outlined by large epigenome consortia such as ENCODE or IHEC. A more recent method, which is becoming increasingly popular, is ATAC-seq, the assay for transposase-accessible chromatin analysis using sequencing (Buenrostro et al., 2013). Intact native chromatin is exposed to a Tn5 transposase pre-loaded with adapters compatible with Illumina next generation sequencing. The transposon cuts and simultaneously integrates the sequencing adapters into accessible DNA sites (Figure 8). Thus, after gap repair, accessible fragments are enriched by PCR, resulting in a sequencing library of open chromatin regions.
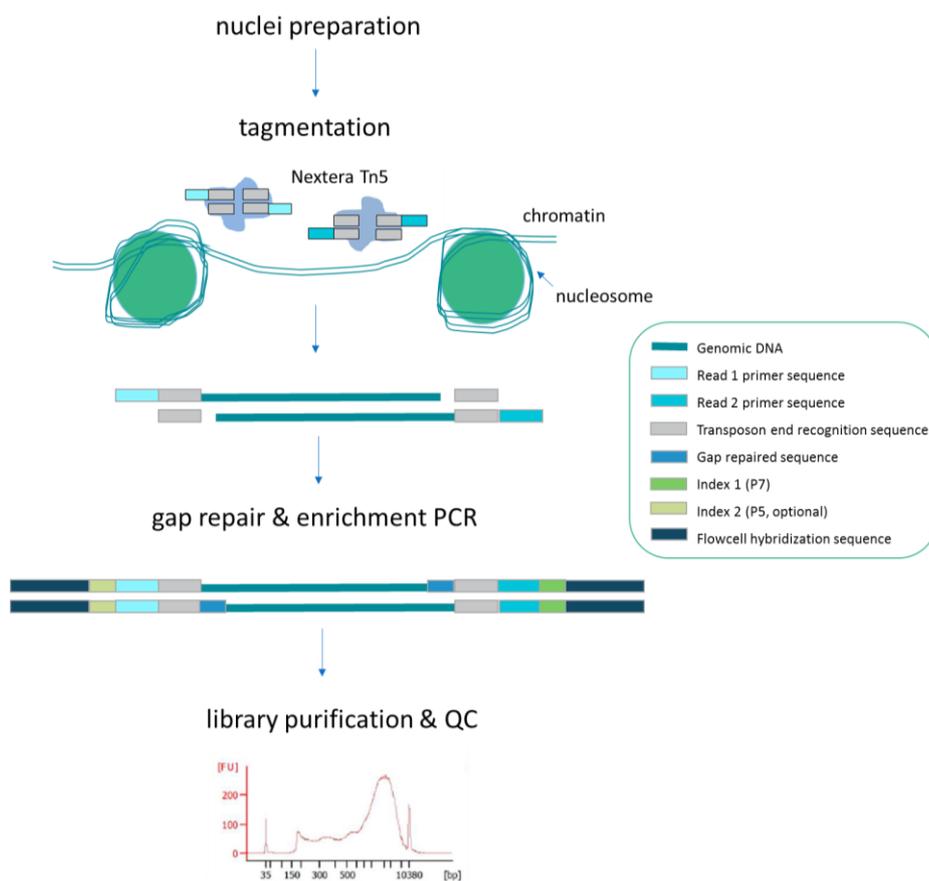
Figure 8: Chromatin accessibility analysis using ATAC-seq (modified after Buenrostro et al., 2013). Isolated nuclei are treated with the Nextera Tn5 transposase (Illumina), which simultaneously fragments and inserts Nextera sequencing adapters into the DNA of accessible chromatin regions. After gap repair, DNA fragments with adapters at 3' and 5' ends are enriched by PCR. The purified sequencing library yields a characteristic fragment size pattern that recapitulates nucleosome phasing.

ATAC-seq offers a fast, straightforward workflow and, in contrast to DNaseI-seq, which requires enzyme titration, transposon concentrations are largely cell-type independent (Buenrostro et al., 2013). An important possible pitfall is, however, the isolation of clean nuclei with intact chromatin, which can be challenging from certain cell-types or tissues. Nuclei that are still attached to cytoplasm often result in an extensive number of reads mapping to the mitochondrial genome, as the transposon preferentially tagments the free mitochondrial DNA (Corces et al., 2017). Since this increases sequencing costs to obtain sufficient genomic coverage, nuclei should be as clean as possible. Relating to chromatin integrity, especially frozen samples represent a major challenge. Degraded chromatin leads to noisy data that are hard to interpret. In some cases this can be overcome by specialized protocols including a very gentle nuclei preparation (Corces et al., 2017). Thus, the proper establishment of nuclei isolation is of utmost importance when starting a new project (Kattler, Master thesis, 2015).

Additionally, ATAC-seq can also be applied on single cells (Buenrostro et al., 2015; Cusanovich et al., 2015). However, single cell ATAC-seq yields extremely sparse data and due to the enrichment based nature of this method, the pure absence of a signal is not conclusive. Thus, the comprehensive analysis of single cell chromatin accessibility data represents an enormous bioinformatical challenge, which is currently mostly overcome by single cell aggregation and machine learning based data imputation (Pott and Lieb, 2015).

1.3 Liver and metabolic diseases

1.3.1 Metabolic liver zonation

Being the largest internal organ, the mammalian liver exhibits a multitude of essential biological functions including regulation of glucose metabolism, bile acid synthesis, fatty acid metabolism and detoxification. Some liver functions, such as gluconeogenesis and glycolysis, are contradictory to each other, which implies that specialized sets of hepatocytes might be involved in the respective processes to ensure their efficiency. The liver consists of four lobes, which are further structured into hexagonal-like hepatic lobules (Figure 9 A). At the borders of each lobule reside the portal triades (Figure 9 B), while the central vein is localized in the lobule's center (Birchmeier, 2016; Rappaport, 1977). The portal triade comprises hepatic artery, portal vein and bile ducts. Oxygen-rich blood flows from the portal triad through the sinusoid capillaries towards the central vein, thereby creating a porto-central oxygen gradient (Kietzmann, 2017).

In human, hepatocytes – the parenchymal liver cells – sit in rows of 15 to 25 cells (McEnerney et al., 2017) along the porto-central axis. Besides hepatocytes, liver tissue consists of a diverse set of non-parenchymal cells (NPCs, Figure 9 B). Cholangiocytes are specialized epithelial cells that frame the bile ducts and modify bile acids produced by hepatocytes (Tabibian et al., 2013). Discontinuous endothelial cells form the sinusoids between portal field and central vein, which harbor liver-resident macrophages, called Kupffer cells (Naito et al., 1997). More recently, several populations of innate lymphocytes such as liver resident NK cells (Peng and Sun, 2017) have been reported as well. Hepatic stellate cells reside in the space of Disse, which is located between sinusoids and hepatocytes. Apparently, they remain in a quiescent state during normal liver homeostasis, but, for instance, play a major role in extracellular matrix deposition during hepatic fibrosis (Eng and Friedman, 2000) as well as in the regulation of liver immunology (Weiskirchen and Tacke, 2014).

Although morphologically similar, hepatocytes are heterogeneous in respect to their metabolic enzymatic activity (Figure 9 B), depending on their localization within the hepatic lobule – a long known phenomenon termed as metabolic zonation (Jungermann and Sasse, 1978). Although metabolic pathways are subject to dynamic changes, due to blood contained factors including nutrients and hormonal stimuli, general gene expression profiles along the porto-central axis can be considered as rather stable zonation patterns (Gebhardt, 2014). During recent years, it was shown in multiple rodent studies that zonation is maintained by short-distanced morphogen gradients. In particular, pericentral Wnt-/β-catenin signaling, which is apparently established by the central vein's endothelial cells (Preziosi et al., 2018), represents a master regulator of metabolic zonation profiles (Gebhardt, 2014). Moreover, the involvement of further signaling pathways such as fibroblast growth factor (FGF), transforming growth factor (TGF) beta, hepatocyte growth factor (HGF), Hedgehog and Notch signaling has been described (Gebhardt, 2014), indicating a complex crosstalk between different morphogen gradients. In addition, the porto-central oxygen gradient, which correlates with the number of mitochondria and oxidative properties, might exhibit modulatory functions, for instance by pericentral stimulation of hypoxia-inducible transcription factors (HIFs) (Kietzmann, 2017).
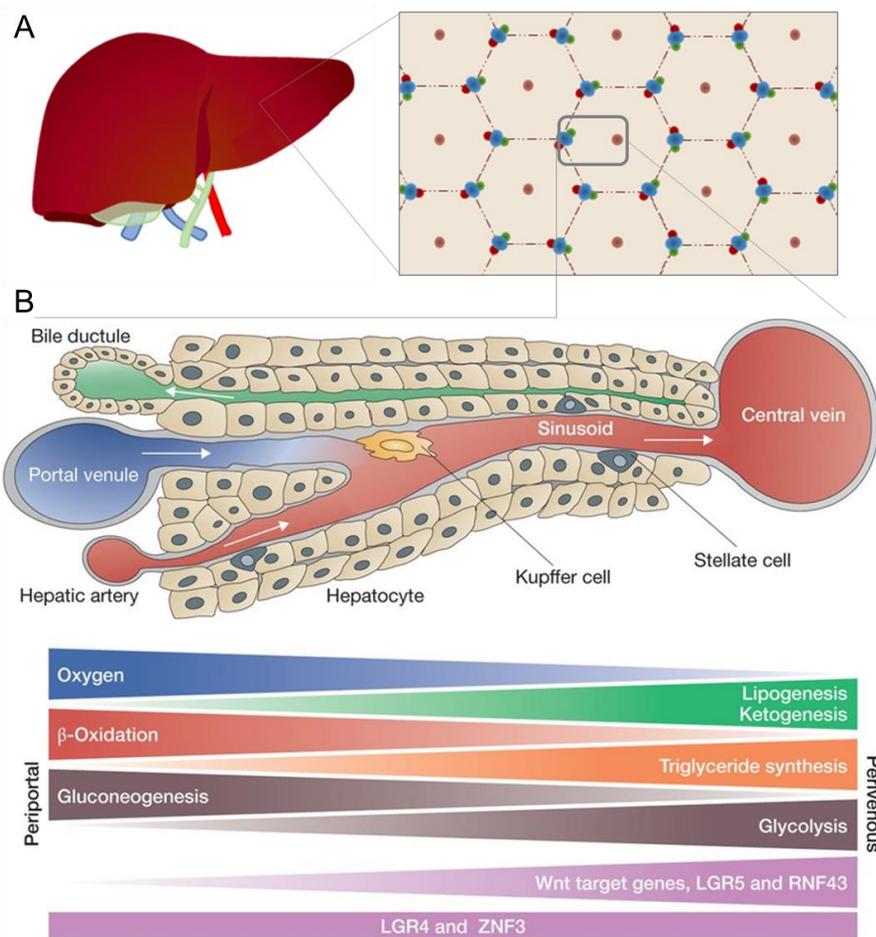
Figure 9: Metabolic liver zonation. A – The mammalian liver consists of hexagonal hepatic lobules with a central vein and peripheral portal triades, consisting out of bile ductile, portal vein and hepatic artery. B – Oxygen-rich blood flows from the portal vein through sinusoids towards the central vein. The resulting oxygen gradient along the porto-central axis is accompanied by gradients of metabolic functions, such as periportal gluconeogenesis or pericentral glycolysis. This so called metabolic zonation is maintained by gradients of Wnt signaling pathway members (Birchmeier, 2016, modified).

In human liver, the zonation of metabolic enzymes, including alcohol dehydrogenase (Sokal et al., 1993), fatty acid binding protein (Suzuki and Ono, 1987), glucose-6-phosphatase (Racine-Samson et al., 1996), glutamine synthetase (Racine-Samson et al., 1996) and several members of the cytochrome P450 family (Palmer et al., 1992; Sano et al., 1989) was already demonstrated by immunohistochemical stainings and activity measurements more than 20 years ago. Moreover, a study using laser capture microdissection coupled to RNA sequencing of normal liver from 3 human donors reported 139 transcripts with differential expression across the hepatic lobule (McEnerney et al., 2017), indicating an upregulation of Wnt signaling in pericentral hepatocytes as reported in mouse (Wang et al., 2015). These findings were recently supplemented by profiling of the human liver transcriptome on single cell level (MacParland et al., 2018). However, this study focused mostly on the classification of non-parenchymal cells, particularly revealing distinct macrophage subpopulations, and did not include a spatial reconstruction of hepatocytes along the porto-central axis, as was previously done in mouse by coupling scRNA-seq with single molecule fluorescence *in situ* hybridization (sm-FISH) of known zonation landmark genes (Halpern et al., 2017). Thus,

more sophisticated studies are necessary to obtain a deeper understanding of principles of zonated morphogenic and metabolic control in human liver, both in normal homeostasis as well as during the development of complex metabolic liver diseases.

1.3.2 Liver development

The understanding of liver development during embryogenesis is crucial for the proper interpretation of experimental findings regarding hepatic regeneration in normal and diseased liver as well as in context of *in vitro* differentiation of hepatocyte-like cells. Generally, it is characterized by a complex interplay of a spatially distinct and sequentially orchestrated set of signaling pathways such as Notch, Wnt, FGF, and BMP signaling, which regulate a cascade of stage-specific transcription factors determining cell fate.

In brief, hepatocytes, but also cholangiocytes, derive from endoderm, the inner germ layer of the early embryo. Definitive endoderm (DE), which in human develops from bipotential mesendodermal cells during the third week of gestation, is further segregated along the anterior-posterior axis by spatially distinct signaling from adjacent mesoderm (Gordillo et al., 2015). For instance, the posterior region of the gut is exposed to FGF4 and Wnt signaling, which results in an endodermal subpopulation expressing CDX2 that promotes the formation of the gastrointestinal tract (Stringer et al., 2012). Contrary, Wnt signaling is inhibited by secreted Sfrp5 in the ventral forgut endoderm, which thus gives rise to a population of bipotent hepato-pancreatic progenitors (Russell and Monga, 2018). Bone morphogenic protein (BMP), which is inhibits the Notch signaling pathway, and low levels of fibroblast growth factor (FGF) signaling from cardiac mesoderm are essential for commitment to hepatic cell fate (Gordillo et al., 2015). The Wnt signaling pathway was also implied to have major impact on hepato-pancreatic cell fate specification (Rodriguez-Seguel et al., 2013). Resulting fetal hepatoblasts, expressing albumin (ALB) and α-fetoprotein (AFP), form the liver bud, which additionally contains endothelial cells, developing stellate cells and, as the fetal liver constitutes a transitional site of hematopoiesis until 11th week of human gestation (Tavian and Péault, 2005), also hematopoietic progenitors. Hepatoblasts, which have the potential to differentiate into hepatocytes or cholangiocytes, are kept in a proliferative stage by TBX3 expression (Suzuki et al., 2008). Moreover, β-catenin, a key factor of the Wnt signaling pathway, is also involved in hepatoblast proliferation and maturation (Tan et al., 2008). A major regulator of the hepatoblast cell fate decision is the Notch signaling pathway. Notch2-mediated signaling, resulting in SOX9 expression, is crucial for cholangiocyte differentiation (Zong et al., 2009), while, interestingly, Notch3 was reported to be involved in hepatocyte lineage decision (Ortica et al., 2014). However, detailed mechanisms of human hepatocyte differentiation are not fully understood (Russell and Monga, 2018). In summary, this underlines the complex spatiotemporal patterns of different signaling pathways during liver organogenesis.

1.3.3 Hepatic regeneration

The mammalian liver exhibits an outstanding regeneration capacity. For example, following partial hepatectomy of two thirds of tissue in rodent models, full original liver mass and function can be restored within 2 weeks (Higgins and Anderson, 1931). However, in normal liver homeostasis only a small fraction of

hepatocytes is actually in replication (Macdonald, 1961). Over the years, the question regarding the origin of hepatic regeneration has been discussed controversially, and several mechanisms have been proposed. The intense proliferation of periportal biliary ductal cells upon severe liver injury, which is referred to as oval cell response, has been studied extensively in different rodent models (Lemire et al., 1991; Petersen et al., 1998; Sell et al., 1981) and was also reported in human (Roskams et al., 1998). Oval cells are atypical ductal cells expressing Epcam or Sox9, both proposed as liver progenitor markers (Dan et al., 2006; Furuyama et al., 2011), and have been considered as bipotent facultative progenitors that have the potential to differentiate into both hepatocytes and cholangiocytes (Kordes and Häussinger, 2013). However, several more recent mouse lineage tracing experiments reported that oval cells do not act as hepatocyte progenitors in common mouse models of chronic liver injury (Schaub et al., 2014; Tarlow et al., 2014). Thus, the oval cell response is now considered as an alternative regenerative mechanism in cases when all hepatocytes are chronically affected, for instance by toxins or viruses (Raven et al., 2017). Indeed, these studies also indicate mature hepatocytes as the predominant source of hepatic regeneration in chronic liver disease. Interestingly, the turn-over rate of mouse hepatocytes under normal conditions is rather low, and the replacement of the whole parenchyma takes about 1 year (Alison and Lin, 2011). Thus, the injury-induced massive proliferation of hepatocytes entails a coordinated cell-cycle entry of the remaining hepatocyte population (Corlu and Loyer, 2012). Earlier studies also refute a periportal to pericentral – or vice versa - hepatocyte streaming (Bralet et al., 1994; Kennedy et al., 1995), substantiating the hypothesis of stem cell activity independent hepatocyte self-renewal capacity.

Still, substantial evidence for the existence of stem cell niches in the adult mammalian liver has been gathered (Kordes and Häussinger, 2013). Recent publications actually suggest two different types of hepatic stem cell niches with specific functionality. A lineage tracing experiment, using Wnt-responsive Axin2 positive cells, in normal mouse liver revealed a subpopulation of diploid cells near the central vein expressing the early liver progenitor marker TBX3. These cells apparently feature a Wnt-signaling regulated self-renewal capacity in homeostatic liver and differentiate into polyploid mature hepatocytes (Wang et al., 2015). Wnt signaling itself is probably maintained by endothelial cells of the central vein (Preziosi et al., 2018). In contrast, lineage tracing using cholangiocyte-enriched SOX9 identified a periportal hepatocyte subpopulation, apparently regulated by Notch signaling (Köhler et al., 2004), that starts extensive proliferation upon mild chronic liver injury and apparently has the potential to restore the entire liver parenchyma (Font-Burgada et al., 2015). Moreover, lineage tracing, using Mfsd2a as periportal zonation marker, showed that periportal hepatocytes do not contribute to normal adult liver homeostasis, but expand and replace pericentral hepatocytes during liver injury. Intriguingly, after recovery these periportal-like cells around the central vein can restore their pericentral metabolic functions, while remaining Mfsd2a positive (Pu et al., 2016). Taken together, these findings suggest – at least in the rodent models – a pericentral Wnt signaling controlled hepatic stem cell niche which is responsible for homeostatic self-renewal, as well as a periportal Notch regulated source of regeneration upon liver damage. While Wnt signaling has also been implied in human liver homeostasis (McEnerney et al., 2017), the situation in human hepatocyte regeneration upon liver injury has not been characterized yet (Gilgenkrantz and Collin de l'Hortet, 2018).

1.3.4 Non-alcoholic fatty liver disease

Non-alcoholic fatty liver disease (NAFLD) is often referred to as a hepatic manifestation of the metabolic syndrome (Kim and Younossi, 2008). Major risk factors for the development of NAFLD are obesity and in its wake also type 2 diabetes mellitus (T2DM), and the combination of NAFLD and T2DM correlates with a poor prognosis (Cernea et al., 2017). Pathological overweight is generally defined as a body mass index (BMI) of more than 25, obesity of more than 30. The most fundamental cause of increasing body weight is a misbalance between calorie uptake and consumption, thus obesity correlates with a sedentary lifestyle as well as the calorie-rich western diet. The prevalence of obesity is increasing, not only in western countries, and was recently described by the World Health Organization as an "escalating global epidemic" (WHO, 2018), as it is also considered as major risk factor for cardiovascular diseases and certain types of cancer. Of major concern is also the rapid increase of obesity, with all its implications, in children during recent years (Abarca-Gómez et al., 2017).

The onset of NAFLD is defined as a liver steatosis (Figure 10), marked by the accumulation of fatty acids in lipid droplets in at least 5 % of hepatocytes (Benedict and Zhang, 2017). Bland steatosis without any inflammation is largely asymptomatic. However, in some cases it can progress to steatohepatitis (NASH), which is characterized by a pronounced infiltration of hepatic tissue with immune cells, particularly monocytes, granulocytes and innate lymphocytes (Narayanan et al., 2016). An additional hallmark of NASH is the so called hepatocyte ballooning, which denotes the appearance of enlarged, lipid droplet enriched hepatocytes that also feature cytoskeletal rearrangements and dilation of the endoplasmatic reticulum (Caldwell et al., 2010). Up to this stage, NAFLD has still the potential to fully recover to normal liver morphology and function, although elevated levels of certain liver enzymes, such as alanine transaminase, are already detectable, indicating a beginning impairment of liver functions (Cohen et al., 2011). Persistent inflammation in late NASH stages can then lead to deposition of excess extracellular matrix, forming fibrotic scar tissue. A widespread fibrosis, referred to as state 4 fibrosis, impacts and alters the normal liver morphology, resulting in a remodeled hepatic lobule architecture and vascularization. This pathophysiological alterations can lead to liver cirrhosis, which represents a severe condition, characterized by portal hypertension, that can eventually lead to liver failure (Schuppan and Afdhal, 2008). It is also frequently seen as a precondition for the development of hepatocellular carcinoma (Ramakrishna et al., 2013). The only curative therapy of late cirrhotic stages is liver transplantation (Pais et al., 2016). Although several drugs are under consideration, to date, there is no approved pharmacological treatment of NAFLD available (Cernea et al., 2017). Indeed, the progression to cirrhotic stages can often be prevented by careful lifestyle adjustment and weight reduction. In severe cases with morbid obesity and T2DM, entailing a poor prognosis, bariatric surgery can markedly improve insulin resistance as well as liver histology (von Schönfels et al., 2018). Thus, the detection of early stages of NAFLD is essential for proper disease management.
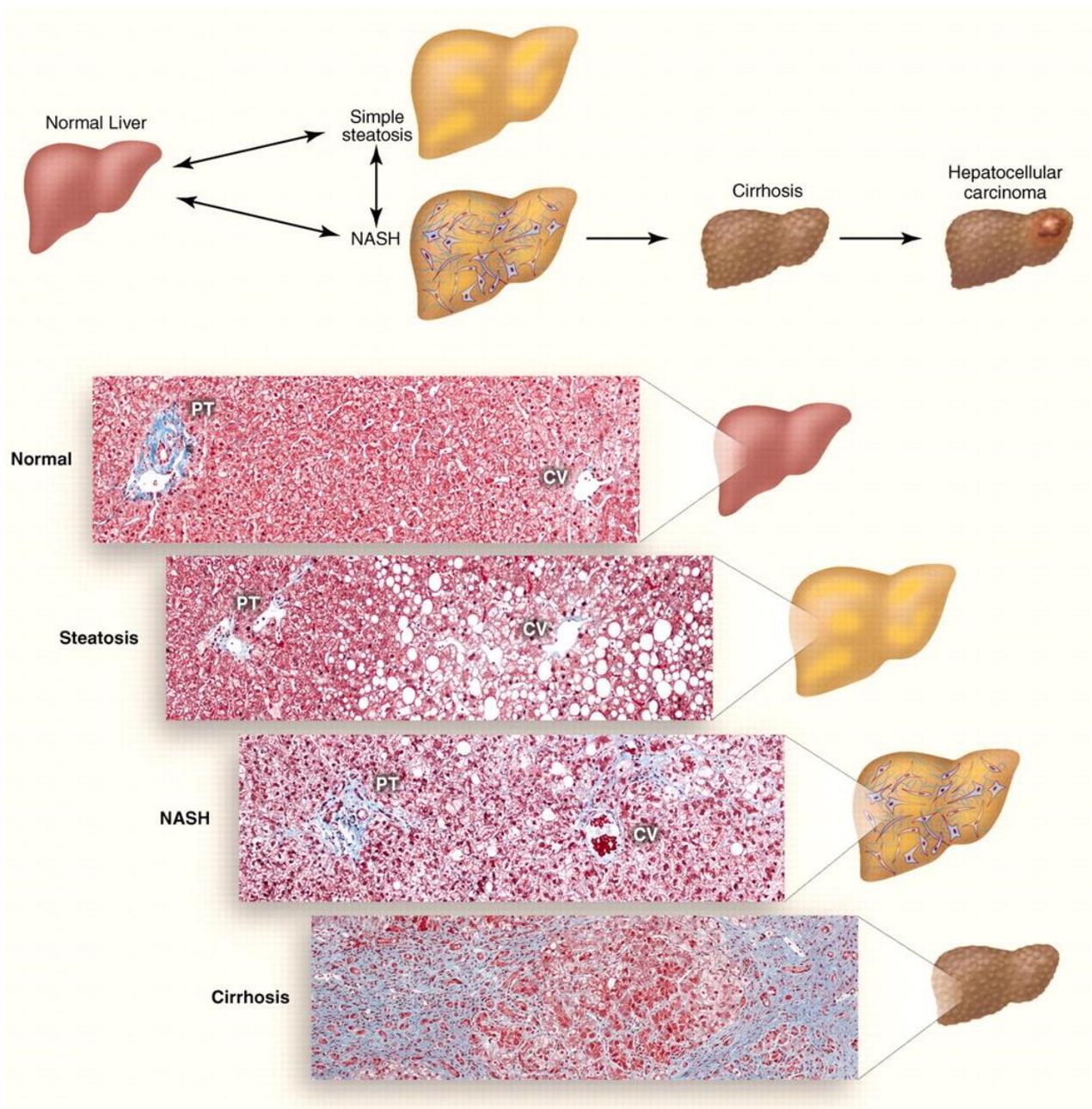
Figure 10: Histology of Non-Alcoholic Fatty Liver Disease progression (Cohen et al., 2011). Steatosis can develop from normal liver by accumulation of triglycerides in hepatocytes as lipid droplets (white) starting around the central vein. A bland steatosis can progress to steatohepatitis (NASH), which is characterized by invasion of inflammatory infiltrate, and further to cirrhosis with pronounced inflammation and fibrosis throughout the whole hepatic lobule. Collagen fibers are colored in light blue (Masson's trichrome staining).

Interestingly, although obesity is clearly associated with NAFLD, there are also obese donors that show normal liver histology and have no indication for fatty acid accumulation in hepatocytes (Machado et al., 2006). Moreover, there are also lean patients that develop NAFLD. Although having a normal BMI, they show metabolic profiles comparable to those of obese patients (Kumar and Mohan, 2017). Insulin resistance is however reduced in lean NAFLD patients, suggesting a different pathophysiological disease development (Kumar et al., 2013). Visceral abdominal adiposity, high fructose intake, increased cholesterol levels that

stimulate *de novo* lipogenesis, as well as genetic factors have been suggested as potential risk factors for non-obese NAFLD, but underlying mechanisms are not well understood (Kim and Kim, 2017).

In human adults, the onset of non-alcoholic steatosis is mostly restricted to pericentral hepatocytes (Figure 10, Cohen et al., 2011). In more progressed disease stages the accumulation of lipid droplets can spread throughout the whole hepatic lobule. Furthermore, the deposition of fibrotic tissue also starts around the central vein (Brunt et al., 1999). These observations argue for the zone-specific analysis of the development and progression of NAFLD. Remarkably, well established mouse models for steatosis or steatohepatitis feature differences regarding the zonal onset of fatty acid accumulation. While, for instance, in the leptin-deficient ob/ob mouse lipid droplet accumulation starts in pericentral hepatocytes (Stacchiotti et al., 2016), the methionine/choline-deficient model is characterized by a periportal steatosis (Koteish and Diehl, 2001). Moreover, both widely used models apparently do not reflect the human NAFLD phenotype in a sufficient manner (Anstee and Goldin, 2006). There is also little overlap comparing differential gene expression during NAFLD progression in human and different mouse models (Teufel et al., 2016). Thus, regarding their application to human NAFLD development, findings derived from mouse studies should be considered with care.

In contrast to NAFLD, alcoholic liver disease (ALD), which is the most prevalent type of chronic liver disease, is caused by excessive alcohol consumption for 1 to 5 years or more. Still, the development of ALD depends on many additional factors, including not only dose and duration of alcohol consumption, but also the type of alcohol, drinking patterns as well as further risk factors such as obesity or iron overload (Osna et al., 2017). Beyond further metabolic effects, permanent ethanol uptake apparently stimulates lipogenesis, resulting in fatty acid accumulation in hepatocytes (Osna et al., 2017). Thus, the spectrum of ALD resembles the progression of NAFLD from steatosis to steatohepatitis, fibrosis, cirrhosis and finally the potential to develop hepatocellular carcinoma, suggesting similarities in pathogenic features (Rowell and Anstee, 2015).

1.4 *in vitro* differentiation of hepatocyte-like cells

One of the liver's plethora of functions is the metabolism of xenobiotics including drugs, toxins and alcohol. Adverse drug effects are a main cause of hepatoxicity, which can lead to liver injury and eventually liver failure. Thus, for the development of new drugs it is of utmost importance to test effects on the human liver prior to clinical studies with human cohorts. Animal experiments for pharmacological drug assessment are still a compulsory part of clinical studies, but have been criticized for years, as rodents, but also larger mammals, do not reflect the human drug metabolism in a conclusive way (Akhtar, 2015; van Meer et al., 2012). A promising road towards animal-free drug assessment is the use of *in vitro* cultured human hepatocytes. To date, there are different alternatives for *in vitro* culturing systems aiming to resemble human liver detoxification.

The current gold standard for toxicological tests and pharmacological screenings are primary human hepatocytes (PHH), which can be maintained in spheroid culture for up to 5 weeks (Bell et al., 2016). In conventional cell culture PHH loose many hepatocyte related functions during hours, while sandwich matrigel cultures and 3D spheroids prevent culturing entailed dedifferentiation for at least 14 days (Bell et

al., 2018). However, their limited availability and inter-individual differences emphasize the need for the establishment of more reliable and versatile *in vitro* systems.

Immortalized cell lines such as the well described hepatocyte-like HepG2 (Sassa et al., 1987) or HepaRG (Guillouzo et al., 2007) lines are readily available, can be maintained in culture over long periods at relatively low costs, and particularly HepaRG cells express many hepatocyte specific genes at levels comparable to PHH (Hart et al., 2010). Although, after induction with hepatotoxic drugs such as phenobarbital or rifampicin, fully differentiated HepaRG cells exhibit CYP activity comparable to PHH, both HepG2 and HepaRG show low predictivity for hepatoxicity assessment of drugs (Gerets et al., 2012). Thus, these cell lines can certainly be used to gain insight into intriguing questions of fundamental research, but do not represent an ideal tool for pharmacological or toxicological studies.
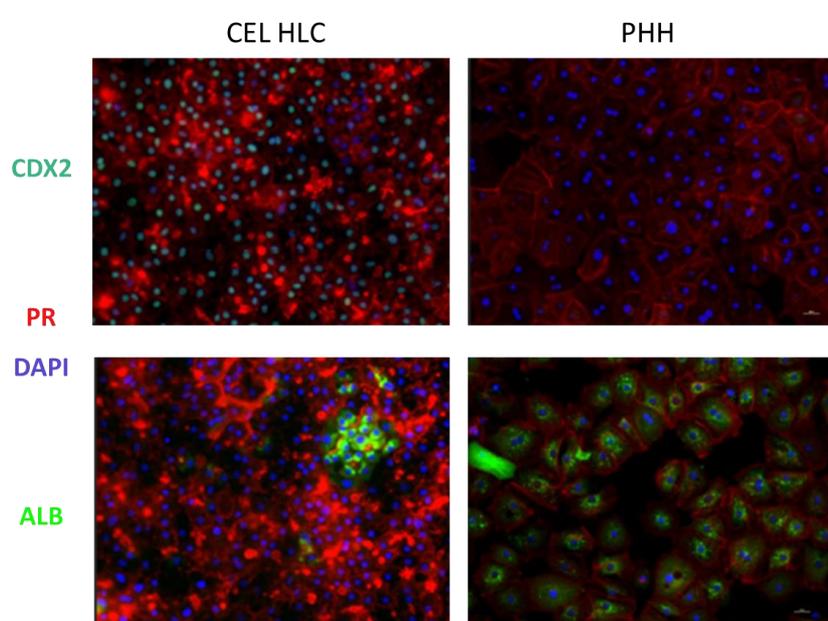
Figure 11: Immunohistological stainings reveal drastic differences within HLC subpopulations as well as in comparison with primary human hepatocytes (PHH). HLCs and PHH were stained for colon-specific marker CDX2 (upper panel) and hepatocyte marker albumin (ALB, lower panel). Nuclei were stained with DAPI (blue) and cellular membranes with an antibody against the membrane-localized progesterone receptor (PR, red). Immunohistological stainings (10 x magnification) were provided by Dr. Patricio Godoy and Patrick Nell (IfADo).

A promising alternative is the *in vitro* differentiation of induced pluripotent stem cells (iPSCs) to hepatocyte-like cells (HLCs) that should resemble expression profiles of PHH. To this end, differentiation protocols based on various combinations of extracellular matrix components with exogenous factors such as cytokines, growth factors and small biologically active molecules have been established. Generally, human iPSCs are differentiated to definitive endoderm (DE) and then further to hepatoblasts and HLCs. However, HLCs generated by different protocols show variable degrees of final differentiation (Godoy et al., 2015). Standardized commercially available solutions such as the Cellartis iPS Cell to Hepatocyte Differentiation System (Takara Bio, Goteborg, Sweden), further on referred to as CEL protocol, might offer more reliable *in vitro* systems. A more cost efficient alternative might also be a detailed protocol, recently published by the group of David Hay, thus further on referred to as HAY protocol, describing a robust iPSC to HLC differentiation system using recombinant human laminins as extracellular matrix (Wang et al., 2017). In 3D

29

culture these cells were reported to exhibit a stable liver-like phenotype for more than one year and even to support compromised liver function in animal models (Rashidi et al., 2018), thus representing a promising *in vitro* system for human liver.

Yet, gene expression profiles deviate severely from those of PHH (Bell et al., 2017; Godoy et al., 2018). Indeed, single cell RNA sequencing of 2D and 3D differentiated HLCs rather implies a fetal hepatoblast-like phenotype of HLCs (Camp et al., 2017). Moreover, a previous analysis of gene regulatory networks identified mixed HLC populations, particularly of liver, intestine, fibroblast and stem cell character (Godoy et al., 2015). This becomes also apparent in immunohistochemical stainings for certain marker genes. For instance, hepatocyte-marker albumin (Alb), which is highly expressed in PHH, is only detected in a subset of HLCs, while colon-specific marker CDX2 is highly abundant in most HLCs (Figure 11, kindly provided by Dr. Patricio Godoy and Patrick Nell, IfADo). Taken together, these findings highlight the need for further optimization of HLC culturing conditions towards a more homogenous PHH-like phenotype.

Besides their potential for general *in vitro* assessment of hepatoxicity and liver disease modelling, iPSC based HLC systems show promise for personalized medicine (Figure 12). In principle, iPSCs can be generated from any somatic cell-type. Thus, patient-specific HLCs could be cultivated for individual drug screenings to identify the most promising treatment with reduced adverse effects, as was already successfully described for familial hypercholesterolemia (Cayo et al., 2017).
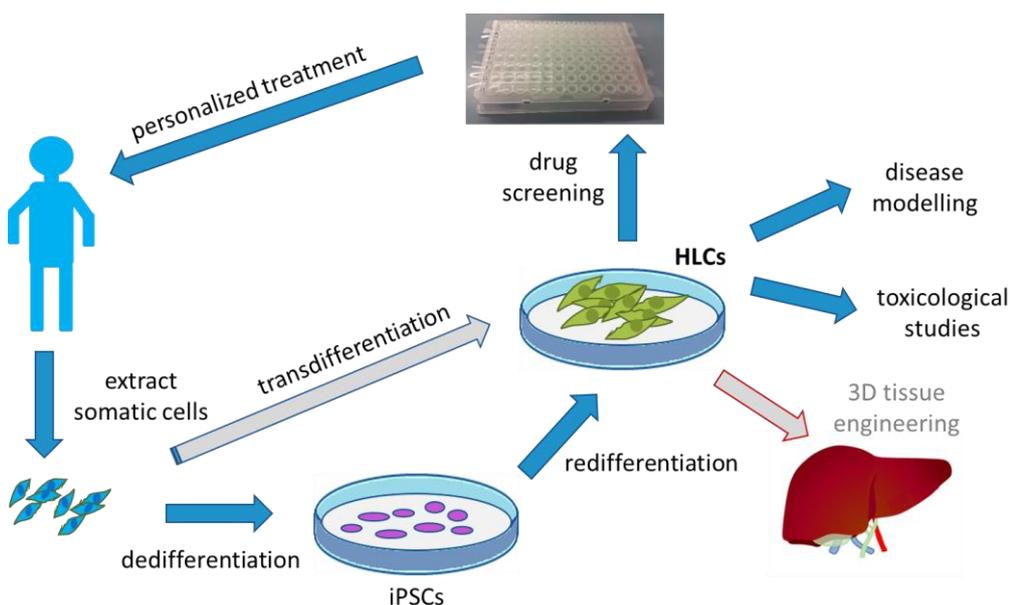
Figure 12: Applications for *in vitro* differentiated hepatocyte-like cells (HLCs). Somatic cells from human donors can be reprogrammed into induced pluripotent stem cells (iPSCs), which can then be re-differentiated into HLCs.

In addition, there are also advances towards tissue engineering (Figure 12). However, the liver is a complex organ with a plentitude of cell-types and structures. Thus, *in vitro* engineering of liver tissue would require extensive co-culturing of different cell-types. A further challenge would be the establishment of known oxygen and morphogen gradients that play important roles in the regulation of hepatocyte functions (chapter 1.3.1). Still, there are already intriguing approaches. For example, 3D HLC bioprinting, which refers to the exact placement of iPSC-derived HLCs into a biocompatible scaffold, mimicking extracellular matrix

components, resulted in metabolically active cells. However, especially the construction of tissue microenvironment and the integration of different cell-types remain major challenges. Thus, it is a long road towards functional liver tissue engineering and the usage of HLC systems for personalized drug screenings represents a more seizable goal.

1.5 Objectives

This thesis focuses on the transcriptional and epigenetic characterization of human hepatocyte subpopulations in health and complex metabolic diseases as well as during *in vitro* differentiation of hepatocyte-like cells, and is based on the generation, analysis and integration of genome-wide sequencing data. Depending on their spatial location within the hepatic lobule, hepatocytes show remarkable metabolic heterogeneity. In rodents, this phenomenon, termed metabolic zonation, is regulated by interacting morphogen gradients that result in the establishment of gene expression gradients along the porto-central hepatic axis. Interestingly, the same morphogens were also implied to have major impact on hepatic regeneration during normal liver homeostasis and chronic liver disease. Particularly in human, the underlying regulatory mechanisms are not well understood. Thus, a major focus of this thesis, in close cooperation with the group of Prof. Dr. Jochen Hampe (TU Dresden), is the characterization of NGS based, zone-specific transcriptome and DNA methylome profiles of human hepatocytes from 3 hepatic zones captured by laser microdissection (LCM). This genome-wide dataset (3.1.1) comprises healthy human liver as well as the early spectrum of NAFLD and provides the first integrated spatial epigenomic map of the human liver, which can be queried regarding multiple fundamental questions of human liver biology including regulation of metabolic zonation and hepatic regeneration. Moreover, the onset of several metabolic liver diseases such as NAFLD or ALD is characterized by the accumulation of lipid droplets in a zone-dependent fashion. This observation strongly suggests the zone-specific analysis of fatty liver disease progression. To this end, the human cohort is extended to patients with more progressed steatohepatitis and severe fibrosis to elucidate the role of hepatic zonation in NAFLD progression, and also includes samples from cirrhotic livers with an alcoholic background representing the end-stage of fatty liver disease. The unique experimental set-up of a pathologically well defined human cohort in combination with sophisticated LCM coupled to NGS based genome-wide transcriptome and DNA methylome analysis of small numbers of isolated cells allows the first spatially resolved analysis of fatty liver disease progression (3.1.2).

While single cell RNA sequencing is becoming more and more prominent, a major drawback is the loss of spatial tissue context. Still, as several studies imply gradient-like expression patterns in liver, single cell resolution, which is not feasible using standard LCM approaches, would contribute to a better understanding of many crucial questions. There are already promising bioinformatical advances regarding the modelling of single cell data into pseudotime. The same principles could be applied for spatial tissue reconstruction and might be further improved by the integration of spatial expression reference maps. Hence, the second part of this thesis demonstrates the spatial reconstruction of a single hepatocyte transcriptome dataset generated from human liver by making use of the reference maps described in chapter 3.1.1, and aims to identify possible applications for future analysis.

Besides the elucidation of regulatory mechanisms of general liver biology in health and disease, the *in vitro* differentiation of hepatocyte-like cells, which are not only valuable tools for hepatoxicity assessment and disease modelling, but might one day become reliable tools for treatment of chronic liver diseases, is of major interest. Although during recent years considerable progress regarding the establishment of HLC culture systems has been made, there is still room for improvement. Previous studies mainly focus on gene expression analysis to characterize the HLC differentiation process and resulting phenotypes. Hepatic differentiation is a complex process, regulated by interacting signaling pathways and stage-specific cascades of transcription factors. Yet, the impact of epigenetic mechanisms remains largely elusive. So, the final part of this thesis deals with the integration of gene expression data with DNA methylation and chromatin accessibility data, with particular emphasis on distinct HLC subpopulations detected by single cell RNA-seq. This approach might lead to the identification of additional factors that could be modulated in order to improve *in vitro* HLC differentiation.

2. Material and Methods

2.1 Material

2.1.1 Chemicals and reagents

Component	Company
1, 4-Dithiothreit (DTT)	Carl Roth (Karlsruhe, Germany)
AG 501-X8 resin	Bio-Rad Laboratories(Hercules, USA)
Agencourt Ampure XP Beads	Beckman Coulter (Krefeld, Germany)
ATP	New England Biolabs (Frankfurt, Germany)
Betaine	Sigma-Aldrich (Taufkirchen, Germany)
Biozym LE Agarose	Biozym Scientific (Hessisch Oldendorf, Germany)
Boric acid	Carl Roth (Karlsruhe, Germany)
Bromphenolblue	Sigma-Aldrich (Taufkirchen, Germany)
$CaCl_2*H_2O$	Sigma-Aldrich (Taufkirchen, Germany)
Chloroform	VWR International(Darmstadt, Germany)
Complete Protease Inhibitor Cocktail Tablets	Roche Diagnostics (Mannheim, Germany)
dATP	Solis BioDyne (Tartu, Estonia)
Desoxyribonucleotids (dNTPs)	Solis BioDyne (Tartu, Estonia)
Dimethylformamide	Sigma-Aldrich (Taufkirchen, Germany)
EDTA	Sigma-Aldrich (Taufkirchen, Germany)
EGTA	Sigma-Aldrich (Taufkirchen, Germany)
Ethidiumbromide	Carl Roth (Karlsruhe, Deutschland)
EtOH (absolute)	Sigma-Aldrich (Taufkirchen, Germany)
Fetal bovine serum (FBS)	Sigma-Aldrich (Taufkirchen, Germany)
Glucose	Sigma-Aldrich (Taufkirchen, Germany)
Glutamine	Sigma-Aldrich (Taufkirchen, Germany)
Glycogen	VWR International (Darmstadt, Germany)
HEPES	ThermoSientific (Darmstadt, Germany)
Isoamylalcohol	VWR International (Darmstadt, Germany)
K2HPO4	Merck (Darmstadt, Germany)
KCl	Merck (Darmstadt, Germany)
MEM Non-Essential Amino Acids Solution	ThermoSientific (Darmstadt, Germany)
MgCl2	Solis BioDyne (Tartu, Estonia)
Na2HPO4	Merck (Darmstadt, Germany)
NaCl	VWR International (Darmstadt, Germany)
NaOAc	Merck (Darmstadt, Germany)
Nonident P-40	Carl Roth (Karlsruhe, Germany)
nuclease-free water	Qiagen (Hilden, Germany)
PIPES	VWR International (Darmstadt, Germany)
Poly(vinylpyrrolidinone)	Sigma-Aldrich (Taufkirchen, Germany)
RNasIn (40 U / µl)	Promega (Fitchburg, USA)
Saccharose	Sigma-Aldrich (Taufkirchen, Germany)
SDS	AppliChem (Darmstadt, Germany)

Spermidine (free base)	MP Biomedicals(Santa Ana, USA)
Sodium butyrate	Sigma-Aldrich (Taufkirchen, Germany)
Tris-HCl	Carl Roth (Karlsruhe, Deutschland)
Triton	VWR International (Darmstadt, Germany)
Ultrapure Phenol : Chloroform : Isoamylalcohol	Invitrogen(Karlsruhe, Germany)
Xylencyanol	Sigma-Aldrich (Taufkirchen, Germany)

2.1.2 Buffers and Solutions

Buffer / Solution	Composition
1X TE (pH 8.0)	10 mM Tris-HCl 1 mM EDTA
2X TD buffer	20 mM Tris-HCl 10 mM $MgCl_2$ 20 % Dimethylformamide
5X TBE	890 mM Tris 20 mM EDTA 890 mM boric acid
10X CutSmart Buffer	New England Biolabs (Frankfurt, Germany)
10X PBS (pH 7.4)	43 mM Na_2HPO_4 14 mM K_2HPO_4 1.37 M NaCl 27 mM KCl
Chloroform-Isoamylalcohol	24 x Chloroform 1 x Isoamylalcohol
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)
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$_2$*H$_2$O |

ChIP lysis buffer	5 mM PIPES (pH 8)
	85 mM KCL
	10 mM sodium butyrate
	0.5 % Nonident P-40

DNA extraction B	10 mM EDTA
	10 mM Tris-HCl (pH 8.0)
	1 % SDS

Gel loading dye	40 % Saccharose
	0.05 % Bromphenolblue
	0.05 % Xylencyanol

| Elution Buffer | Qiagen (Hilden, Germany) |

Permeabilization Buffer A (pH 8.0)	60 mM KCl
	15 mM Tris-HCl
	15 mM NaCl
	1 mM EDTA
	0.5 mM EGTA
	0.5 mM Spermidine (free base)
	1 x PIC (Protease Inhibitor Cocktail Tablets)

| Qiagen PCR Buffer | Qiagen (Hilden, Germany) |

| RRBS Lysis Buffer (pH 8.0) | 10 mM Tris-HCl |
| | 5 mM EDTA |

Smartseq2 Lysis Buffer	25 µl RNasin (40 U / µl, Promega)
	470 µl 0.2 % (v/v) Triton X-100
	5 µl DTT (100 mM)

2.1.3 Enzymes

Enzyme	Cat.Number	Company
AluI (10 U/µl)	R0137S	New England Biolabs (Frankfurt, Germany)
Lyophilized Collagenase Type 1	C2674-1G	Sigma-Aldrich (Taufkirchen, Germany)
HaeIII (50 U/µl)	R0108T	New England Biolabs (Frankfurt, Germany)
HotStar Taq DNA Polymerase (5 U/µl)	203203	Qiagen (Hilden, Germany)
Klenow Fragment (3`→5` exo-, 5 U/µl)	M0212S	New England Biolabs (Frankfurt, Germany)
Nextera Tn5 Transposase	FC-121-1030	Illumina(San Diego, USA)
Proteinase K	19133	Qiagen (Hilden, Germany)
RNase A	19101	Qiagen (Hilden, Germany)

RNase-free DNase	79254	Qiagen (Hilden, Germany)
Superscript II RT (200 U/µl)	18064071	Invitrogen(Karlsruhe, Germany)
T4 DNA Ligase (2,000 U/µl)	M0202T	New England Biolabs (Frankfurt, Germany)

2.1.4 Reaction kits

Reaction Kit Name	Cat. number	Company
Agilent High Sensitivity DNA Kit	5067-4626	Agilent Technologies (Santa Clara, USA)
Direct-zol RNA MiniPrep Plus	R2070	Zymo Research(Irvine, USA)
ERCC RNA Spike-In Mix	4456740	Invitrogen(Karlsruhe, Germany)
EZ DNA Methylation-Gold Kit	D5006	Zymo Research(Irvine, USA)
KAPA HiFi HotStart ReadyMix PCR Kit	KK2602	Roche(Basel, Switzerland)
MinElute PCR Purification Kit	28006	Qiagen (Hilden, Germany)
NEBNext® High-Fidelity 2X PCR Master Mix	M0541L	New England Biolabs (Frankfurt, Germany)
Nextera DNA Library Prep Kit	FC-121-1030	Illumina(San Diego, USA)
Nextera XT DNA Library Preparation Kit	FC-131-1096	Illumina(San Diego, USA)
Nextera XT Index Kit	FC-131-1002	Illumina(San Diego, USA)
Nextera XT Index Kit v2 Set D	FC-131-2004	Illumina(San Diego, USA)
PerfeCTa® NGS Quantification Kit	733-2300	Quanta Biosiences(Gaithersburg, USA)
Qubit® dsDNA HS Assay Kit	Q32854	Invitrogen(Karlsruhe, Germany)
TruSeq DNA Single Index Set A	20015960	Illumina(San Diego, USA)
TruSeq DNA Single Index Set B	20015961	Illumina(San Diego, USA)
TruSeq Dual Index Sequencing Primer PE	FC-121-1003	Illumina(San Diego, USA)
TruSeq PE Cluster Kit v3-cBot-HS	PE-401-3001	Illumina(San Diego, USA)
TruSeq SBS Kit v3 - HS (200 Cycles)	FC-401-3001	Illumina(San Diego, USA)
TruSeq SR Cluster Kit v3-cBot-HS	GD-401-3001	Illumina(San Diego, USA)

2.1.5 Oligonucleotides

RP-HPLC purified and lyophilisated oligonucleotides (Table 1) were obtained from biomers.net GmbH (Ulm, Germany) with exception of the TSO which ordered from Eurogentec Proteomics (Cologne, Germany).

Table 1: Oligonucleotide sequences (B = 5' biotin, 7 = LNA g, 9 = RNA-G, V = A/C/G, N = any base, i5 and i7 = 8 index read bases as described in the Nextera xt index kit (Illumina)).

Name	Sequence [5' → 3']
Oligo-dT primer	AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
TSO	B-AAGCAGTGGTATCAACGCAGAGTACAT997
IS PCR primer	AAGCAGTGGTATCAACGCAGAGT
V1 short TruSeq adapter 1	TCTTTCCCTACACGACGCTCTTCCGATCT
V1 short TruSeq adapter 2	GATCGGAAGAGCAC ACGTCTGAACTCCAGTCAC
V1 indexed TruSeq PCR primer i5	AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCCGAT
V1 indexed TruSeq PCR primer i7	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC[i7]ATCTCGTATGCCGTCTTCTGCTTG
V2 TruSeq PCR primer 1	AATGATACGGCGACCACCGAGATCTACAC
V2 TruSeq PCR primer 2	CAAGCAGAAGACGGCATACGAGAT

Nextera PCR Primer i5	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC
Nextera PCR Primer i7	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG

2.1.6 Primary human tissue and cells

Clinical characteristics of primary human tissues used for laser capture microdissection experiments (chapter 3.1) are described in Table 2. Moreover, cryopreserved primary human hepatocytes from 3 male donors (DJJ, AFJ and IAN; Cat.Number: M00995-P) were obtained from BioreclamationIVT (Baltimore, USA) for single cell RNA-seq (chapter 3.3).

Table 2: Classification of surgical liver samples obtained from human donors (donor, n=45) for laser capture microdissection of hepatic zones. Samples were grouped regarding their phenotype into normal controls (NC), healthy obese controls (HO), bland steatosis (STEA), early NASH with mild inflammation (EARLY), progressed NASH (NASH) and fibrotic NASH (FIB3). Moreover, cirrhotic samples from donors with alcoholic fatty liver disease (ACI) are included. The table also contains information about sex (f = female, m = male), NAS score (NAS), NAS ballooning score (Bal), fat content (Fat, in %), inflammation score (Infl), fibrosis grade (Fib), age and body mass index (BMI). Insulin (Ins [nmol/l]) values were used to obtain HOMA indices to assess insulin resistance (IR = HOMA > 2.5). Diabetes refers to a known diabetes diagnosis. Initially sample 7522 was classified as cryptogen. Due to corresponding fibrosis grade and high RNA-seq data correlation it was assigned to FIB3 samples.

Donor	Group	Sex	NAS	Bal	Fat	Infl	Fib	Age	BMI	Ins	HOMA	IR	Dd
7012	NC	F	0	0	0	0	0	54	22	0.0	0.3	no	no
7173								54	26	0.0	0.7		
7373					3			57	20.6	0.0	0.6		
7213								44	NA	0.1	4.4	yes	
7279					0			59	26.5	0.1	2.6		
7194		M						62	27.1	0.0	1.2	no	
6758	HO	F	1	0	5	0	0	36	40.9	0.0	1.1	no	no
7519					0			31	45.6	0.0	0.9		
6963					3			35	62	0.2	6.7		
6975			0					44	42.5	0.1	4.1	yes	
6976					0			40	55.9	0.1	4.3		
7230								36	40.2	0.2	4.2		
7252								29	45.4	0.1	3.7		
6922		M	1		5	1		68	31.7	0.0	1.5	no	yes
6967	STEA	F	2		40			46	46	0.1	9.1	yes	yes
7172			1		35			53	61.9	0.4	16.8		
7181			3		70		0	46	43.5	0.1	4.2		no
7468			2	0	60	0		29	62.1	0.2	6.3		
7484					70			37	38.9	0.1	3.8		
7137		M	3		90	1		67	33.2	0.1	9.3		yes
7251					70		0	34	48.1	0.5	18.7		no
7041	EARLY	F	2	0	40	1	0	44	50.6	0.1	5.1	yes	no
7188				0				51	45.1	0.1	3.1		
7344			3	0	50			58	51.2	0.3	20.3		yes
7425				0	60			45	42.4	0.7	30.9		

ID		Sex											
6610		M		1	50			41	59.7	0.1	5.2		
7157			4	1	40		1	44	40.3	0.1	7.4		
7415			3	0			0	19	55.8	0.1	2.5		no
7185	NASH	F	7	1	70	0	0	33	53.8	0.1	1.4		no
7196			6	2	90		1	54	48.9	0.0	1.0	no	no
7322				1	80	0		51	69.1	0.0	0.4		yes
7286			5	1				47	38.6	0.1	3.2		no
7299				1	70	1		51	57.5	0.5	36.8	yes	yes
7319				1	80			39	46.8	0.1	8.7		yes
7203	FIB3	F	NA	NA	50		1	57	49.1	0.4	8.3	yes	NA
7522					40	1	3	82	27.7	0.2	8.1		NA
7197		M	6	2	100			53	50.8	0.1	4.6	yes	yes
7326			NA	NA	40	2		51	58	0.3	11.5		NA
7263	ACI	F	NA	NA	NA			66	27.1	NA			
7479					0			64	NA				
7490					10	1		51					
7259		NA			NA		4	39	28.6				
7505		M			0			59	20.8				
7265					0	3		63	24.9				
7383		NA			0	2		NA	28.4				

2.1.7 Cultured Cells

Cell culture (Table 3) was performed at the Leibniz Research Centre for Working Environment and Human Factors (IfaDo, Dortmund) by Patrick Nell and David Feuerborn.

Table 3: Sample overview of *in vitro* differentiation from induced pluripotent stem cells (iPSCs) to definite endoderm (DE) to hepatocyte-like cells (HLCs).

Cell-type	Differentiation protocol	Replicates	Assay
iPSC	CEL	4	RNA-seq
iPSC	CEL	4	RRBS
iPSC	CEL	4	ATAC-seq
DE	CEL	4	RNA-seq
DE	CEL	4	RRBS
DE	CEL	4	ATAC-seq
HCL	CEL	4	RNA-seq
HCL	CEL	4	RRBS
HCL	CEL	4	ATAC-seq
HCL	CEL	3	scRNA-seq
iPSC	HAY	1	RNA-seq
iPSC	HAY	1	RRBS
iPSC	HAY	1	ATAC-seq

DE	HAY	1	RNA-seq
DE	HAY	1	RRBS
DE	HAY	1	ATAC-seq
HCL	HAY	1	RNA-seq
HCL	HAY	1	RRBS
HCL	HAY	1	ATAC-seq
HCL	HAY	1	scRNA-seq

2.1.8 Equipment

Besides general molecular biology equipment the following devices were used:

Device	Company
Agilent 2100 Bioanalyzer	Agilent Technologies(Santa Clara, USA)
cBot	Illumina(San Diego, USA)
DynaMag	Invitrogen(Karlsruhe, Germany)
HiSeq 2500	Illumina(San Diego, USA)
NanoDrop 2000c	ThermoSientific (Darmstadt, Germany)
Qubit Fluorometer	Invitrogen(Germany, Karlsruhe)

2.1.9 Software and Databases

ChIPseek - chipseek.cgu.edu.tw

Cytoscape - https://cytoscape.org

DAVID - https://david.ncifcrf.gov

deepTools - http://deeptools.ie-freiburg.mpg.de

Draw Venn Diagram - http://bioinformatics.psb.ugent.be/webtools/Venn

ENCODE at UCSC database - https://genome.ucsc.edu/ENCODE

Galaxeast - galaxeast.fr

Galaxy - https://galaxyproject.org

GeneTrail2 - https://genetrail2.bioinf.uni-sb.de

HOMER Motif Analysis - homer.ucsd.edu/homer/motif/

IGV browser - https://software.broadinstitute.org/software/igv

jupyter notebook - https://jupyter.org

Microsoft Office Standard 2013 -

pathview - https://pathview.uncc.edu

python (version 3.6) - https://www.python.org

R (version 3.4.3) - http://www.R-project.org

R studio - https://www.rstudio.com

RegulatorTrail - https://regulatortrail.bioinf.uni-sb.de/

Sequence Read Archive (SRA) - https://www.ncbi.nlm.nih.gov/sra

Additionally, numerous R and python packages, detailed in chapter 2.2.2, were used.

2.1.10 External Datasets

Single cell RNA-seq data (Camp et al., 2017) with SRA accession numbers for raw fastq files:
GSE81252 (SRP075475)
GSE96981 (SRP102427)

2.2 Methods

2.2.1 Experimental methods

2.2.1.1 Reduced representation bisulfite sequencing on laser capture microdissected material (LCM-RRBS)

Laser capture microdissection of three zones along the hepatic proto-central axis was performed on human liver cryosections (Table 2, 2.1.6 Primary human tissue) by Luise Obermann in Jochen Hampe's group (TU Dresden) as previously described (Brosch et al., 2018). Subsequently, microdissected cells were shipped on dry ice and stored at - 80 °C. A Reduced Representation Bisulfite Sequencing (RRBS) library preparation protocol (Schillebeeckx et al., 2013) was modified as described in Brosch et al. (2018) to prepare adequately concentrated Next Generation Sequencing libraries from the roughly 100 hepatocytes typically obtained by microdissection. Cell lysis was carried out directly in the caps of the capturing tubes. Cells were suspended in 8 µl lysis buffer (10 mM Tris-HCl, 5 mM EDTA) and 2 µl Proteinase K (diluted to 1 mg per ml in lysis buffer, Qiagen) and tubes were incubated upside down in a thermal shaker (200 rpm) at 55 °C for 3 hours. Then tubes were briefly centrifuged to collect the lysate at the bottom of the tubes. Proteinase K activity was inhibited with 0.5 µl Pefabloc (21 mM, Sigma) and incubation for 1 hour at room temperature. The CpG methylation insensitive enzymes HaeIII (25 U, NEB) and AluI (5 U, NEB) were used for DNA restriction in a double digest complemented with 1.5 µl CutSmart buffer (10X, NEB), 1 µl MgAce (5mM, Sigma) and 1 µl yeast tRNA (100 ng, ThermoFisher Scientific) at 37 °C for 18 hours. Restriction enzymes were heat-inactivated at 80 °C for 20 min. A-Tailing of DNA fragments was performed by addition of 1 µl Klenow Fragment (3' → 5', exo −, 5 U per µl, NEB) and 1 µl dATP (10 mM, NEB) at 37°C for 30 min followed by enzyme inactivation at 75°C for 20 min. Ligation of methylated adapters was achieved with 1 µl adapter, 0.5 µl T4 Ligase (2,000 U per µl, NEB) and 2 µl ATP (10 mM, NEB) at 16 °C for 22 hours. Initial samples (Donor IDs 7041, 7157, 7188, 7344, 6610, 6758, 6922, 7230, 7213, 7252, 7012, 7173, 7194, 7279, 6967, 7137, 7172, 7181 and 7251) were ligated using short pre-annealed TruSeq compatible adapters (100 µM, V1_short_TruSeq_adapter_1 and 2, 2.1.5 Oligonucleotides) without index. Later samples were ligated with 1 µl of original indexed TruSeq adapter (TruSeq DNA Single Index Set B, Illumina) diluted to 0.5 µM. After bisulfite conversion with the EZ DNA Methylation-Gold Kit (Zymo Research) DNA was eluted in 10 µl M-elution buffer (Zymo Research). Enrichment polymerase chain reaction (PCR) of adapter ligated fragments was performed using 3 µl PCR Buffer (10X, Qiagen), 1.2 µl MgCl2 (25 mM, Qiagen), 2.4 µl dNTPs (10 mM)

and 0.5 µl Hot Start Taq (5 U/µl, Qiagen) in a 30 µl reaction. Initial samples (LCM-RRBS_1 to 57) were amplified with each 0.5 µl of indexed TruSeq primers (10 µM, V1 TruSeq_i5 and TruSeq_i7, 2.1.5 Oligonucleotides), remaining libraries with each 0.5 µl non-indexed RRBS primers (10 µM, V2 TruSeq_PCR_Primer_1 and 2, 2.1.5 Oligonucleotides). After initial denaturation at 95 °C for 15 min, thermo-cycling 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. Libraries were purified with 0.8 X Ampure XP Beads (Beckman Coulter) and eluted in 10 µl elution buffer (Qiagen).

2.2.1.2 RRBS

Genomic DNA was isolated by phenol chloroform isoamylalcohol extraction. It should be noted that only fixed frozen cell pellets, originally prepared for ChIP-seq, were available. Thus, a modified protocol was developed for proper isolation of intact genomic DNA from fixed material.

Frozen cell pellets were thawed in 700 µl ChIP lysis buffer (without protease inhibitor cocktail) and resuspended on a vortexer. Then, 350 µl of the suspension were transferred into a separate tube and used for parallel chromatin preparation, which was however not used within the scope of the present thesis. The remaining 350 µl were used for extraction of genomic DNA. To this end, 350 µl DNA extraction buffer B were appended, mixed, supplemented with 3 µl RNase A (20 mg/ml, Qiagen) and samples were incubated at 37 °C for 30 min. Protein digestion was performed using 14 µl Proteinase K (10 mg/ml, Qiagen) at 55 °C for 1 hour and then 65 °C over night in a shaking device. For DNA extraction 700 µl Ultrapure Phenol-Chloroform-Isoamylalcohol (PCI, 25:24:1, v/v/v, Invitrogen) were added. Samples were rotated for 15 min at room temperature and centrifugation was carried out at 13,000 rpm for 5 min. The aqueous phase was transferred to a new tube. PCI extraction was repeated, followed two times by extraction with Chloroform-Isoamylalcohol (24:1). For DNA precipitation the final aqueous phase was distributed into two 2 ml tubes, each appended with 775 µl ice-cold absolute ethanol and 33 µl 3 M NaOAc, inverted and stored at – 20 °C over night. Visible DNA threads were fished with a 10 µl tip, washed 3 times by dipping in 70 % ice-cold ethanol, dried at room temperature and solved in 100 µl 1 X TE at 50 °C for 5 hours. DNA quality was assessed by electrophoresis on a 1.5 % agarose gel in 0.5 X TBE stained with ethidium bromide solution. DNA concentration and purity was measured on a NanoDrop 2000c.

RRBS library preparation was carried out similar as described for LCM-RRBS (2.2.1.1) with some modifications. In brief, at least 180 ng (183 to 458 ng) purified genomic DNA were digested using 1 µl HaeIII (50 U, NEB) in a 30 µl reaction supplemented with 3 µl 10 X CutSmart Buffer (NEB) at 37 °C. After 2 hours 0.5 µl HaeIII (50 U/µl) were appended 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 µl 10 mM dATP at 37 °C for 30 min followed by enzyme inactivation at 75 °C for 20 min. Ligation of indexed methylated TruSeq 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 hours. Bisulfite conversion, enrichment PCR (15 Cycles) and final library purification were performed as described for LCM-RRBS (2.2.1.1).

2.2.1.3 single cell RNA-seq

Cryopreserved primary human hepatocytes (PHH) were thawed in a water bath at 37 °C, transferred into 5 ml PBS supplemented with 10 % fetal bovine serum and carefully mixed by inversion.

Adherent hepatocyte-like cells (HLCs) in matrigel culture dishes were washed twice with Dissociation Buffer 1 and incubated for at least 15 min in Dissociation Buffer 2 containing collagenase (200 mg / ml, Sigma) at 37 °C. As soon as the monolayer started to detach cells were carefully scraped with a cell scraper. Given that HLCs were rather clumpy the cell suspension was pipetted up and down about 10 times to reduce clumps. Finally, HLC suspensions were transferred to 1.5 ml tubes, centrifuged for 5 min at 300 g and 4 °C, washed twice in cold PBS and resuspended in PBS containing 0.1 % PVP.

PHH and HLC suspensions were diluted in PBS containing 0.1 % PVP and single cells were manually isolated using a glass capillary into 2 μl Smartseq2 lysis buffer and stored at − 80 °C. cDNA preparation was performed as previously described (Picelli et al., 2013) with modifications. To each well 1 μl 10 mM dNTPs, 0.5 μl Oligo-dT primer (2.1.5 Oligonucleotides) and 0.5 μl ERCC RNA Spike-In Mix (Invitrogen, diluted 1:400,000) were added and 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 (2.1.5 Oligonucleotides) 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 (2.1.5 Oligonucleotides) 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) and elution in 7 μl elution buffer (Qiagen). After quality control of 11 randomly picked wells per replicate on an Agilent Bioanalyzer with the HS DNA Kit, in average 450 pg of cDNA were used as input for dual-indexed Nextera XT library preparation (Illumina) following the manufacturer's recommendations with 9 cycles for library amplification. All 96 individual single cell libraries per replicate were pooled in equal amounts prior to 1X Ampure XP Beads (Beckman Coulter) purification of 100 μl pooled libraries with final elution into 15 μl elution buffer (Qiagen).

2.2.1.4 RNA-seq

RNA was extracted from approximately 1 million *in vitro* cultured cells stored at − 80 °C in 1 ml TRIzol (ThermoFisher) using the Direct-zol RNA MiniPrep Plus Kit (Zymo Research) including DNaseI treatment following the manufacturer's recommendations. RNA concentration and purity was assessed by NanoDrop 2000c (ThermoSientific). cDNA synthesis was carried out as described for single cells (2.2.1.3) but adapted for bulk RNA sequencing. Thus, 100 to 237 ng total RNA in 2.3 μl nuclease-free H_2O were combined with 1 μl 10 mM dNTPs and 1 μl of 10 μM Oligo-dT primer (2.1.5 Oligonucleotides), incubated at 72 °C and placed on ice. Reverse transcription and preamplification were carried out as described in chapter 2.2.1.3 except for applying only 5 cycles in the preamplification PCR. cDNA was quantified using the Qubit dsDNA HS Assay Kit (Qiagen). Nextera libraries were prepared by combining 4 ng cDNA with 0.5 μl Tn5 (Nextera DNA

Library Prep Kit, Illumina) in 1 X TD Buffer and incubation at 55 °C for 10 min. Tagmented DNA was purified with the MinElute PCR Purification Kit (Qiagen) and final elution in 11 µl elution buffer (Qiagen). Libraries were amplified using 15 µl NEBNext® High-Fidelity 2X PCR Master Mix (NEB), 1 µl Nextera PCR Primer i5, 1 µl Nextera PCR Primer i7 (2.1.5 Oligonucleotides) in a 30 µl reaction with the following thermo-cycling program: 72°C for 5 min and 98 °C for 30 sec, followed by 8 cycles of 98 °C for 10 sec, 63 °C for 30 sec, 72 °C for 1 min and final elongation at 72 °C for 5 min. Final library purification was achieved with 0.9 X Ampure XP Beads (Beckman Coulter) and eluted in 10 µl elution buffer (Qiagen).

2.2.1.5 ATAC-seq

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) on *in vitro* cultured HLCs and cryopreserved PHH was performed as described by Buenrostro et al. (2013) with modifications. After initial establishment of cell-type specific conditions nuclei preparation, tagmentation and following purification of HLCs was carried out by Patrick Nell at IfADo (Dortmund) and shipped on dry ice for further processing.

Single cell suspensions in PBS were prepared as described for scRNA-seq (2.2.1.3). Prior to nuclei preparation centrifugation was carried out at 300 g and 4 °C for 5 min. For nuclei preparation cells were resuspended in 1 ml Permeabilization Buffer A (pH 8.0) supplemented with 0.05 % Nonident P-40 and incubated on ice for 10 to 15 min until isolated nuclei were visible by microscopic control. HLC nuclei were generally very clumpy, thereby not allowing proper counting. Thus, nuclei counts were estimated on the expected cell density of the used culture dish and aliquoted respectively to obtain approximately 50,000 nuclei per reaction. PHH nuclei were counted in a Neubauer chamber and aliquoted accordingly. Nuclei were centrifuged, washed with 1 ml 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. Purification, amplification (9 cycles) and final library purification were performed as described for RNA-seq Nextera library preparation (2.2.1.4).

2.2.1.6 Library quality control and Next Generation Sequencing

Prior to sequencing libraries were quantified with the Qubit dsDNA HS Assay Kit (Invitrogen) and fragment size distribution for exemplary libraries was checked on the Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit (Agilent). Libraries of the same type (Nextera respective TruSeq) were pooled per flow cell lane according to required sequencing depth. Library pools were eventually quantified with the PerfeCTa NGS Quantification Kit (Quanta Biosiences) and normalized for clustering on a cBot (Illumina). LCM-RRBS, RRBS, ATAC-seq and RNA-seq libraries were sequenced on a HiSeq2500 (Illumina) using TruSeq SBS Kit v3 – HS Chemistry in single read runs with 94 bp read length. scRNA-seq libraries were sequenced accordingly, except for requiring dual index reads thereby limiting the read length to 88 bp. In case of Nextera sequencing (scRNA-seq, RNA-seq and ATAC-seq) read 1 primer hp8 was replaced by hp12. Final quantification, clustering and sequencing was performed by Dr. Gilles Gasparoni.

2.2.2 Bioinformatical methods

2.2.2.1 Data processing

Processing of raw FASTQ files was done by Dr. Karl Nordström. Initial raw data quality controls was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). 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 as well as adapter sequences. RRBS reads were trimmed in RRBS mode. For the EpiTriO project, in the first round of analysis (V0.1, chapter 3.1.1), reads were aligned to the 1000 genomes version of the human GRCh37 reference and further processed as described below. For subsequent analysis of the extended EpiTriO dataset (V0.2, chapter 3.1.2), all samples (including V0.1 samples) were re-processed and aligned to the more recent human genome reference GRCh38. All further datasets used in this thesis (chapter 3.2 and 3.3) were mapped to GRCh38, accordingly.

RRBS reads were aligned with the BWA (Li and Durbin, 2009, v0.6.2) wrapper methylCtools (Hovestadt et al., 2014, v0.9.2). Samtools (Li et al., 2009, v1.3) and Picard tools (v1.115) (http://broadinstitute.github.io/picard/) was 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 methylation calls.

ATAC-seq reads were aligned with the GEM mapper (Marco-Sola et al., 2012, v1.376β), and samtools (Li et al., 2009, v1.3) 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 (Feng et al., 2012, v2.1.0.20140616), applying the following parameters: --nomodel, --shift -125, --extsize 250. Coverage files normalized for library size were generated using bamCoverage from deepTools (Ramirez et al., 2014, v1.5.9.1). Normalized read counts were counted using featureCounts (Liao et al., 2014, v1.5.0-p3) in 100 bp and 10,000 bp bins for subsequent analysis.

Bulk RNA-seq reads were aligned using STAR (Dobin et al., 2013, v2.5.2a) with per sample 2-pass mapping strategy, that processes all reads in both passes as described in the tool vignette. PCR duplicates were detected using MarkDuplicate from Picard tools (version 1.115; http://broadinstitute.github.io/picard/). Reads aligned to GRCh37 were summarized to v19 Gencode gene models (Harrow et al., 2012) with featureCounts (Liao et al., 2014, v1.5.0-p3) counting primary alignments only. Gene and transcript-wise read counts for data mapped to GRCh38 were estimated based on Gencode release 30 (GRCh38.p12) using RSEM (Li and Dewey, 2011, v1.3.1). Smartseq2 generated scRNA-seq reads were processed accordingly. For UMI based scRNA-seq reads obtained by the 10X Genomics Chromium Single Cell Gene Expression Solution, the zUMIs pipeline, including Star alignment to GRCh38, was applied as previously described (Parekh et al., 2018), returning an R object containing different count matrices such as raw exon-intron counts or exon-only counts.

2.2.2.2 NGS data analysis

2.2.2.2.1 RNA-seq data

RNA-seq data quality was judged by number of detected genes and transcripts per sample, and the 1000 most variable genes, calculated based on expression values normalized as log CPM + 1 (counts per million), were used for Principal Component Analysis (PCA).

The initial script for differential analysis (V0.1) of RNA-seq was provided by Dr. Karl Nordström and served as template for following analysis. EdgeR (Robinson et al., 2010, v3.16.5) was used to detect differentially expressed genes with significance thresholds of maximal FDR adjusted p-value of 0.01 and minimal absolute logFC of 1, based on gene wise negative binomial generalized linear models. For the EpiTriO project genes for which fewer than five samples had a CPM value above 0.5 were excluded. Inter-individual variance of expressed genes within groups was calculated based on log (CPM+1) after filtering of unexpressed genes using the matrixStats R package (https://github.com/HenrikBengtsson/matrixStats). After calculation of normalization factors, dispersion was robustly estimated using a model adjusting for phenotype, sex, age, BMI, donor and the prevalence of diabetes (V0.1). As for ACI samples no clinical data were available regarding the prevalence of diabetes, zonal comparisons including this phenotypic group were only adjusted for sex, age, BMI and donor (V0.2). For samples with missing clinical data (age or BMI), the value was set to the mean of the remaining samples in the respective group. Differential analysis between phenotypes was done without donor adjustment. In the StemNet dataset (chapter 3.3) genes for which row means of raw counts were below 1 were excluded, and DEG analysis was perfomed as described above, using a model without covariate adjustment.

2.2.2.2.2 RRBS data

General quality of DNA methylation data was analyzed by RnBeads (Assenov et al., 2014). Selection criteria for CpG sites analyzed by PCA are explained in respective figure legends. Differential DNA methylation analysis of the v0.1 EpiTriO dataset was performed by Dr. Karl Nordström using MethylKit (Akalin et al., 2012, v1.3.1). All further analysis were performed accordingly. MethylKit was modified to treat missing sites as zero-covered sites in the aggregation step of tiled analysis. After merging both strands, calls were filtered to sites with at least 10X coverage. Then, the tiled dataset (500 bp tiles) was searched for differentially methylated regions (step-size: 500 bp), allowing for missing sites in no more than 75 % of samples per group (e.g. EpiTriO V0.1: 14 out of 19 samples need to be covered, StemNet: 3 out of 4 replicates need to be covered), with at least 3 CpGs, a maximum FDR adjusted p-value of 0.01 and minimal methylation difference of 5% (EpiTriO) or 20% (StemNet). Mitochondrial, X and Y chromosomes were excluded from the analysis. Resulting differentially methylated regions (DMRs) were annotated to closest genes and genomic features (promoter, exon, intron or intergenic) based on Gencode v19 (EpiTriO v0.1 samples) or Gencode release 30 (GRCh38.p12) gene models (Harrow et al., 2012) using bedtools (Quinlan and Hall, 2010, v2.20.1) or GenomicRanges (Lawrence et al., 2013, v1.36.0).

Genomic segmentation of RRBS data was performed by Abdulrahman Salhab with the R package MethylSeekR (Burger et al., 2013) as previously described (Salhab et al., 2018).

Calculations of DMR enrichment and DNA methylation gradients at TFBS was performed by Alexander Hermann (TU Dresden). The detailed method is provided in our publication (Brosch et al., 2018).

For the EpiTriO dataset, mitochondrial abundance was deduced from the ratio of mitochondrial versus genomic reads in the RRBS data. Significance of differences between hepatic groups was tested by Wilcoxon rank sum test with the wilcox.test function in R. P-values less than 0.05 were considered as significant.

2.2.2.2.3 ATAC-seq data

ATAC-seq quality was assessed by visual inspection of coverage tracks using the IGV browser (Robinson et al., 2011), by number of detected peaks and fraction of reads in peaks (FRiP score) as well as the ratio of mitochondrial reads. Bin-wise counts (generated using featureCounts as described in 2.2.2.1) were summarized as log (CPM + 1) and the 50,000 most variable 100 bp bins were selected for PCA.

For differential chromatin accessibility analysis of ATAC-seq data, the mitochondrial chromosome and a set of mostly telomeric and centromeric blacklist regions with exceptionally high read counts (sum of all 18 samples in the dataset > 10,000 reads / 100 bp) were excluded. Bin-wise counts (generated using featureCounts as described in 2.2.2.1) were summarized as log (CPM + 1). Differential analysis was then performed using csaw (Lun and Smyth, 2016, v1.16.1) combined with edgeR (Robinson et al., 2010, v3.16.5). To adjust background differences between samples, counts of 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. Moreover, average log2 CPM values were calculated for merged windows. Annotation to the nearest gene was achieved as described for DMRs (chapter 2.2.2.2.2).

2.2.2.2.4 Smartseq2 scRNA-seq data

Quality of Smartseq2 single cell RNA-seq data was assessed as described for bulk RNA-seq (chapter 2.2.2.2.1). In addition, the Ambion® ERCC Spike-In Control spike-in was used to assess expression profiling efficiency. The Smart-seq2 scRNA-seq dataset was extended by a compatible external dataset (Camp et al., 2017). Thus, batch effects were assessed by PCA visualization of the 1,000 most variable genes, PCA regression, average silhouette width calculation, and kBET rejection rate (Büttner et al., 2019) as outlined in the kBET vignette. In addition, batch correction was performed by mutual nearest neighbor correction (Haghverdi et al., 2018) and ComBat implemented within the R package Surrogate Variable Analysis (Leek et al., 2017). Corrected data were assessed again, but as no major batch effects were observed and corrections did not yield major improvements (Sup. Figure 12), original data were used for further analysis. Data were normalized as log (CPM + 1).

Pseudotime analysis was performed using monocle2 (Trapnell, 2017) in semi-supervised mode with CDX2, ALB and AFP as marker genes for anchoring of the analysis as described in package vignette. Regulon analysis was done using a recent python implementation of the Single-CEll regulatory Network Inference

and Clustering (SCENIC) workflow (Aibar et al., 2017), called pySCENIC (https://github.com/aertslab/pySCENIC). Regulons with activity scores above automatically assigned AUC score threshold in at least 30 % of single cells were considered as active in the respective cell cluster. Binarized regulon activities were used for hierarchical clustering and correlation based t-SNE visualization of single cells. Single cell variance was calculated using the R package matrixStats (https://github.com/HenrikBengtsson/matrixStats) with the function RowVars on normalized expression data after removal of unexpressed genes.

2.2.2.2.5 High-throughput 10X Genomics scRNA-seq data – spatial reconstruction

Single cell RNA-seq data generated by the high-throughput 10X Genomics platform were analyzed by SEURAT (Butler et al., 2018, v2.3.4). Cells were filtered by number of detected genes (500 – 2,000), UMI count (2,000 – 12,500) and percentage of mitochondrial reads (< 0.15). High confidence primary hepatocytes were identified by PCA on highly variable genes (HVGs), which was used to remove outliers. Further filtering was performed based on ALB (log2 normalized expression level > 4) and CD47 expression (log2 normalized expression level < 4).

Spatial reconstruction of single high confidence hepatocytes was in principle performed by dimensionality reduction of highly variable genes (HVGs), which were detected by SEURAT using the function FindVariableGenes that controls for the high correlation between dispersion and average expression level. First, all HVGs detected in the dataset were subject to several dimension reduction algorithms. Diffusion maps were generated by the R package destiny (Haghverdi et al., 2015, v2.6.2), PCA and t-SNE were calculated within SEURAT. Uniform Manifold Approximation maps (UMAP) were generated in python using the UMAP package (Becht et al., 2019). Then, dimension reductions were performed on the 107 HVGs that featured zonated expression profiles in normal human liver tissue (Brosch et al., 2018). All dimension reduction plots were generated with the ggplot2 package (Wickham, 2009).

As by visual inspection of marker gene distribution, diffusion maps showed the best performance, further analysis focused on diffusion map based spatial reconstruction. Several diffusion parameters, such as changing the local Gaussian kernel width sigma from local to an optimal global sigma of 2.9 or testing different numbers between 100 and 7,912 (number of cells – 1) for k, did not visibly impact the diffusion map results. Thus, all displayed diffusion maps were generated with destiny's default parameters. It should be noted that strict filtering of low abundance transcript was not necessary, as the diffusion algorithm was apparently driven by highly expressed genes.

For interspecies comparisons, a spatially reconstructed external mouse liver scRNA-seq dataset generated by MARS-seq (Halpern et al., 2017) was used. Reconstructed mouse zones 1 and 2 were considered as pericentral, zones 4 and 5 as intermediate and zones 8 and 9 as periportal. Spearman rank statistics of sequencing depth normalized mRNA expression levels (log(CPM+1)) were calculated for mouse single cell data (Halpern et al., 2017), human LCM-RNAseq data and human single cell RNA-seq data with one-to-one matching orthologues detected in all three datasets. The comparison was restricted to genes with high confidence expression values, corresponding to a CPM > 0 in the mouse dataset after background subtraction as previously described (Halpern et al., 2017). The detection of zonated gene expression from

reconstructed scRNA-seq datasets was performed as described by Halpern et al. (2017), with Kruskal-Wallis test q < 0.2 between CV and PV considered as significantly zonated.

2.2.2.2.6 Integrative data analysis

Association of DMRs or DARs to DEGs was performed by annotating the closest genes as described above (chapter 2.2.2.2.2). The overlap of DMRs, respective DARs, was analyzed using bedTools (Quinlan and Hall, 2010) implemented on the GalaxEast Server (http://galaxeast.fr). Correlations between datasets were calculated with the R cor.test function. If not mentioned otherwise correlation values (r) are Pearson correlations. The use of Spearman rank statistics is specifically mentioned.

Hg38 annotated bed files were lifted to genome version hg19 using LiftOver from UCSC (https://genome.ucsc.edu/cgi-bin/hgLiftOver) to allow the further annotation of genomic features and HOMER (Heinz et al., 2010) based TF motive enrichment analysis by ChIPseek (Chen et al., 2014).

Gene ontology overrepresentation analysis was done using DAVID (Jiao et al., 2012) with at least 3 genes per GO term and p < 0.05. Gene Set Enrichment Analysis (GSEA) was performed by Kolmogorov-Smirnov Test with enrichment score > 0.5 and Benjamini-Yekutieli FDR adjusted p-value < 0.01 using Genetrail2 (Stöckel et al., 2016). Interaction networks of differential genes were generated by String (Szklarczyk et al., 2019) and visualized in using Cytoscape (v3.5.1, Shannon et al., 2003), using continuous mapping of additional parameters such as expression logFC or methylation difference for coloring.

2.2.2.2.7 Data visualization

In general, dimension reduction plots, scatterplots, violin plots, boxplots and volcano plots throughout this thesis were generated using the R package ggplot2 (Wickham, 2009). Bar diagrams and colored tables were generated in Microsoft excel. Pie charts were either produced in Microsoft excel or taken from the output of the ChIPseek online tool (Chen et al., 2014). Venn diagrams of overlapping gene IDs were generated by DrawVennDiagram (http://bioinformatics.psb.ugent.be/webtools/Venn), while Venn diagrams of overlapping genomic regions were calculated using the R package GenomicRanges (Lawrence et al., 2013).

ATAC-seq coverage patterns were prepared using deeptools (Ramirez et al., 2014) based on normalized coverage tracks (bigwig) by the functions compute matrix in 4 kb windows around region centers as reference points and visualized by plotHeatmap. All other heatmaps were produced by the R package pheatmap (https://CRAN.R-project.org/package=pheatmap), with underlying values being either original values (if applicable, normalized to sequencing depth) or scaled z-scores. Clustered heatmaps were clustered either by euclidean, manhattan or canberra distance, as outlined in the respective figure legends. Genomic tracks of expression, DNA methylation and ATAC-seq data (bigwig), normalized by sequencing depth, were visualized in the IGV browser (Robinson et al., 2011). DNA methylation tracks were set to a range of 0 to 100, while DNA methylation coverage tracks, ATAC-seq and RNA-seq tracks were scaled to the highest detected coverage of the visualized samples.

For illustration of expression levels of exemplary genes across three hepatic zones, expression data were normalized to the zone with the highest expression for each individual gene. Plots were generated in Microsoft excel. It should be noted that dashed lines are for visualization only.

Bubble plots for GO term enrichment of hepatic zones were produced using the R package GOplot, while other GO term enrichment visualizations were achieved by bar-plots of enrichment scores using ggplot2. KEGG pathways were visualized using Pathview (Weijun Luo, 2017) with continuous mapping of expression logFC or methylation differences. Protein interaction networks, which were generated by String as described in chapter 2.2.2.2.6, were visualized by the java based application Cytoscape (v3.5.1, Shannon et al., 2003), using continuous mapping of expression logFC and methylation difference as fill and border colors to integrate differential gene expression or DNA methylation into the network.

3. Results

3.1 Epigenomic analysis of microdissected human liver zones

The transcriptional and epigenomic analysis of microdissected human liver zones was carried out in cooperation with the group of Prof. Dr. Jochen Hampe (TU Dresden) as part of the BMBF funded EpiTriO project (Figure 13). Laser capture microdissection of hepatocytes from three zones along the hepatic porto-central axis (Figure 14, A) was performed on human liver cryosections by Luise Obermann, Fabian Reichel and Dr. Mario Brosch (TU Dresden), and shipped on dry ice for preparation of RRBS libraries. Moreover, RNA-seq data generated from adjacent cryosections was provided for integrated data analysis.
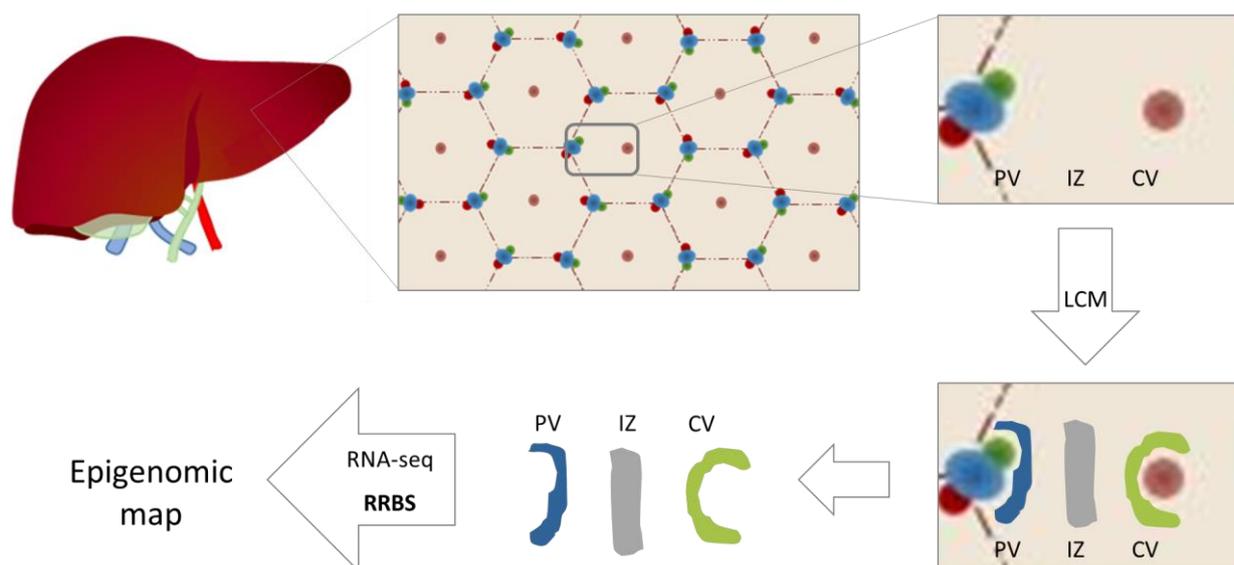
Figure 13: Workflow for the generation of epigenetic reference maps from zonal hepatocytes. Hepatocytes from three hepatic zones (PV, IZ and CV) were isolated using laser capture microdissection (LCM) on cryosections of human liver samples. Microdissected cells from adjacent cryosections were used for RNA-seq and RRBS. PV: periportal zone, IZ: intermediate zone, CV: pericentral zone.

During RRBS protocol establishment for low cell numbers it became apparent that RRBS using HaeIII as sole restriction enzyme yielded functional libraries (Figure 14, B), but CpG coverage was rather poor. In contrast, AluI/HaeIII double digested RRBS libraries resulted in high quality data with broad coverage of about 10 million CpGs per sample. As LCM of zonal hepatocytes yielded approximately 90 cells per sample, all LCM-RRBS libraries were generated based on AluI/HaeIII double digest, thereby allowing the comprehensive analysis of DNA methylation profiles. RRBS libraries were typically sequenced to at least 40 million reads per library or an average CpG coverage of 5 (Sup. Table 1). Sequencing of RRBS libraries was carried out by Dr. Gilles Gasparoni and processing of raw RNA-seq and RRBS data was performed by Dr. Karl Nordström as outlined in chapter 2.2.2.1.
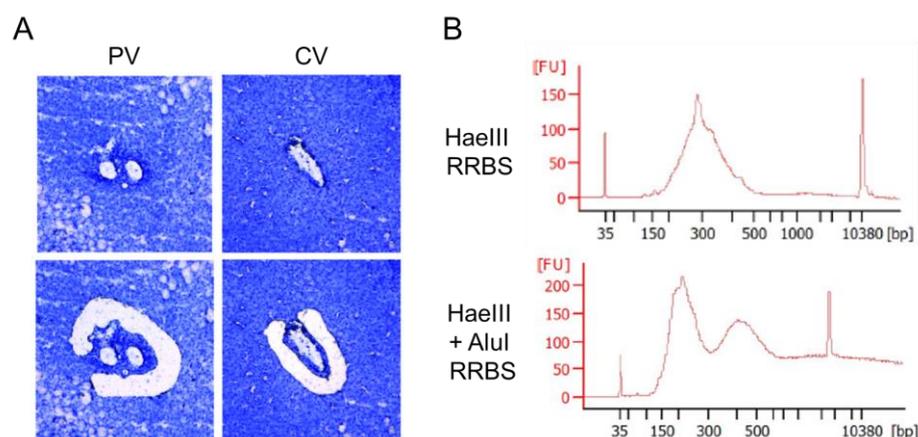
Figure 14: A – Exemplary histological Cresyl Violet staining of a periportal (PV) and pericentral (CV) region (Brosch et al., 2018). B – Bioanalyzer tracks representing the typical fragment length distribution of HaeIII RRBS and HaeIII/AluI double digested RRBS on about 90 hepatocytes obtained from microdissection.

The present study comprises a human cohort of surgical liver samples from 38 donors including normal lean (NC) and healthy obese controls (HO), together with NAFLD patients spanning the entire spectrum of non-alcoholic fatty liver disease (NAFLD) from bland steatosis (STEA), early NASH (EARLY), progressed NASH (NASH) to NASH with fibrosis grade 3 (FIB3). As no NASH samples with higher fibrosis grade were available, we additionally included samples of 7 donors with an alcoholic cirrhosis of fibrosis grade 4 (ACI), representing the irreversible state of chronic steatohepatitis (Figure 15). Control samples (NC) were gained from patients undergoing liver resection for metastasis assessment, while NAFLD samples were obtained from bariatric surgery. Cirrhosis samples derive from surgical liver resections. Importantly, all samples were flash frozen within less than 40 sec after surgery, ensuring high cellular integrity. Control and NAFLD samples were categorized by standard liver histological parameters (Kleiner et al., 2005) into phenotypic groups according to pathologically determined NAS scores. Donor information including sex, age, BMI, NAS score and further histological and clinical parameters are summarized in Table 2 (chapter 2.1.6).

The first round of analysis (V.01, n = 19) focused on zonation signatures, thus it was limited to controls (NC, HO) and the early spectrum of NAFLD (STEA, EARLY) with low amounts of inflammatory infiltrate and no fibrosis. The main findings were recently published (Brosch and Kattler et al., 2018) and are presented in a more detailed fashion in chapter 3.1.1, with analysis produced by other co-authors specifically outlined. Already published figure panels are cited accordingly. In a second analysis (V.02, n = 45) we extended our dataset by further replicates of controls and early NAFLD samples to obtain more statistical power, but importantly also included donors with more progressed and fibrotic NASH (FIB3) as well as alcoholic cirrhosis samples (Figure 15). Here, the analysis was focused on zone-specific deregulation and its impact on hepatic zonation patterns during disease progression. The results are described in chapter 3.1.2.
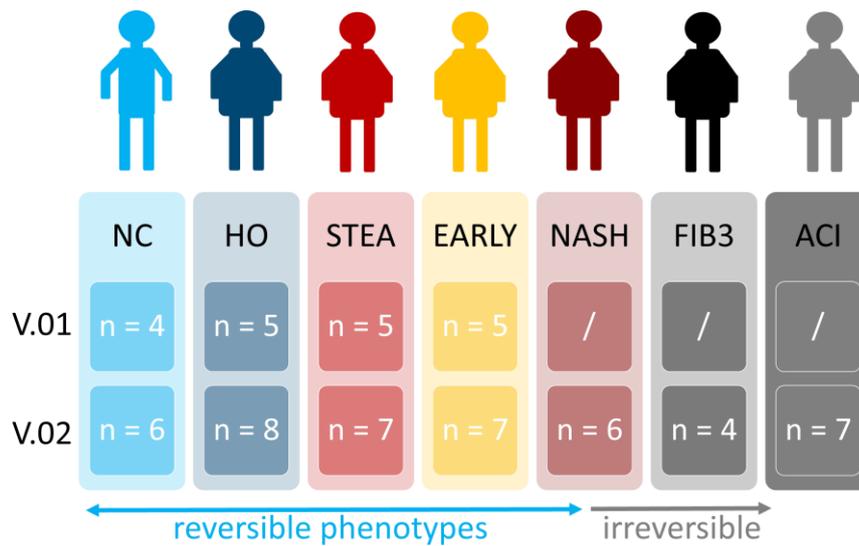
Figure 15: Human cohort (n = 45) comprising surgical liver samples of normal controls (NC), healthy obese controls (HO) as well as donors with bland steatosis (STEA), early NASH (EARLY), progressed NASH (NASH), fibrotic NASH (FIB3) and alcoholic cirrhosis (ACI). The first stage of analysis is based on 19 donors of controls and the early spectrum of NAFLD (V.01), in the second round (V.02) 26 additional donors including NASH, FIB3 and ACI were added. Donor information including sex, age, BMI, histological factors such as NAS score and further clinical parameters are summarized in Table 2 (chapter 2.1.6).

3.1.1 Hepatic zonation signatures in health and early NAFLD

3.1.1.1 Transcriptional zonation profiles

For the analysis of transcriptional zonation profiles of the human liver, full-length mRNA-seq was performed on hepatocytes (n = 19) isolated from pericentral, intermediate and periportal hepatic zones by laser microdissection (Table 2). The resulting RNA-seq dataset comprises normal controls (NC, n = 4), healthy obese donors (HO, n = 5), patients with bland steatosis (STEA, n = 5) and early steatohepatitis (EARLY, n = 5). Samples were sequenced to a mean depth of 20 million reads (Sup. Table 2). Raw fastq files were provided by Prof. Dr. Jochen Hampe's group and analyzed as described in chapter 2.2.2.

On average, 17,922 genes respectively 103,780 transcripts were expressed, and high intragenic (0.933) and particularly exonic rates (0.861) confirm high data quality with efficient expression profiling (Sup. Table 2). Potential contamination with non-parenchymal cells such as immune cells, endothelial cells or fibroblasts was assessed by expression analysis of known cell-type marker genes (Sup. Figure 1 A) and revealed neglectable fractions of non-parenchymal cells, ensuring the predominant presence of hepatocytes. A principal component analysis (PCA) of the 1000 most variable transcripts (Figure 16 A) yielded significant correlations between hepatic zones and the first two principal components (PC1: r = 0.47, p= 2.0×10^{-4}; PC2: r = 0.86, p < 2.2×10^{-16}), explaining 91.78 % of total variability in the dataset (Sup. Figure 2 A). This indicates zonation as a major driver of variance in the RNA-seq dataset.
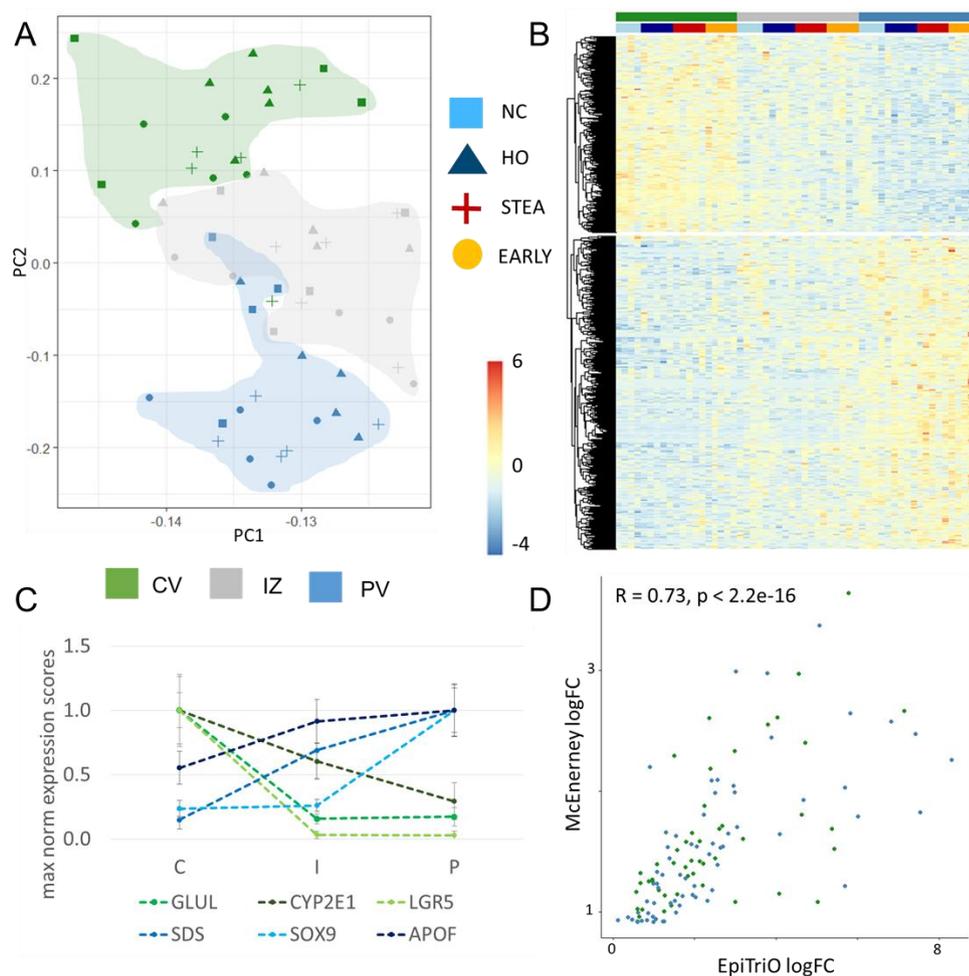
Figure 16: Transcriptional zonation along the porto-central axis revealed by lcm-RNA-seq on microdissected hepatic zones of the human liver (n = 19). Pericentral (CV) is colored in green, intermediate (IZ) in grey and periportal (PP) in blue. The annotation legend with corresponding symbols displays controls (NC) in light blue, healthy obese (HO) in dark blue, steatosis (STEA) in red and early NASH (EARLY) in orange. A – Principal component analysis of the 1000 most variable transcripts (Brosch et al., 2018). B – Expression z-scores of zonated genes (805) at pericentral, intermediate and periportal zones determined by edgeR with |log2FC| > 1 and FDR < 0.01 between pericentral and periportal samples. The heatmap is clustered by Pearson correlation between genes (Brosch et al., 2018). C – Maximum normalized expression scores (including standard deviation) in human hepatic zones of exemplary genes with previously described strong zonation gradients in mouse liver (Halpern et al., 2017). Dashed lines are for visualization and do not necessarily reflect the actual transcriptional gradient between zones (Brosch et al., 2018). D – Comparison with a recent lcm-RNA-seq based study on pericentral and periportal hepatocytes from 3 human donors (McEnerney et al., 2017). The scatterplot displays the Pearson correlation of log2 fold changes of genes with zonal expression (120) detected by McEnerney et al.

Differential expression analysis between zones was performed using edgeR with a paired model correcting for phenotype, sex, age, BMI, and diabetes diagnosis. In total, 805 genes were differentially expressed (|log2FC| > 1 and FDR < 0.01) between the periportal and the pericentral zone. Interestingly, we did not detect any genes with highest expression in the intermediate zone, but rather transcriptional gradients along the porto-central axis (Figure 16 B). While 317 genes showed upregulated expression around the central vein, 488 genes with predominant periportal expression were detected. We did not observe any significant

alternative splicing between zones (data not shown). Moreover, it should be noted that differential expression analysis using the HOMA index, a more precise measure for insulin resistance than mere diabetes diagnosis, as parameter for covariate adjustment results in the detection of 987 zonated genes, due to improved FDR correction. The subsequently presented analysis is, however, based on the set of 805 zonated genes to ensure consistency with our already published findings (Brosch et al., 2018).

In our human dataset, known zonation landmark genes described in mouse (Halpern et al., 2017) showed corresponding zonation profiles (Figure 16 C). A previous study, limited on 3 human donors, revealed 68 genes with periportal and 52 genes with pericentral predominant expression (McEnerney et al., 2017). A comparison of our RNA-seq data showed high Pearson correlation (r = 0.73, p < 2.2e-16) between the log fold changes of the genes detected as zonated by McEnerney et al (Figure 16 D). However, our dataset tends to have higher fold changes, which might be due to more stringent microdissection of the hepatic zones. Moreover, our analysis extends the analysis of McEnerney et al. as we detected considerably more zonal genes due to higher statistical power.

Reducing our dataset to samples with normal liver histology (NC, HO, n = 9) only 470 genes with statistically significant zonation profiles were detected. Comparing log2 fold changes of zonated genes in the limited (n = 9) and the full analysis (n = 19) including steatotic samples (STEA, EARLY), Pearson correlations from 0.80 to 0.98 (all p < 2.2 x 10^{-16}, Figure 17) were detected, indicating an impact of pairwise effect sizes on FDR cut-offs due to limited power in the reduced dataset, instead of biological variance between control and steatosis samples. It should be noted that differential expression analysis between steatotic and control samples for pericentral, intermediate and periportal zone revealed in total a deregulation of 467 genes (Sup. Figure 3), however largely irrespective of hepatic zone. Thus, although there are phenotype-related transcriptional differences, the relative zonation patterns apparently remain stable in the early phases of NAFLD (Figure 16, B), which will be described more detailed in chapter 3.1.2.
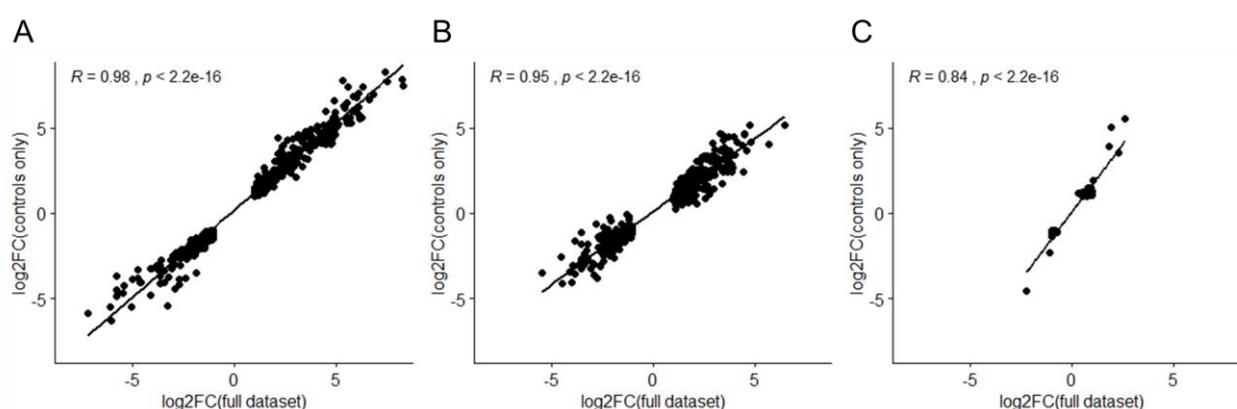
Figure 17: Transcriptional zonation in the full dataset (including steatosis and early NASH) compared to zonation detected in normal lean and healthy obese controls only. Scatterplots of log2 fold changes of zonated genes detected in both comparisons (A), only in the full dataset (B) respective only in the control dataset (C) with according Pearson correlations (Brosch et al., 2018).

3.1.1.2 Epigenetic zonation profiles
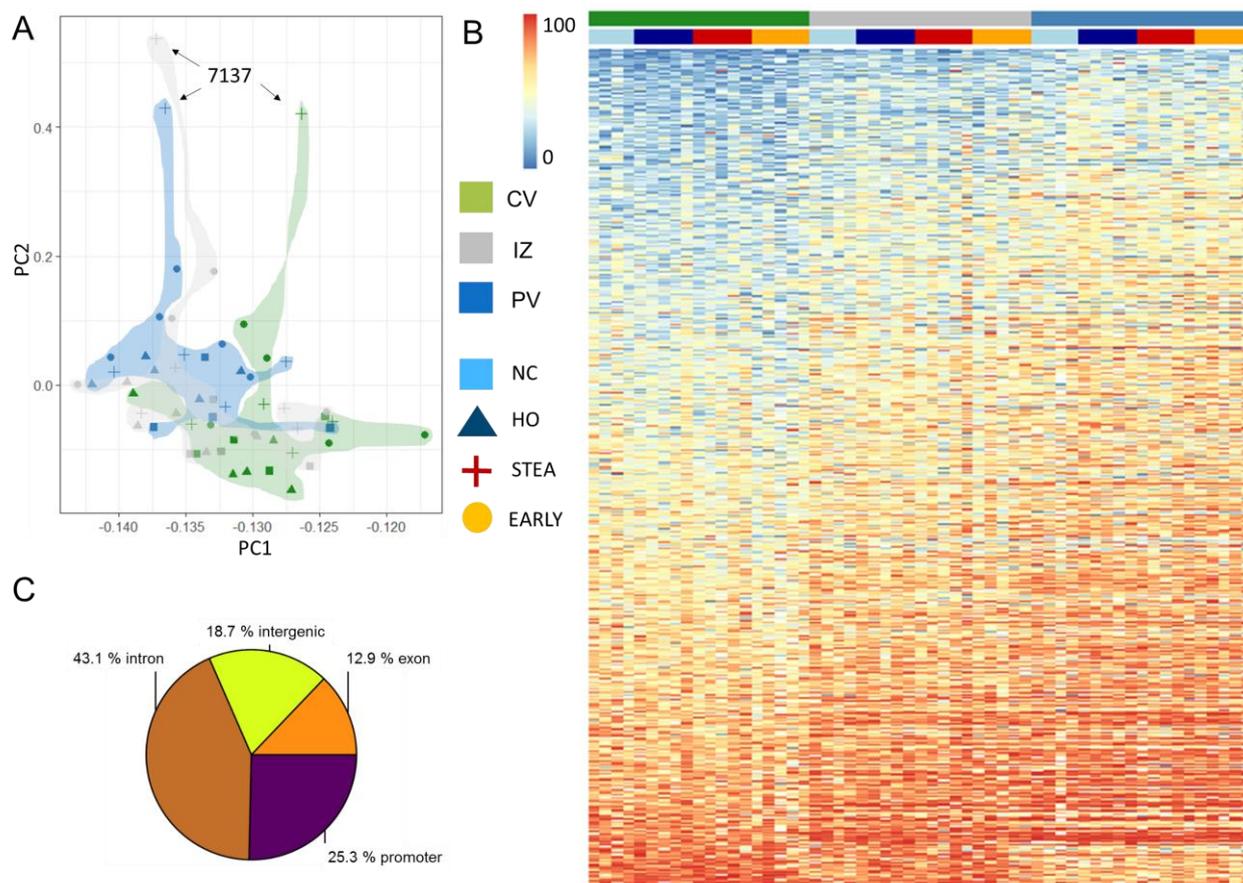
Figure 18: DNA methylation gradients along the porto-central axis detected by lcm-RRBS on microdissected hepatic zones of the human liver (Brosch et al., 2018). Samples (n = 19) are annotated as described in Figure 16. A – Principal component analysis of the 5000 most variable CpGs with coverage > 5 in all samples. B – Average methylation values at the top 1000 differentially methylated regions detected with methylkit (CpG coverage > 10, 500 bp tiles, at least 3 CpGs, FDR adjusted p-value < 0.01) between pericentral and periportal that were also covered in the intermediate zone. Average CpG methylation differences between CV and PP range between 12.4% and 49.2%. Blue corresponds to low and red to high average methylation values per DMR. C – Genomic annotation of all 17,862 DMRs between CV and PV.

AluI/HaeIII double digested RRBS libraries matching the RNA-seq dataset (chapter 3.1.1.1) were generated using adjacent cryosections to profile DNA methylation across hepatic zones. Libraries were at least sequenced to a mean CpG coverage of 5 (Sup. Table 1). Although less pronounced as for transcription, a PCA of the 5,000 most variable CpG sites showed separation by hepatic zone, with highly significant correlation of PC1 and zonation (r = 0.44, p = 6.7 x 10^{-4}, Sup. Figure 2 A), indicating an impact of zonation on methylation patterns. Covariate adjusted analysis of differentially methylated regions (DMRs) using MethylKit (coverage > 10, ≥ 3 CpGs per 500 bp tile, chapter 2.2.2.2) between the pericentral and periportal zone yielded 17,862 DMRs with at least 5 % methylation difference (FDR < 0.01). Strikingly, only 23 significant DMRs with the highest methylation difference in the intermediate zone were detected between PV or CV and IZ. The vast majority of DMRs exhibits a methylation gradient along the porto-central axis (Figure 18, B), mostly with a pericentral hypomethylation (70.36 % of detected DMRs). On the level of

partially methylated domains no significant methylation differences were detected between zones (data not shown). Certainly also attributed to the restriction motives of the chosen restriction enzymes, the majority of DMRs was found in a gene-context, and only 18.7 % of DMRs were located in intergenic regions (Figure 18, C).

3.1.1.3 Integration of zonal gene expression, DNA methylation and transcription factor binding

Particularly in case of pronounced anti-correlation between promoter CpG methylation and gene expression, there might be a decisive regulatory function of DNA methylation on expression levels. Moreover, intragenic or further distal regions featuring zonal DNA methylation patterns could impact gene expression in a more complex manner. To assess correlations between zonal gene expression and DNA methylation patterns, DMRs were annotated to their closest gene. DMRs related to differential expression (Figure 19 A) were predominantly located in introns (46.4 %), exons (18.7 %) and promoter regions (24.7 %, Figure 19 B). In total, 44.3 % (357) of zone-specific DEGs were associated with zonal DMRs (1094, Figure 19 A), in which 205 DEGs corresponded to multiple DMRs. Gene expression and associated DNA methylation differences were largely anti-correlated ($r = −0.37$, $p < 2.2 \times 10^{-16}$). In the pericentral zone 132 genes showed transcriptional upregulation corresponding to 430 hypomethylated DMRs, while 139 periportal DEGs were associated with 291 hypomethylated DMRs.

A representative example is the Wnt signaling pathway member Palmitoleoyl-Protein Carboxylesterase (NOTUM), which is exclusively expressed in pericentral hepatocytes and exhibits a cluster of six promoter DMRs with pronounced pericentral hypomethylation (Figure 19 C). Publicly available DNaseI-seq and ChIPseq data from the hepatocyte-like cell line HepG2 (ENCODE V3) indicate regulatory activity for these DMRs, suggesting a potential effect of DNA methylation in the regulation of zonal gene expression. Yet, we also detected 186 DEGs positively correlating with DNA methylation changes (Figure 19 A), and multiple DMRs associated with the same DEG did not necessarily behave similarly. A total of 99 DEGs exhibited mosaic DNA methylation patterns, suggesting differences in regulatory potential of DMRs, for example by affecting transcription factor (TF) binding dynamics.

Indeed, subsequent *in silico* analysis of TF binding motives revealed enrichment of distinct sets of TFs for negatively and positively DEG correlating DMRs (Figure 19 D). The most significant TFs for hypomethylated DMRs corresponding to upregulated expression include hepatic nuclear factor 4a (HNF4α), involved in the regulation of numerous hepatic genes (Odom et al., 2004), TCF7L2, a key factor downstream of the Wnt signaling pathway (Yi et al., 2005), and Foxo1, which regulates metabolic homeostasis in response to oxidative stress and promotes gluconeogenesis in hepatocytes (Barthel et al., 2005). In contrast, binding motives of ligand-dependent Aryl Hydrocarbon Receptor (AHR), which plays a major role in the regulation of drug metabolizing enzymes such as cytochrome P450 (Ramadoss et al., 2005), and the C/EBP family, essential for many liver functions like nutrient metabolism and liver regeneration (Takiguchi, 1998), were enriched in hypomethylated DMRs associated with reduced gene expression. Interestingly, TCF7L2, HNF4α and C/EBP were enriched only in DMRs hypomethylated in the pericentral zone (Sup. Table 3), while AHR and Foxo1 were more strongly enriched in periportal hepatocytes (Sup. Table 4), suggesting potential involvement in the establishment of zonal expression patterns.
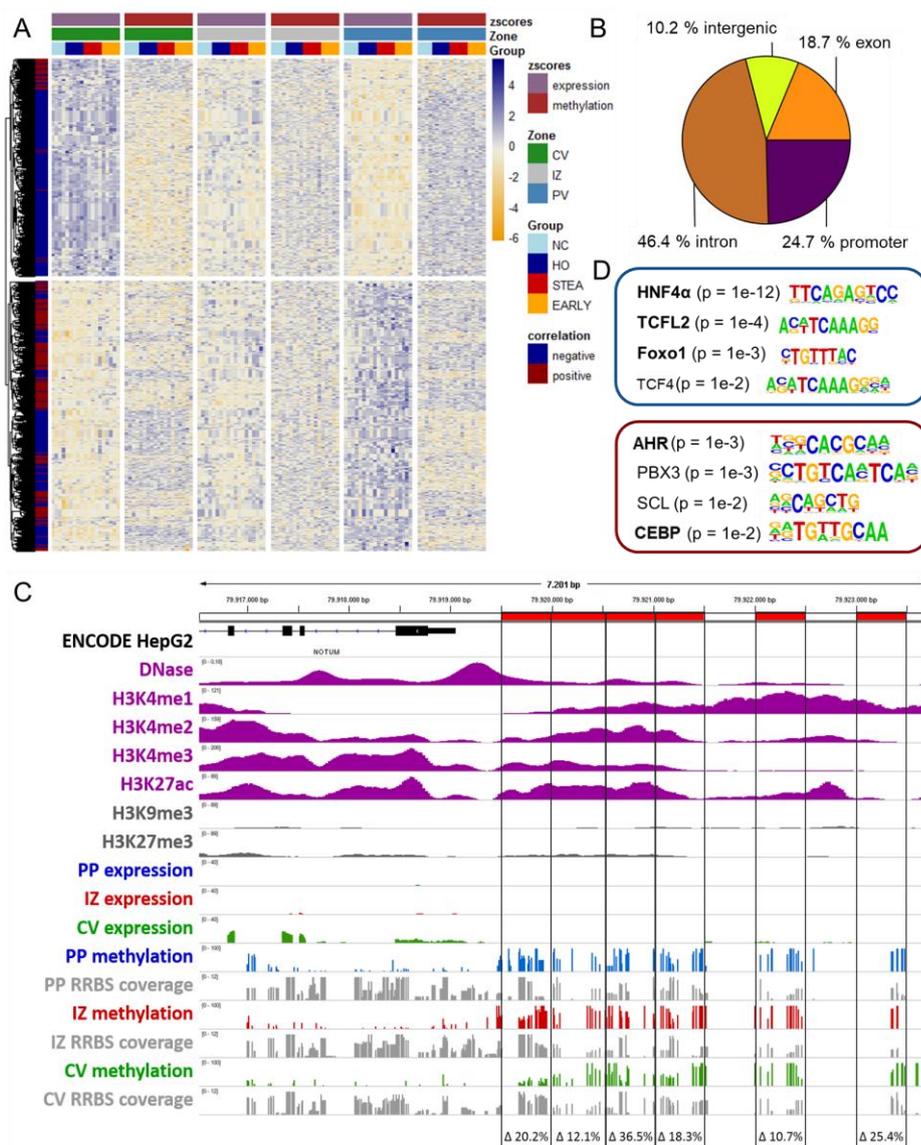
Figure 19: Integration of gene expression and DNA methylation profiles across hepatic zones. A – Visualization of gene expression (purple) and DNA methylation (dark red) *z*-scores of 1094 DEG-associated DMRs and corresponding genes at pericentral (green), intermediate (red) and periportal zone (blue). Yellow heatmap values corresponds to low *z*-scores (reduced expression or DNA methylation) and navy to high *z*-scores (increased expression or DNA methylation). Negative or positive Pearson correlation between expression and methylation z-scores are annotated in blue and red, respectively. B – Genomic annotation of DEG-associated DMRs (Brosch et al., 2018).C – Exemplary genome browser representation of expression and DNA methylation signatures across hepatic zones (Brosch et al., 2018). Expression of the Wnt signaling pathway member NOTUM is restricted to the pericentral zone. The 5 kb promoter region features 6 DMRs with methylation differences from 10.7 % to 36.5 % anti-correlated with gene expression. A representative healthy obese sample was used for visualization. Pericentral tracks are displayed in green, intermediate in red and periportal in blue. RRBS coverage tracks are colored in grey. Potential regulatory regions (purple) are illustrated by HepG2 ENCODE data for chromatin accessibility generated by DNase-seq and the histone modifications H3K4me1, H3K4me2, H3K4me3 and H3K27ac by ChIP-seq. DMRs are marked by black boxes with mean methylation difference between pericentral and periportal hepatocytes calculated from all samples (n = 19) considering only CpG sites with coverage > 10. D – *In silico* analysis of transcription factor binding motives in negatively (outlined in blue) and positively (red) expression correlated DMRs using HOMER. Presented are the 4 most significant TFs with p-value and respective binding motives. TFs printed in bold have known key functions in hepatic regulations.

Hence, together with Alexander Hermann and Prof. Dr. Jochen Hampe (TU Dresden), porto-central DNA methylation differences at experimentally verified TFBS of 59 TFs with publicly available ChIP-seq data for the hepatocyte-like cell line HepG2 (ENCODE V3) were analyzed to further substantiate these findings.
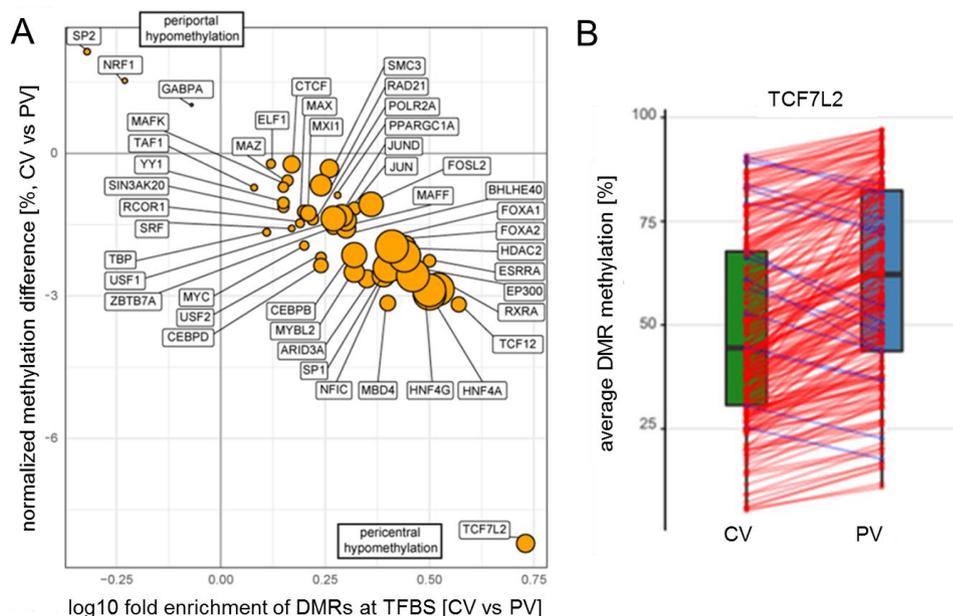
Figure 20: DNA methylation gradient of transcription factor binding sites (TFBS) along the porto-central hepatic axis (Brosch et al., 2018). A – Bubble plot of 46 TFs displaying significant ($p < 2.8 \times 10^{-4}$) DNA methylation gradients at their binding sites. The y-axis depicts the average methylation difference, normalized for the mild genome-wide CV hypomethylation (-1.27 %), between pericentral (CV) and periportal (PV) at DMR overlapping TFBS, with negative values corresponding to CV hypomethylation. The x-axis depicts the log10 fold enrichment of DMRs among TFBS. Bubble sizes correlate with log10 p-values, calculated by Fisher's exact test, with large bubbles referring to especially low p-values. Genome-wide TFBS are derived from HepG2 ChIP-seq data (ENCODE V3). B – Exemplary DNA methylation gradients of DMRs containing binding sites for TCF7L2. Plots are based on calculations of DMR enrichment and DNA methylation gradients at TFBS performed by Alexander Hermann. The detailed method is provided in our publication (Brosch et al., 2018).

Indeed, 46 TFs showed highly significant ($p < 2.8 \times 10^{-4}$) DNA methylation gradients between pericentral and periportal hepatocytes at their binding sites (Figure 21 A). Intriguingly, these TFs were uniformly expressed along the porto-central axis, but their target genes frequently showed zonal expression patterns, which might be attributed to differential signaling efficacy of respective TFs. As certain TFs may have differing binding preferences for methylated and unmethylated DNA (Spruijt and Vermeulen, 2014), presumably contributing to the establishment of TF binding dynamics, information regarding binding preferences of the 46 TFs showing DNA methylation gradients (Figure 21 A) was gathered from current literature (Sup. Table 5). For several TFs only *in silico* predicted data were reported, thus these findings were validated by analyzing average DNA methylation at respective TFBS in HepG2 using publicly available WGBS data (DEEP), indicating the same binding preferences as previously reported (Sup. Table 5). While most of the analyzed TFs preferentially bind unmethylated DNA, there are several TFs, such as RXRa and CEBPB, specifically binding methylated CpGs. Both RXRa and CEBPB binding sites are enriched in DMRs hypomethylated in pericentral hepatocytes, respectively hypermethylated in the periportal zone (Figure 20

A), thus suggesting an increased periportal binding efficiency. RXRa is a key factor for the activation of fatty acid oxidation (Hardwick et al., 2009), which is known to be localized in periportal hepatocytes (Figure 9). Likewise, periportal hypermethylation of CEBPB binding sites is in accordance with its involvement in hepatic liver regeneration (Jakobsen et al., 2013), which is probably driven by a periportal hepatocyte subpopulation (Font-Burgada et al., 2015). Contrary, TFs binding unmethylated DNA would be expected to be enriched in hypomethylated DMRs. For instance, TCF7L2, which actually features the most pronounced zonal methylation gradient (Figure 21, B) and is strongly enriched in hypomethylated DMRs correlating with upregulated gene expression (Figure 20 D and Figure 21 A), strongly indicates higher pericentral signaling efficiency. This is in line with the detected pericentral expression of alleged TCF7L2 target genes such as AXIN2, LGR5, CYP1A1 or OAT (Norton et al., 2014). Taken together, these observations indicate that DNA methylation gradients may elicit the zonated activity of otherwise uniformly expressed TFs by regulating differential binding efficiency.

3.1.1.4 Morphogen gradients regulating hepatic zonation
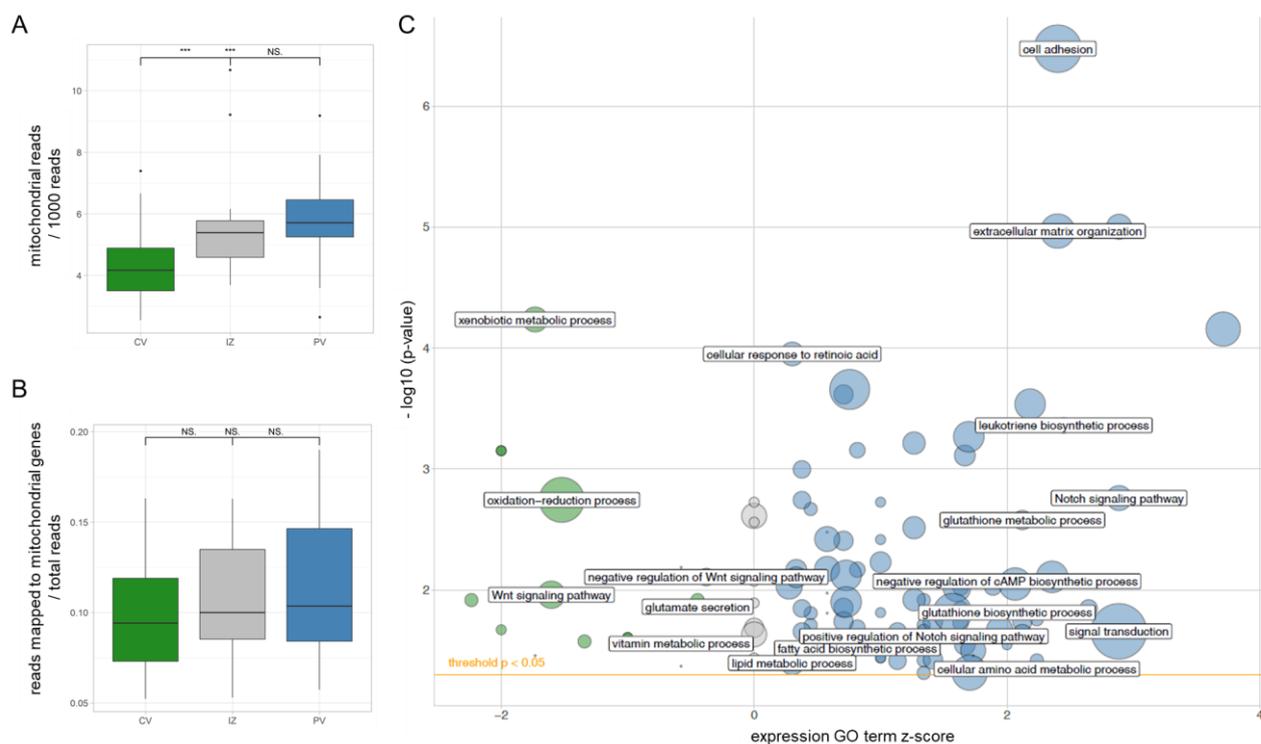
Figure 21: Mitochondrial, metabolic and morphogen zonation. A – Mitochondrial abundance gradient deduced from the fractions of mitochondrial reads in the RRBS data (n = 19) in pericentral (green), intermediate (grey) and periportal (blue) zones (Brosch et al., 2018). Significant differences between zones detected by Wilcoxon test are marked by asterisks (p < 0.001). B – Mitochondrial expression gradient identified by the ratio of lcm-RNA-seq reads mapping to mitochondrial genes. Differences between zones were not significant (NS). C – Visualization of significantly enriched GO terms (p < 0.05) of genes with zonated expression signatures (Brosch et al., 2018). Negatively scored GO terms (green) correspond to pericentral GO enrichment, positive scores reflect periportal GO terms (blue). The circle size represents the number of zonated genes related to the respective term. Exemplary GO terms are labelled. The full GO enrichment analysis is shown in Sup. Table 6 and 7.

Besides gradients of gene expression and DNA methylation, a significant mitochondrial gradient along the porto-central hepatic axis, deduced from the ratio of mitochondrial reads in the RRBS data, was observed (Figure 21 A). In line with the known oxygen gradient and resulting oxidative capacity, mitochondrial abundance is highest in the periportal zone and decreases towards the pericentral zone. A trend to zonation is also evident in mitochondrial gene expression (Figure 21 B) and likely reflects zonal metabolic activity (Figure 21 C), identified by gene ontology enrichment analysis of zonal expression and DNA methylation patterns (Sup. Table 6 and 7).

Moreover, zone-specific enrichment of signaling pathways was detected (Figure 21 C). In the periportal zone an accumulation of Notch signaling pathway related genes was observed (e.g. GO:0007219, p = 1.7×10^{-4}, Figure 23 A), while there was a strong enrichment of Wnt signaling among the pericentrally upregulated genes (e.g. GO:0016055, p = 5.7×10^{-4}, Figure 22 A). Notch and Wnt signaling pathways are of major interest regarding regulation of hepatic zonation as well as liver regeneration. Both on the transcriptional and epigenetic level, networks of key pathway members showed zonal differences (Figure 22 A, Figure 23 A), while other morphogen pathways relevant for hepatocyte functions, such as FGF, TGF beta, HGF and BMP were neither epigenetically nor transcriptionally zonated (FDR < 0.01).
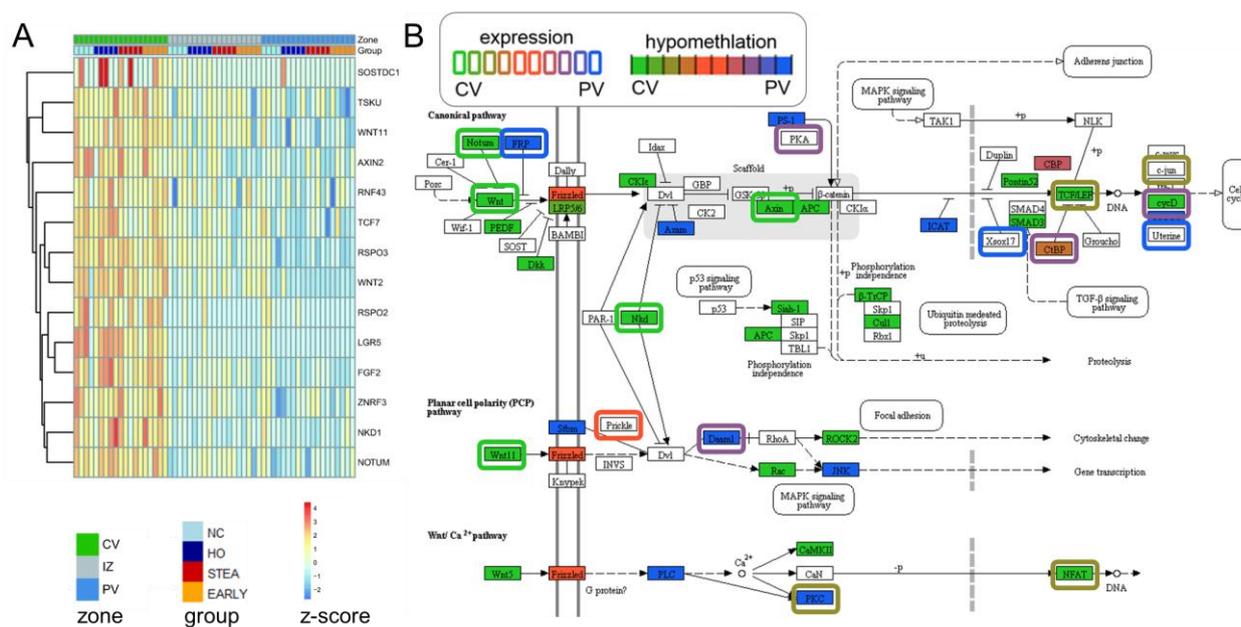
Figure 22: Wnt signaling pathway. A – Expression and DNA hypomethylation signature of the Wnt signaling KEGG pathway. Genes colored in green are associated with pericentrally hypomethylated DMRs, while blue corresponds to hypomethylated DMRs in the periportal zone. Accordingly, genes framed in green are pericentrally expressed and blue frames illustrate periportal expression. B – Expression z-scores of Wnt signaling pathway genes (GO:0016055, GO:0090263 and GO:0030178) along the porto-central axis (Brosch et al., 2018). The annotation legend displays hepatic zones and phenotypes as described in Figure 16.

Classical Wnt agonists such as WNT3A or WNT3 as well as genes of the export machinery like PORCN1 or WLS do not show zonated expression profiles. Instead, Wnt signaling enhancers RSPO2, RSPO3 and LGR5 are almost exclusively expressed in pericentral hepatocytes, and negative Wnt regulators including AXIN2, NOTUM, NKD1, RNF43 and ZNRF3 are also strongly upregulated in the pericentral zone (Figure 22 B). Most zonal expression differences are accompanied by matching methylation changes (Figure 24 A),

such as a pericentral hypomethylation of DMRs associated to AXIN2 (Δmeth = 5.0–29.3%), NOTUM (Δmeth = 10.7–36.5%, Figure 19 C) and LGR5 (Δmeth = 9.2%). Yet, several Wnt signaling pathway members, which do not exhibit zonal expression profiles, show still zonated DNA methylation. Moreover, binding sites of the key downstream effector of Wnt signaling TCF7L2 show a more than 5 fold enrichment at pericentral DMRs (Figure 20 B). Taken together, these gradients of zonated Wnt signaling members suggest the sensitization of pericentral hepatocytes for canonical Wnt signals in a tightly controlled signaling environment, presumably representing a key regulatory mechanism for the establishment and maintenance of hepatic zonation.
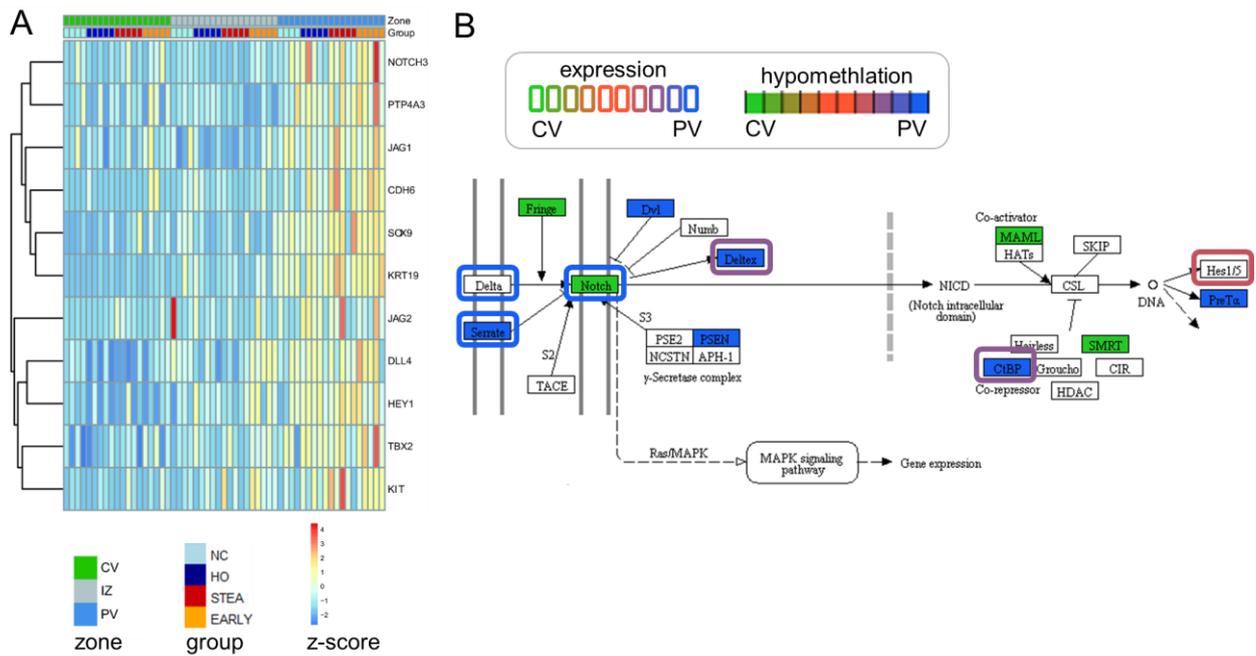
Figure 23: Notch signaling pathway. A – Expression and DNA hypomethylation signature of the Notch signaling KEGG pathway annotated as described in Figure 23. B – Expression z-scores of Notch signaling pathway genes (GO:0007219) along the porto-central axis (Brosch et al., 2018).

Interestingly, LGR5 (logFC = 7.15, FDR = 4.88 × 10^{-26}), which shows the strongest pericentral zonation profile, and AXIN2 (logFC = 2.24, FDR = 6.17 × 10^{-5}) expression are strongly correlated with expression of the fetal liver progenitor marker TBX3 (pairwise expression correlation to AXIN2: r = 0.51, p = 5.36 × 10^{-5} and LGR5: r = 0.86, p < 2.2 × 10^{-16}, Figure 24 A). The genomic landscape associated to TBX3 displays 11 pericentrally hypomethylated DMRs (Δmeth = 6.9–43.6%), reflecting the highly significant pericentral upregulation of TBX3 expression (logFC = 2.26, FDR = 4.1 × 10^{-59}, Figure 24 A). These observations are in line with previous reports in mouse, suggesting AXIN2 and LGR5 positive hepatocytes as potential source of homeostatic regeneration (Wang et al., 2015).

Periportal Notch signaling was also apparent on transcriptional and epigenetic level (Figure 23 A and B). For example, NOTCH ligand JAG1 (logFC = 2.03, FDR = 1.64 × 10−5, Δmeth = 5.3 to 10.0%) displayed pronounced upregulation that correlates with hypomethylation of associated DMRs in periportal hepatocytes (Figure 24 A). Moreover, the liver stem cell marker EPCAM, potentially marking a periportal hepatic progenitor cell niche (Yoon et al., 2011), showed a particularly strong periportal expression (logFC = 5.42, FDR = 1.5 × 10−9), even though no zonal DNA methylation differences were detected (Figure 24 A).
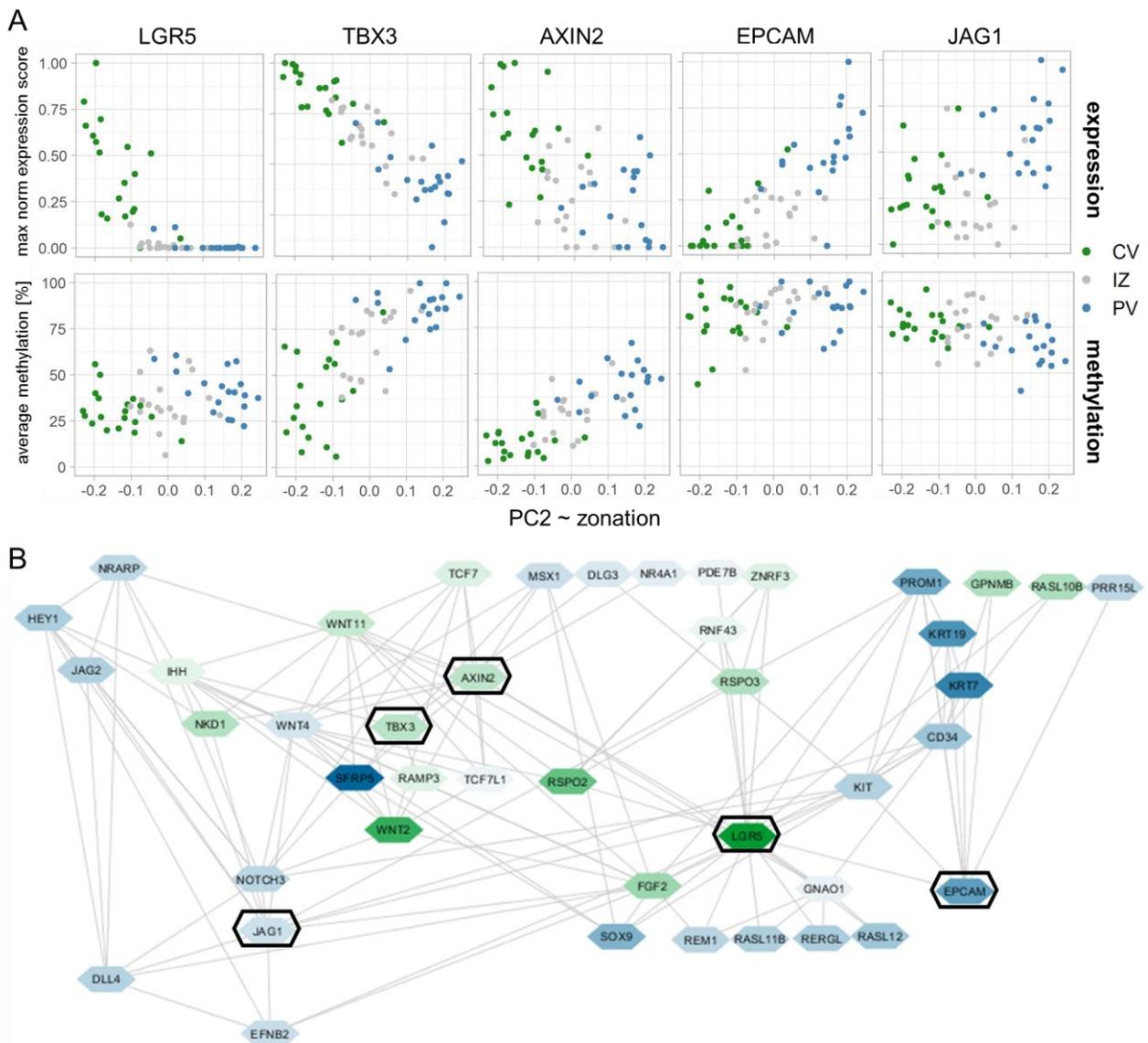
Figure 24: Morphogen gradients and resulting regulatory networks along the porto-central axis. A – Transcriptional and DNA methylation profiles of LGR5, AXIN2, TBX3, EPCAM and JAG1, projected against principal component 2 as a measure for zonation (Figure 16, A). For better visualization the PC2 axis was inverted. Transcription values are depicted as maximum normalized scores. Methylation (%) refers to the DMR with the strongest absolute average methylation difference between pericentral and periportal. Green corresponds to pericentral (CV), grey to intermediate (IZ) and blue to periportal (PV) samples. B – String interaction network of LGR5, AXIN2, TBX3, EPCAM and JAG1 and their direct interaction neighbors with zonated transcription. The color gradient refers to zonal expression profiles with green illustrating predominant pericentral expression and blue corresponding to periportal upregulation. Key morphogens are highlighted by black boxes.

Integrating zonal expression of Wnt and Notch signaling pathway members in a network based on known protein interactions (Figure 24 B) suggests an intricate interplay of both pathways and underlines their presumable importance for the maintenance of liver zonation in the human.

In summary, the zonal expression and DNA methylation analysis of liver samples with normal histology as well as with bland steatosis and early steatohepatitis revealed pronounced, largely anti-correlated transcription and methylation gradients along the hepatic porto-central axis. While donors with early non-

alcoholic fatty liver disease show consistent gene expression differences corresponding to the severity of the disease across all zones, the relative zonated gene expression and DNA methylation patterns remain unchanged. Prominent gradients of DNA methylation at binding sites of uniformly expressed transcription factors suggest a zonally differential binding efficiency of those TFs, representing an epigenetic layer in the regulation of hepatic zonation. Moreover, zonation patterns are probably established and maintained by the interplay of Wnt and Notch morphogen signaling pathways, which might also play a role in hepatic regeneration. The pericentral epigenetic and transcriptional Wnt signature supports the concept of a pericentral hepatocyte regeneration pathway under steady-state conditions, while a periportal Notch signaling might be indicative of a stem cell niche activated by chronic liver injury. To further substantiate these findings, the second part of this project (V0.2) is dedicated to the analysis of more progressed NASH and cirrhosis samples, which might contribute to the better understanding of Wnt and Notch signaling pathways in chronic liver injury and regeneration.

3.1.2 Hepatic zonation in progressed NASH, fibrosis and alcoholic cirrhosis

In an attempt to unravel hepatic zonation in progressed NASH and alcoholic cirrhosis, the dataset (V0.2) was extended with zonal samples from donors with inflammatory steatohepatitis (NASH), steatohepatitis with grade 3 fibrosis (FIB3), and from patients with severe alcoholic cirrhosis (ACI). Moreover, additional replicates of controls and earlier NAFLD stages were included to obtain higher statistical power (Figure 15). Samples from more progressed disease stages turned out to be rather challenging regarding the preparation of RRBS sequencing libraries. Filtering of low quality samples yielded in total 139 zonal transcriptomes (Sup. Table 2) and 106 DNA methylomes (Sup. Table 1), resulting in 104 matching samples with high data quality. Differential gene expression analysis was based on the full transcriptome dataset (n = 139) to ensure a sufficient number of replicates. Integrative epigenomic analysis was then performed on the core set of 104 matching samples.
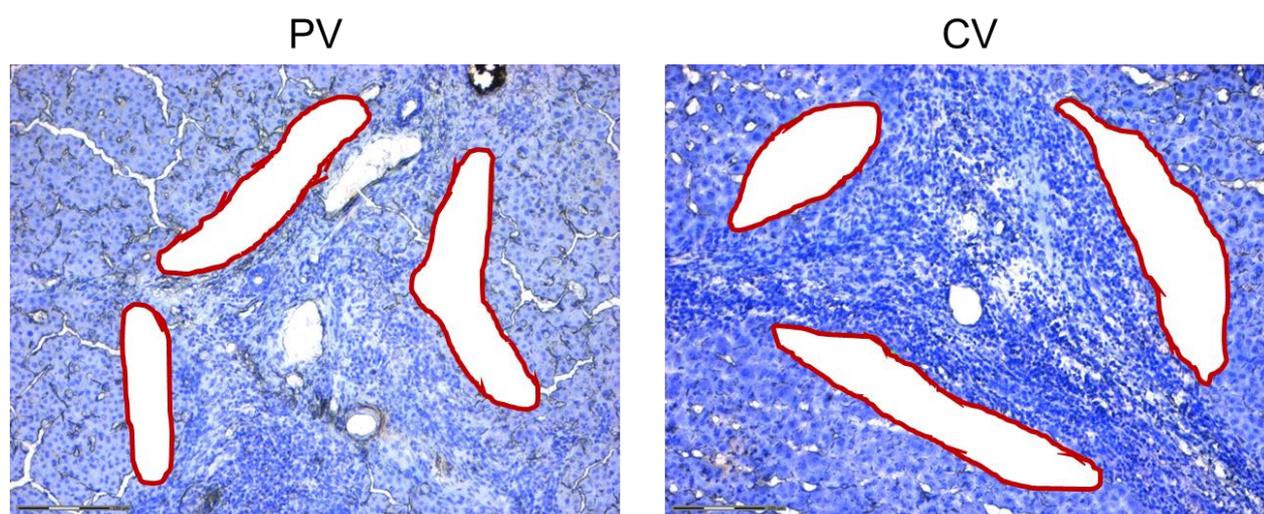
Figure 25: Exemplary histological Cresyl Violet staining of a periportal (PV) and pericentral (CV) region in cirrhotic liver tissue. Microdissected regions are marked in red. Images were kindly provided by Dr. Mario Brosch and Fabian Reichel.

It should be noted that fibrotic and cirrhotic liver tissues exhibit fibrous septa, which consist out of type I collagenous fibrotic material and are infiltrated with different cell-types such as fibroblasts and immune cells. In fatty liver diseases, they are accumulated around the central vein, but can also bridge across the hepatic lobule in progressed stages (Younossi et al., 2018). To ensure the analysis of hepatocyte signatures, the fibrotic material itself was excluded from microdissection. Thus, only hepatocytes bordering fibrous septa close to central vein (CV) or portal field (PV) were sampled for NGS library preparation of FIB3 and ACI samples (Figure 25).

Potential contamination with non-parenchymal cells was again estimated based on expression levels of known cell-type marker genes (Sup. Figure 1 A, V0.2). Most NPC markers showed comparably low expression levels as in the first round of data generation (V0.1). However, higher expression of the macrophage marker CD68 was observed, which was apparently attributed to a random increase of macrophage contamination in few libraries comprising NC, EARLY and ACI samples (Sup. Figure 1 B). In addition, expression levels of cholangiocyte markers, and to a lesser extent also fibroblast markers, were slightly elevated in cirrhotic samples, indicating a higher amount of NPC contamination in these samples

(Sup. Figure 1 C). On the other hand, during severe liver injury hepatocytes have the potential to transdifferentiate into cholangiocytes (Michalopoulos et al., 2005), which may explain the observed cholangiocyte signatures in the microdissected hepatocytes. Nevertheless, marker gene expression levels were still in an acceptable low range, allowing the analysis of predominant hepatocyte signatures across all phenotypes in the dataset.

Principal Component Analysis (PCA) of the 1000 most variable transcripts in the full dataset (V0.2) revealed a pronounced separation of cirrhosis samples from controls and NAFLD samples (Figure 26 A). While in controls and NAFLD clustering by hepatic zone was still perceivable, cirrhosis samples did not show any separation by zonation. Thus, in this setting (V0.2), the phenotype was identified as a major driver of transcriptional variance (Sup. Figure 2 B). To obtain a more precise impression regarding hepatic zonation in individual phenotypic groups, separate group-wise transcriptome PCAs were performed (Sup. Figure 4). Indeed, principle components of controls (NC, HO) and NAFLD samples (STEA, EARLY, NASH) showed significant correlations with hepatic zones, although correlations were already reduced in inflammatory NASH. Contrary, principle components of fibrotic NASH (FIB3) and cirrhosis samples (ACI) correlated only with donors, indicating a deregulation of hepatic zonation patterns in progressed fatty liver disease (Sup. Figure 4).
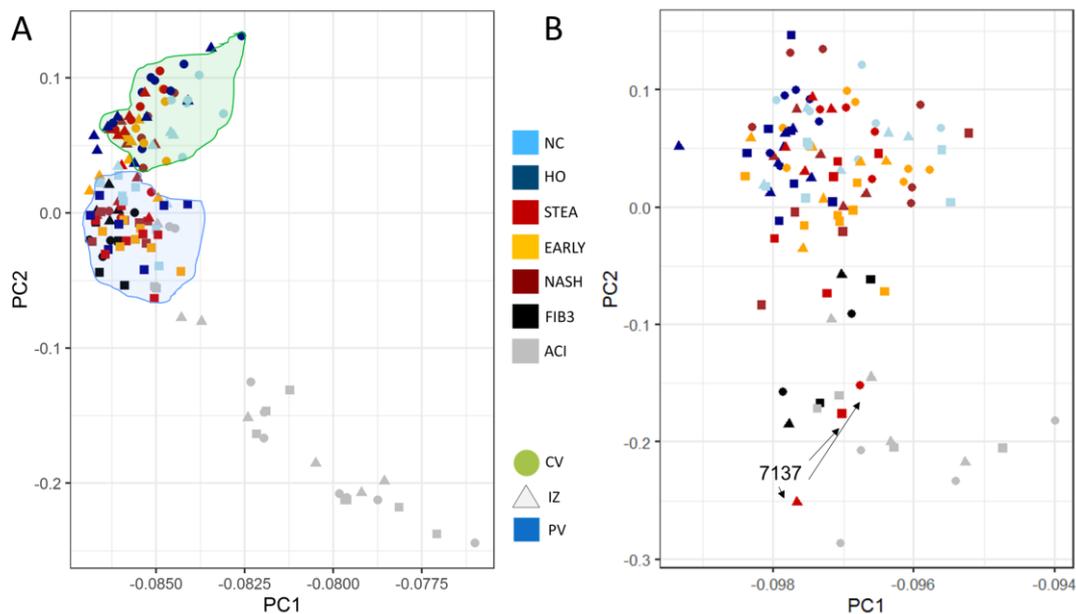
Figure 26: Principal Component Analysis of the 1000 most variable transcripts (A, n = 139) respective 3,621 CpGs with coverage above 5 in all samples (B, n = 106) detected in zonal hepatocytes isolated by LCM from human donors spanning the whole spectrum of fatty liver disease from healthy controls to cirrhosis. Hepatic zones (CV, IZ and PV) are depicted as symbols. The annotation legend displays controls (NC) in light blue, healthy obese (HO) in dark blue, non-alcoholic steatosis (STEA) in red, early NASH (EARLY) in orange, progressed NASH (NASH) in dark red, fibrosis (FIB3) in black and alcoholic cirrhosis (ACI) in grey. In panel B a steatotic outlier (donor 7137) is marked by arrows.

As PCA only considers data points present in all samples, it was restricted to the 3,621 CpGs with coverage above 5 in all 106 samples of the DNA methylation dataset. It showed a separation of fibrosis and cirrhosis samples from controls and NAFLD samples without obvious relation to hepatic zone (Figure 26 B). The first three PCs of the DNA methylation dataset did not show any correlation with hepatic zone, but rather with

phenotype and less pronounced with donor, age, HOMA index as indicator for type 2 diabetes, and also slightly with CpG coverage and batch (Sup. Figure 2 B). Due to experimental setup, batches (V0.1 and V0.2) themselves correlated with phenotype (r = 0.531, p = 1.8 x 10^{-11}). As correlations between PCs and batches were considerably lower, there are probably no major batch effects, both in transcriptome and methylome datasets (Sup. Figure 2 B). Thus, no batch corrections were performed prior to more detailed data analysis. It should be noted that steatotic donor 7137, which was already found to be an outlier in the initial methylome analysis (Figure 16 A), clustered with fibrotic and cirrhotic samples. Based on liver pathology and clinical data (Table 2), this sample was clearly defined as bland steatotic. Interestingly, even though there were no signs of inflammation, it exhibited a mild fibrosis (grade 1). However, several EARLY and NASH samples also featured a fibrosis grade 1 and still did not cluster with the more progressed fibrotic samples (grade 3). Thus, clustering of the outlier 7137 is probably not explained by fibrosis grade. Taken together, both on transcriptional and epigenetic level, PCAs suggest that progressed stages of fatty liver disease contribute more to variability than hepatic zonation and further indicate a fibrosis associated disruption of hepatic zonation profiles.
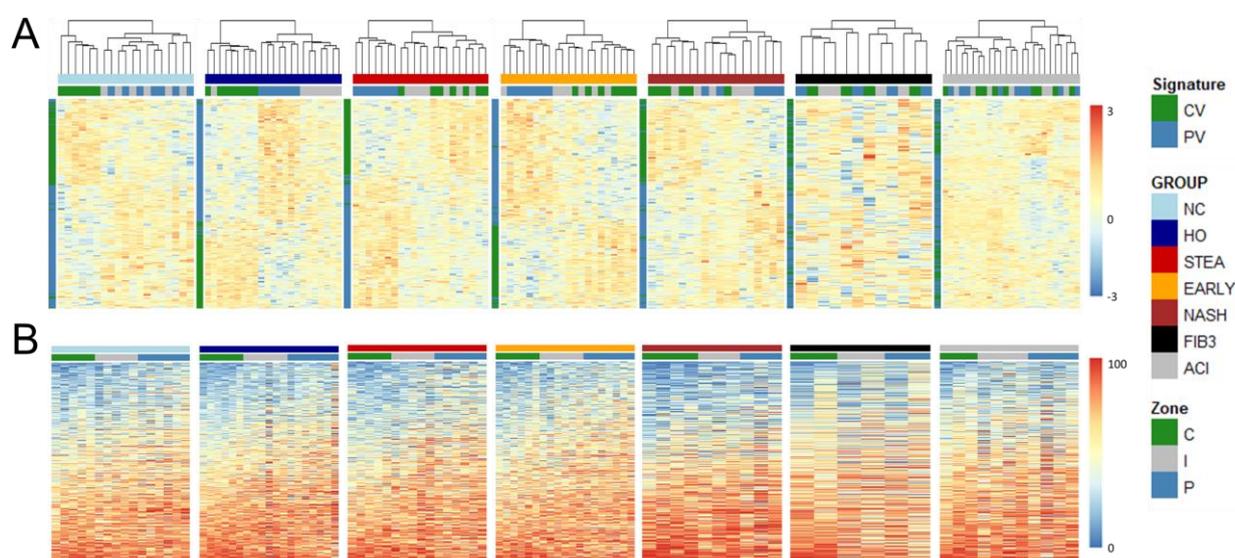
Figure 27: Group-wise expression and DNA methylation heatmaps of genes (A) and DMRs (B) zonated in normal liver tissue. A – Clustered expression z-score heatmap (manhattan clustering distance). The row annotation (Signature) relates to predominant pericentral (CV: green) or periportal (PV: blue) expression in healthy liver. B – The heatmap shows average methylation of the 1000 most variable zonal DMRs (500 bp tiles) and is sorted by average pericentral methylation. Columns are sorted by hepatic zone. DMRs not covered in all samples of the displayed group are excluded.

Prior to more detailed analysis of progressed fatty liver disease it should be noted that differential gene expression analysis between pericentral and periportal samples of the combined dataset (V0.1 and V0.2), excluding FIB3 and ACI samples to prevent disease associated variability, yielded due to higher statistical power a total of 1,073 zonated genes. Fold changes were highly correlated with the original analysis of dataset V0.1 (r = 0.953, p < 2.2 x 10^{-16}). To ensure consistency with previous findings, the original set of zonally expressed genes (Brosch et al., 2018) was used for further analysis of hepatic zonation in progressed disease stages.
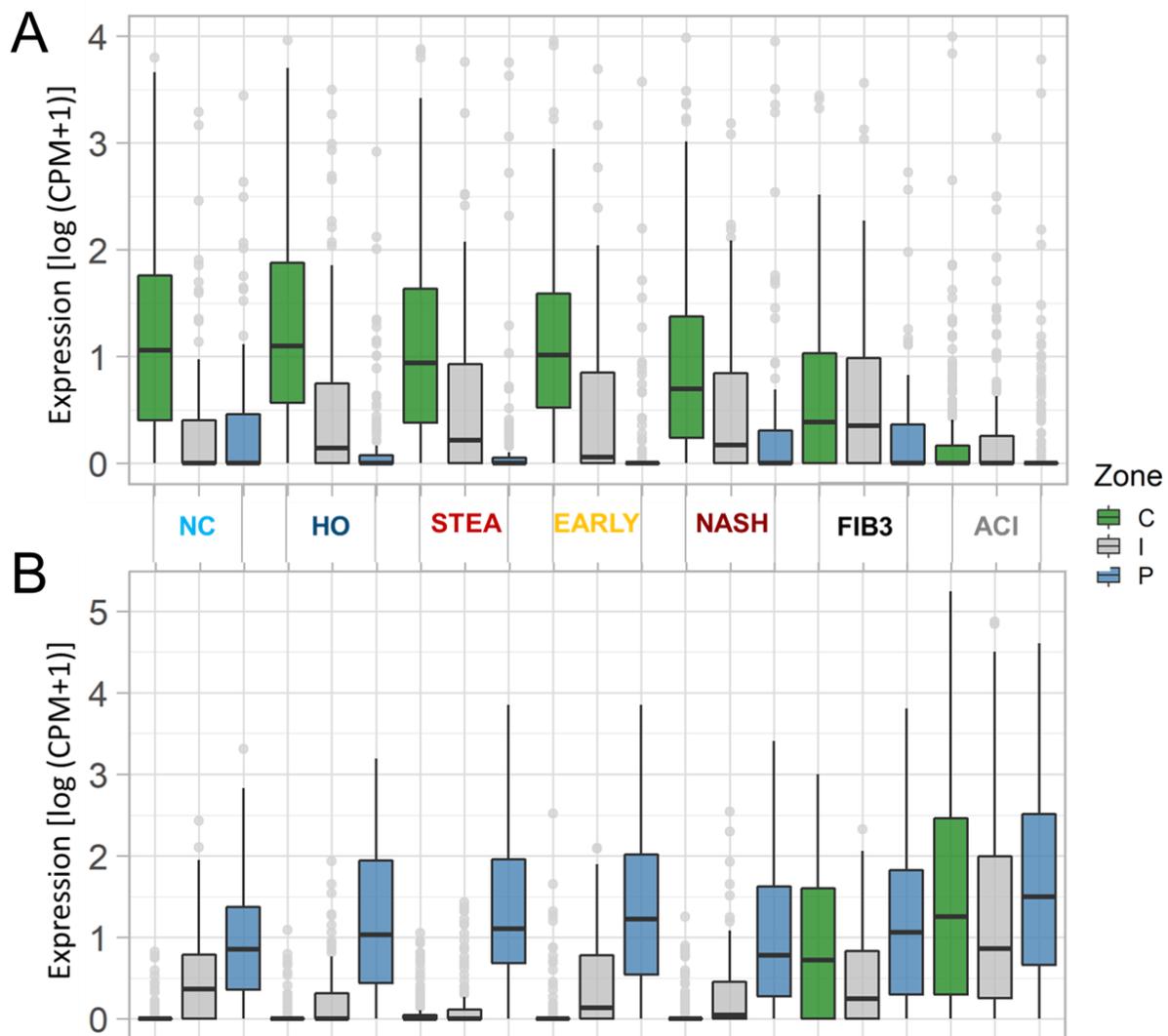
Figure 28: Zone-specific expression levels of the top 20 pericentrally (A) and periportally (B) zonated genes during fatty liver disease progression from normal controls (NC) to cirrhosis (ACI). Expression levels are displayed as log (CPM + 1). The analysis is based on the zonated gene list described in chapter 3.1.1 (805 genes).

To obtain a first impression regarding the impact of fatty liver disease progression on hepatic zonation signatures, expression patterns of the genes previously detected as zonated in normal liver tissue (V0.1), as well as DNA methylation patterns at zonated DMRs (V0.1), were visualized for individual phenotypes (Figure 27 A and B). Expectedly, this analysis confirmed stable expression and DNA methylation zonation patterns in steatosis and early steatohepatitis. Intriguingly, it also suggests a progressive loss of transcriptional and epigenetic zonation signatures in fibrosis and cirrhosis. In this context, especially genes with pronounced expression differences between hepatic zones exhibit considerably deregulated expression across zones in late disease stages (Figure 28). From late NASH to fibrosis and most pronounced in alcoholic cirrhosis a progressive loss of pericentrally expressed genes was detected (Figure 28 A). Simultaneously, an upregulation of periportal genes in these pericentral hepatocytes was observed (Figure 28 B). Though less pronounced these effects were also found when assessing all 806 zonated genes (Sup. Figure 5). Taken together, these observations gave rise to the hypothesis that during late stages of

fatty liver disease pericentral hepatocytes might become more similar to periportal hepatocytes. This phenomenon was termed *portalization* of pericentral hepatocytes.
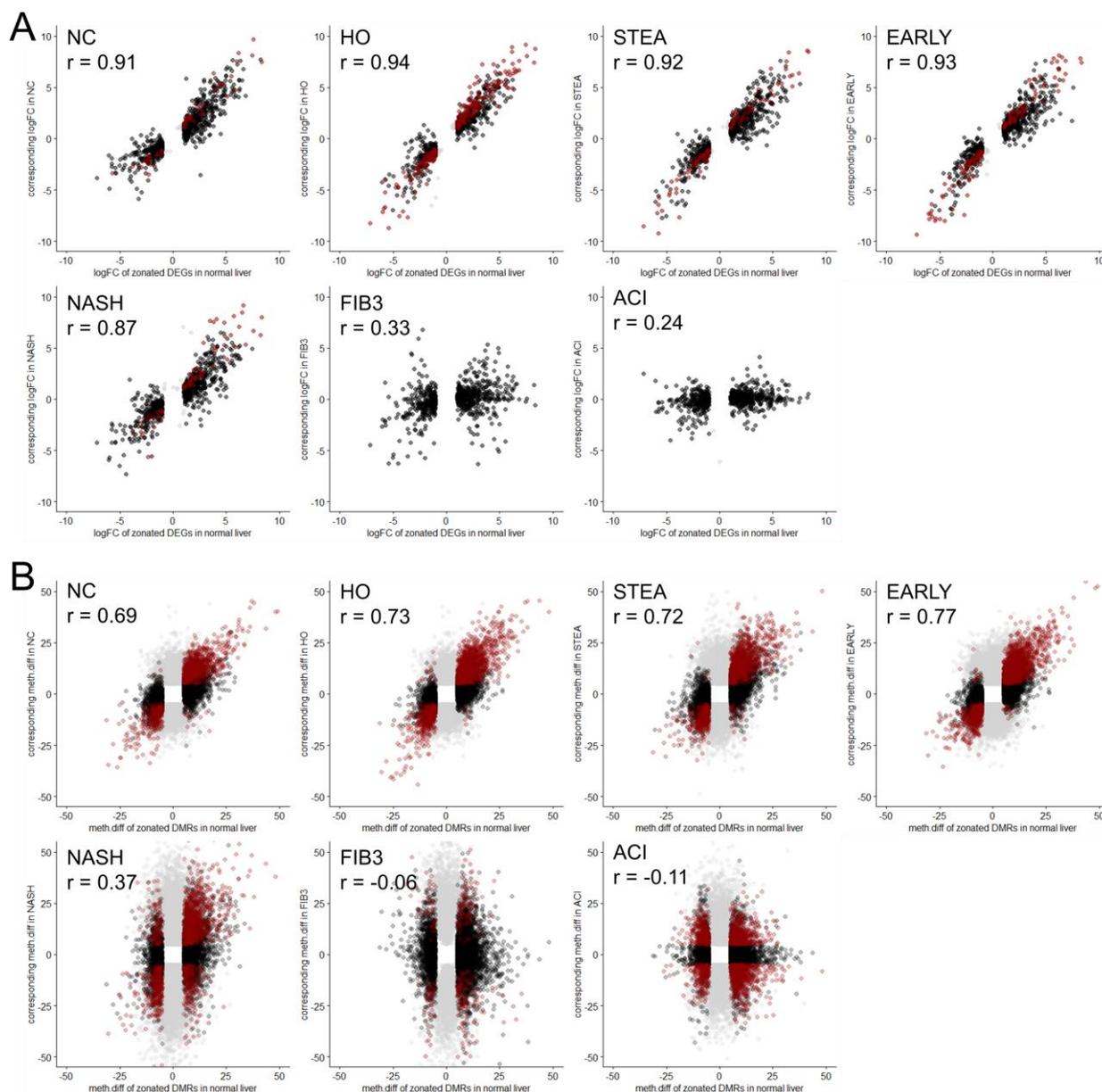
Figure 29: Pearson correlation based comparison of zonated gene expression (A) and DNA methylation (B) between fatty liver disease stages and healthy liver tissue. Displayed Pearson correlation coefficients (r, p < 8.2 x 10^{-6}) were calculated for logFC of genes zonated in healthy liver. DEGs and DMRs colored in red were detected as significantly zonated both in healthy liver and the displayed phenotypic group, while those only zonated in the original healthy liver tissue analysis (V0.1) are colored in black. In addition, zonated DEGs and DMRs detected uniquely in the phenotype-specific analysis between pericentral and periportal hepatocytes are illustrated in grey.

To provide more systematical evidence for the progressive loss of transcriptional zonation, zonal expression log fold changes in normal liver (V0.1) were correlated with those detected individually in each phenotypic group (Figure 29 A). While controls (NC, HO) and early stages of NAFLD (STEA, EARLY) showed high correlations (r = 0.91 – r = 0.94) with the original zonation analysis, correlations decreased from NASH (r =

0.87) to FIB3 (r = 0.33) and further to ACI (r = 0.24). In NASH, 75 genes with zonated expression between CV and PV hepatocytes were detected, while in FIB3 and ACI no zonal gene expression was observed anymore.

Likewise, zonal DMR methylation patterns of individual phenotypes featured high correlations with normal liver (V0.1) in controls and early NAFLD (r = 0.69 – r = 0.77) and decreased correlation in progressed NASH (r = 0.37), while in FIB3 (r = -0.06) and ACI (r = -0.11) zonated DNA methylation profiles did not correlate with epigenetic zonation in normal liver (Figure 29 B). Contrary to transcriptional zonation, also in late disease stages a considerable amount of significantly zonated DMRs was shared with normal liver (NASH = 1,317, FIB3 = 292, ACI = 2,026). Interestingly, in NASH 21.0 %, in FIB3 54.6 % and in ACI 58.2 % of these shared DMRs showed zonation profiles opposed to normal liver. Pericentral hypomethylation in healthy tissue turned into periportal hypomethylation in progressed disease, and vice versa, which indicates a pronounced epigenetic remodeling of zonation profiles during the progression of fatty liver disease. Moreover, in all analyzed phenotypes additional zonated DMRs were detected (NC = 5,797, HO = 5,797, STEA = 8,653m EARLY = 14,578, NASH = 29,156, FIB3 = 4,121, ACI = 10,139), which were not found in the original zonation analysis (V0.1). These DMRs showed low overlap between samples (9 % - 25 %), which is only partially explained by variability of CpG coverage in the RRBS libraries. Furthermore, it should be noted that the phenotype-specific zonal DMRs might represent low confidence DNA methylation changes. However, the standard deviations of these DMRs were not higher than of the DMRs shared between phenotypes. As a rather limited amount of replicates was available for DNA methylation analysis (n = 2 – n = 7, Figure 31) and DNA methylation patterns in human are subject to inter-individual variability (Gunasekara et al., 2019), increasing replicate numbers and also higher sequencing depth might lead to a better overlap of zonal DMRs between phenotypes. In summary, the analysis of zonation signatures in normal liver in comparison with early and late stages of fatty liver disease reveals a progressive loss of transcriptional zonation, together with a pronounced deregulation of zonal DNA methylation patterns.

In addition, zone-wise differential analysis was performed between disease phenotypes and samples with normal liver histology in order to assess also phenotype-specific differences not necessarily directly related to hepatic zonation patterns. Normal lean controls (NC) were mostly obtained from colon cancer patients. In contrast, besides from being severely obese, healthy obese donors (HO) did not suffer from any diagnosed disease. Both NC and HO donors showed normal liver histology, and clinical parameters such as ALT, AST (alanine- and aspartate aminotransferase) and ᵧGT (γ-glutamyltransferase) were in a normal range. Differential expression analysis between HO and NC revealed in total 102 DEGs in any of the three zones. The majority of these genes did not exhibit zonated expression profiles (89.2 %), and for zonated genes (11) such as SAA1 the relative zonation profiles across the hepatic lobule were stable between the two groups (Sup. Figure 6 D). Significantly enriched GO terms (p < 0.05) include liver function associated terms such as xenobiotic metabolic process, canonical glycolysis, steroid metabolic process and response to nutrient, but also the inflammatory related terms acute-phase response and immune response. These observations confirm previous findings, reporting low-grade upregulation of inflammation associated genes in liver tissue of obese patients whit normal liver histology (Bertola et al., 2010). However, in NC several genes such as ZNF667-AS1, FGL1 and TM4SF5, which were implicated to play a role in tumor and metastasis development (Chen et al., 2018; Lee et al., 2010; Zhang et al., 2018), were upregulated in comparison to HO. Given that

NC samples were mostly derived from colon cancer patients, and up to 35 % of them develop liver metastasis (Valderrama-Trevino et al., 2017), an impact of colon cancer on normal liver homeostasis cannot be fully excluded, even though in NC donors no hepatic metastasis were detected yet.
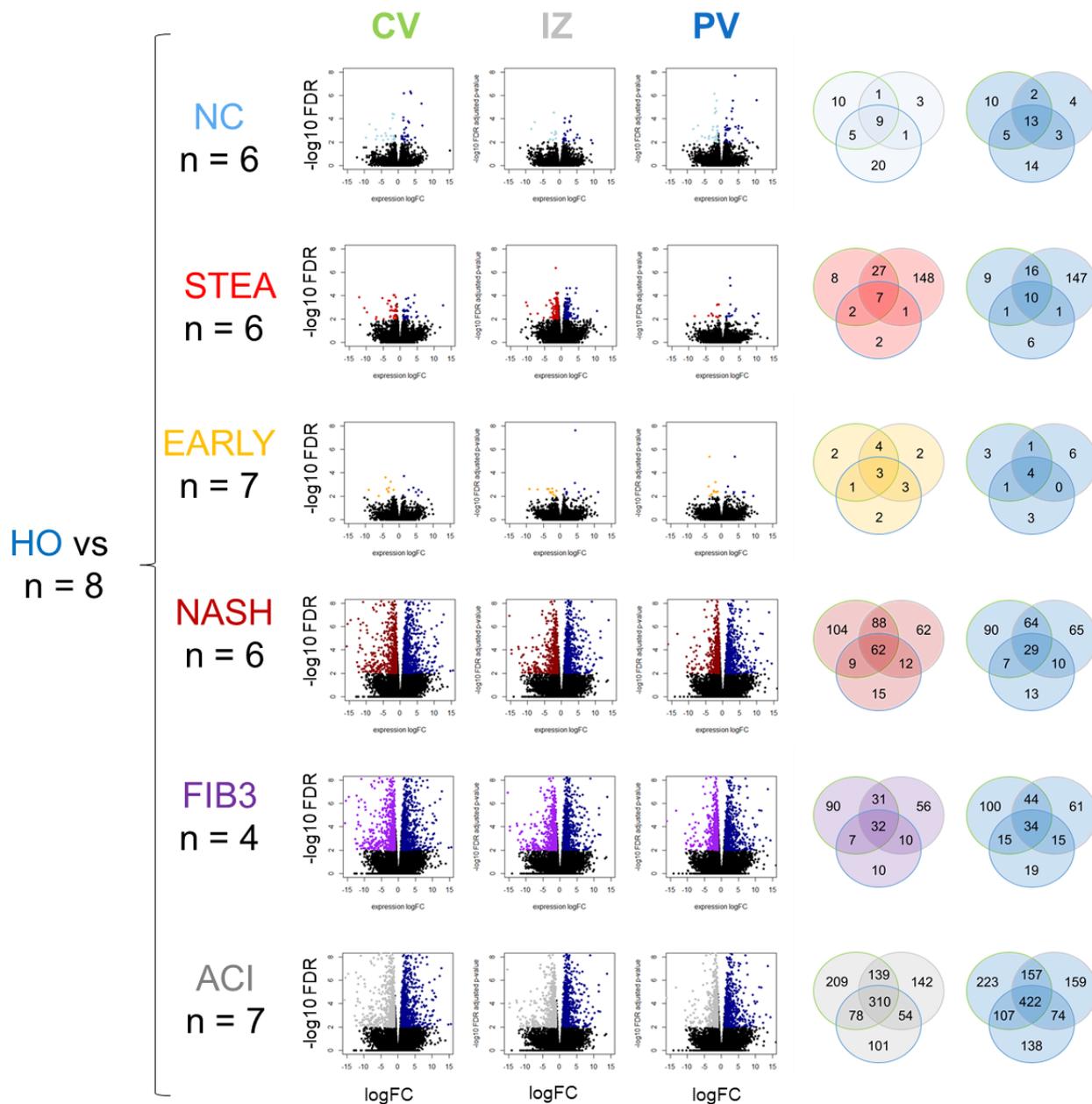
Figure 30: Volcano plots of zone-specific differential expression analysis between HO and the other phenotypic groups (NC, STEA, EARLY, NASH, FIB3 and ACI). The x-axis visualizes expression log₂ fold changes (logFC), while the y-axis depicts negative log10 of FDR adjusted p-values. Significantly upregulated (|logFC|>1, FDR < 0.01) genes and respective Venn diagrams are colored by corresponding group (NC = light blue, HO = dark blue, STEA = red, EARLY = orange, NASH = dark red, FIB3 = purple, ACI = grey). Venn diagrams show the overlap of DEGs detected in pericentral (CV, green line), intermediate (IZ, grey line) and periportal (PV, blue line) zones.

Thus, both NC and HO do not represent an ideal healthy state. However, due to ethical reasons, it is not possible to obtain surgical liver samples from human donors without any medical necessity. As the impact of obesity on hepatic gene expression appears to be rather small and all analyzed NAFLD donors were

obese as well, HO was considered as best control group for zone-wise differential analysis of disease phenotypes (Figure 30 and 31). Yet, it should be noted that ACI samples were derived from lean patients, which could potentially bias the analysis. Thus, ACI samples were also compared to NC as control group, which yielded highly similar $\log_2$ fold changes of genes significantly deregulated in ACI (r = 0.937, p < 2.2e^{-16}), suggesting a negligible impact of obesity associated gene expression differences in the comparison of histologically normal liver with cirrhotic tissue. Hence, further analysis focused on HO based differential analysis of disease phenotypes (Figure 30 and 31). As this analysis consists of a complex set of multiple comparisons, first an overview of phenotype-specific transcriptome and methylome changes is described, which is then followed by a more detailed assessment of individual phenotypes.

Comparing the full range of fatty liver disease including alcoholic cirrhosis, the 18 zone-specific transcriptome comparisons between phenotypes yielded in total 2,845 significant DEGs (|logFC| > 1, FDR < 0.01) with 1,673 genes detected in multiple comparisons (Figure 30). The number of DEGs increased with disease progression. In addition, DEG counts of individual zones indicate more pronounced changes in the pericentral region. However, it should be noted that the difference in DEG counts might arise from mere FDR adjusted p-value cut-off based statistical effects due to different replicate numbers and inter-sample variability of individual comparisons. Canberra clustering of log fold changes of all DEGs detected in the dataset, without consideration of FDR adjusted p-values, confirms the increase of differential gene expression in progressed disease states (Sup. Figure 6 A). Furthermore it indicates phenotype related changes to be more similar in pericentral and intermediate hepatocytes than in periportal hepatocytes. Moreover, differential gene expression was more similar between the zones of a phenotypic group than between phenotypes (Sup. Figure 6 A), suggesting besides the loss of transcriptional zonation also a dominant lobule-wide deregulation of gene expression. Out of all DEGs, 205 genes were previously found to have zonated expression profiles in normal liver tissue (V0.1). Thus, 25.4 % of zonated genes were detected as differentially expressed during fatty liver disease progression. Along with the overall number of DEGs, the number of DEGs that show zonated expression patterns in healthy liver increased with progressing disease stage from NASH to fibrosis and finally to cirrhosis (Figure 32 A). Interestingly, most zonation-related expression changes were detected in the pericentral zone. As was already indicated in Figure 28, they constitute predominant upregulation of genes with strong periportal expression in healthy liver as well as a strong downregulation of pericentral genes. Thus, in cirrhosis expression patterns of these genes are rather similar across hepatic zones, while they show strong zonated expression in healthy liver (Figure 32 B). These findings add further evidence to the hypothesis of progressive portalization of pericentral hepatocytes.

The complementary differential analysis of zone-wise DNA methylation patterns during fatty liver disease showed an increase of DMR counts in late disease stages, together with considerably higher effect sizes (Figure 31). No major differences regarding the number of hypomethylated and hypermethylated DMRs were observed. Interestingly, also standard deviations of significant methylation differences were higher in FIB3 and ACI than in controls and NAFLD samples (Sup. Figure 7 A). It should ne noted that overlap of DMRs between zones was low (Figure 31), which was frequently explained by differences in CpG coverage between samples. The genomic annotation of DMRs was similar across all comparisons with the majority of DMRs being annotated to introns (35 – 36 %), exons (12 – 15 %) and promoter regions (29 – 31 %, Sup.

Figure 7 B). In line with transcriptional deregulation (Sup. Figure 6 A), differential methylation levels in pericentral hepatocytes and intermediate hepatocytes seem to be more similar than in periportal hepatocytes (Sup. Figure 7 C), even though intermediate and periportal hepatocytes share more significant DMRs (Figure 31). Moreover, correlations of covered DMRs between individual zones within a distinct phenotype were rather low (r = 0.63 – 0.73), which suggests zone-specific differences on the epigenetic level and is in line with the phenotype-specific zonation patterns of DNA methylation described in Figure 29.
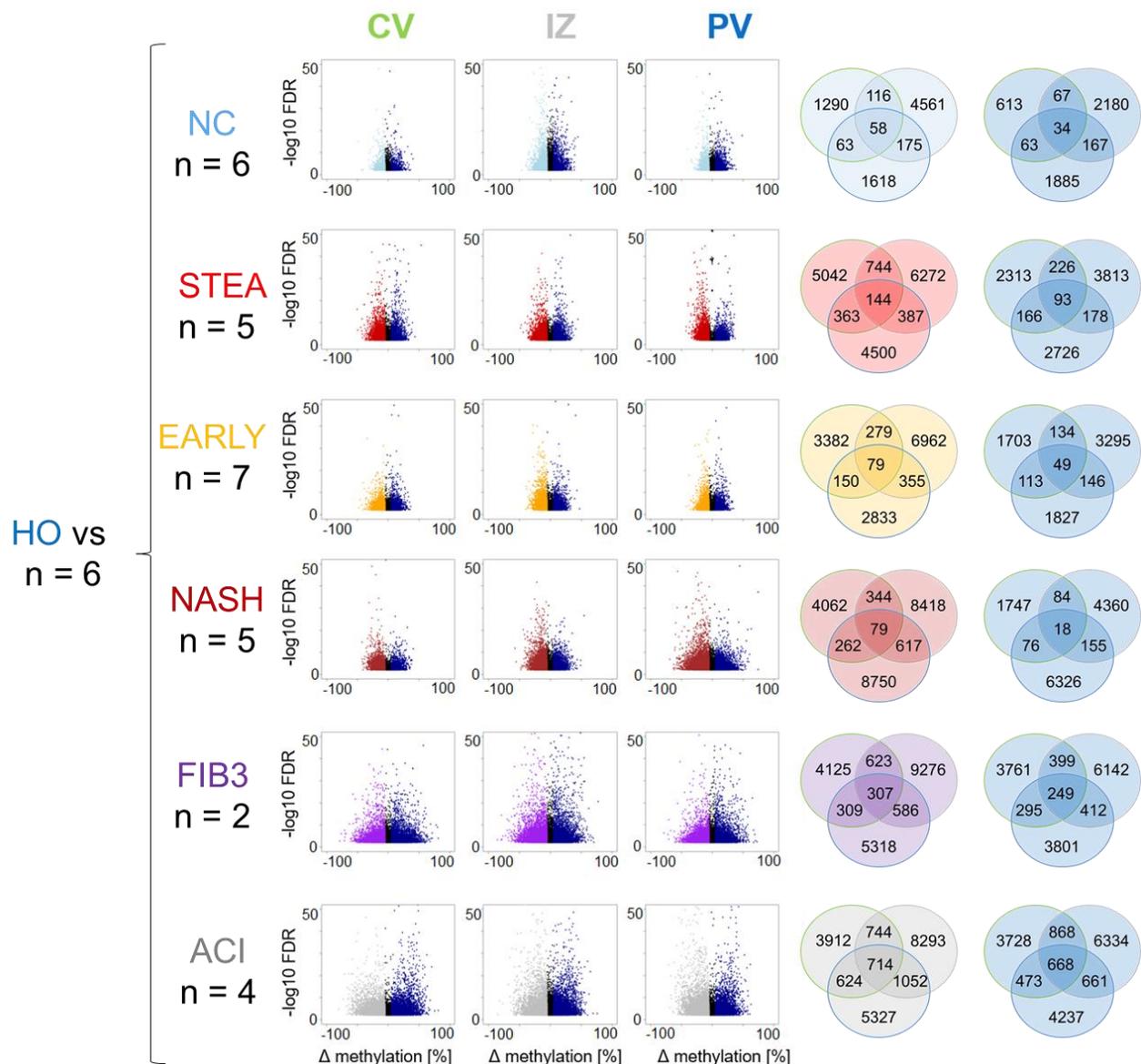
Figure 31: Volcano plots of zone-specific differential analysis of DNA methylation between HO and the other phenotypic groups (NC, STEA, EARLY, NASH, FIB3 and ACI). The x-axis visualizes the methylation difference (%), while the y-axis depicts negative log10 of FDR adjusted p-values. Significant DMRs (|Δ methylation| > 5%, FDR < 0.01) and corresponding Venn diagrams are colored by the hypomethylated group (NC = light blue, HO = dark blue, STEA = red, EARLY = orange, NASH = dark red, FIB3 = purple, ACI = grey). Venn diagrams show the overlap of DMRs detected in pericentral (CV, green line), intermediate (IZ, grey line) and periportal (PV, blue line) zones.
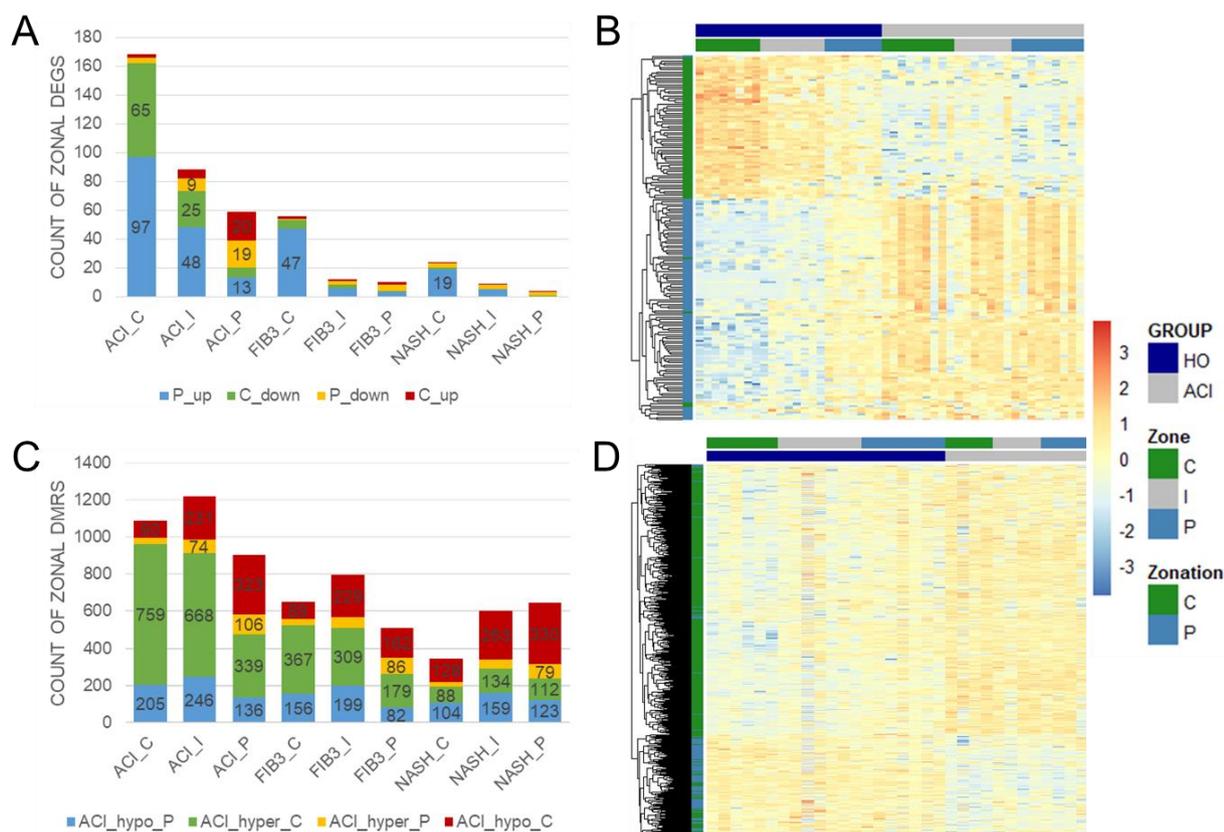
Figure 32: Impact of fatty liver disease progression on zonated genes and DMRs. A – Stacked barplot of zone-specific DEG counts derived from the comparisons between ACI, FIB3 or NASH with healthy obese (Figure 30) of genes with zonal expression profiles in normal liver tissue. P_up (blue) refers to upregulation in disease of genes with periportal expression pattern in healthy liver, C_up (red) equivalently to upregulation of pericentral genes in disease, while C_down (green) and P_down (yellow) mark respective zonal genes downregulated in disease. B – Exemplary expression z-score heatmap of zonal DEGs between ACI_C (grey) and HO_C (dark blue) across all three zones. Red corresponds to high expression levels, blue to low or no expression. Pericentral samples are labelled in green, intermediate in grey and periportal in blue. Heatmap rows represent genes with periportal (blue) or pericentral (green) expression profiles in healthy liver tissue (Zonation). Rows are clustered using manhattan distance metrics. C – Stacked barplot of zone-specific DMR counts derived from the comparisons between ACI, FIB3 or NASH with healthy obese (Figure 30) of regions with zonal methylation signatures in normal liver tissue. ACI_hypo_P (blue) refers to hypomethylation in disease of regions with periportal hypomethylation in healthy liver, ACI_hypo_C (red) equivalently to pericentral hypomethylation in disease of regions with pericentral hypomethylation in healthy liver, while ACI_hyper_C (green) and ACI_hyper_P (yellow) mark respective zonal regions hypermethylated in disease. D - Exemplary methylation z-score heatmap of zonal DMRs between ACI_C and HO_C across all three zones, annotated as outlined in panel B.

Out of 37,635 significant DMRs in ACI in any zone, 29,968 were covered by all comparisons. In total, 1,239 DMRs were associated with genes with zonated expression in healthy tissue. However, the strongest DNA methylation differences were apparently not directly associated with zonated gene expression, but were rather detected uniformly across all hepatic zones (Sup. Figure 7 D and E). Still, few selected zone-specific DNA methylation differences might exhibit a decisive regulatory potential. Thus, phenotype-specific DMRs, which overlapped zonated DMRs in healthy liver (V.01), were assessed in more detail. The most intriguing differences were observed in regions hypomethylated in pericentral hepatocytes of normal liver tissue. In progressed disease stages, they were frequently hypermethylated, particularly in the pericentral zone, but

partially also featured increased hypomethylation in intermediate and periportal zones (Figure 32 C). These observations entail the disruption of normal zonated methylation signatures in progressed fatty liver disease (Figure 32 D) as already described in Figure 29 B.

It should be noted that even though high numbers of DEGs and DMRs were detected within the pairwise comparisons between HO controls and progressed disease stages (Figure 30, Figure 31), the majority of these DMRs (98.6 %) was not annotated to DEGs (Sup. Figure 7 F). Thus, only 22.1 % of DEGs were associated with differential DNA methylation. Contrary, in zonation of healthy liver 44 % of DEGs show also DNA methylation differences. Possible implications of this observation are discussed in chapter 4.1.2.

Taken together, this overview of phenotype-specific zone-wise deregulations during fatty liver disease progression supports the hypothesis of a progressive disruption of hepatic zonation, both on transcriptional and epigenetic level. In the following, the observed disease associated deregulations (Figure 30, Figure 31) are characterized for individual phenotypes in more detail, to allow the interpretation of observed DEG and DMR patterns also in a more functional context.

Consistent with previous results (Sup. Figure 3), in bland steatosis (STEA) and early steatohepatitis (EARLY) the expression of 413 genes was deregulated. It should be noted that the comparison between STEA and HO of the intermediate zone yielded considerably higher DEG counts than for the other zones (Figure 30). This can be explained by the lower inter-individual variability observed within this sample group (Sup. Figure 6 B), presumably due to differences in technical variability. Differential expression in early NAFLD stages (STEA, EARLY) only affected 13 genes with zonated expression profiles, such as SAA1 or CYP2C19, which were consistently deregulated across hepatic zones and stages of fatty liver disease (Sup. Figure 6 C). In line with our previous findings (Brosch et al., 2018), the relative zonation of these genes remains stable in early stages of NAFLD.

In more progressed steatohepatitis (NASH) zone-specific GO term enrichment analysis of differential gene expression revealed for instance liver function related terms such as triglyceride catabolic process (p-value$_{CV}$ = 0.014, p-value$_{PV}$ = 0.012) and lipid transport (p-value$_{CV}$ = 0.020), which indicate deregulation of fatty acids. Moreover, immune response related terms like acute phase response (p-value$_{CV}$ = 0.03, p-value$_{IZ}$ = 0.03, p-value$_{PV}$ = 0.014) and type I interferon signaling pathway (p-value$_{CV}$ = 0.009) underline the inflammatory state of progressed steatohepatitis. Interestingly, the canonical Wnt signaling pathway was significantly downregulated in the intermediate zone (p-value$_{IZ}$ = 0.005).

Likewise, in fibrotic samples (FIB3) the regulation of lipid metabolic processes (p-value$_{CV}$ = 0.010, p-value$_{IZ}$ = 0.016, p-value$_{PV}$ = 0.025) was deregulated in all hepatic zones, while the disruption of triglyceride homeostasis (p-value$_{CV}$ = 0.010, p-value$_{IZ}$ = 0.049) was only detected in pericentral and intermediate hepatocytes. In line with the progressed disease state, positive regulation of inflammatory response (p-value$_{CV}$ = 0.037, p-value$_{PV}$ = 0.012) and acute phase response (p-value$_{CV}$ = 5 x 10^{-5}, p-value$_{IZ}$ = 1.8 x 10^{-4}, p-value$_{PV}$ = 0.002) were upregulated across all zones.

Further detailed analysis focused on alcoholic cirrhosis samples (ACI), as they represent the most severe phenotype analyzed in this dataset and showed the most pronounced transcriptional and epigenetic differences compared to HO controls (Figure 30, Figure 31). Figure 33 displays the top GO terms either up- or downregulated in ACI by individual zone. Several GO terms, including steroid hormone mediated signalling pathway and acute phase response, were detected in pericentral (Figure 33 A) and periportal

DEGs (Figure 33 B). The most upregulated GO term, both in pericentral and periportal hepatocytes, was the tumour necrosis factor mediated signalling pathway, which was previously shown to play a role in hepatic stellate cell mediated fibrogenesis by remodelling of the extracellular matrix (Tarrats et al., 2011). However, more frequently, top GO terms upregulated in ACI were distinct for hepatic zones.
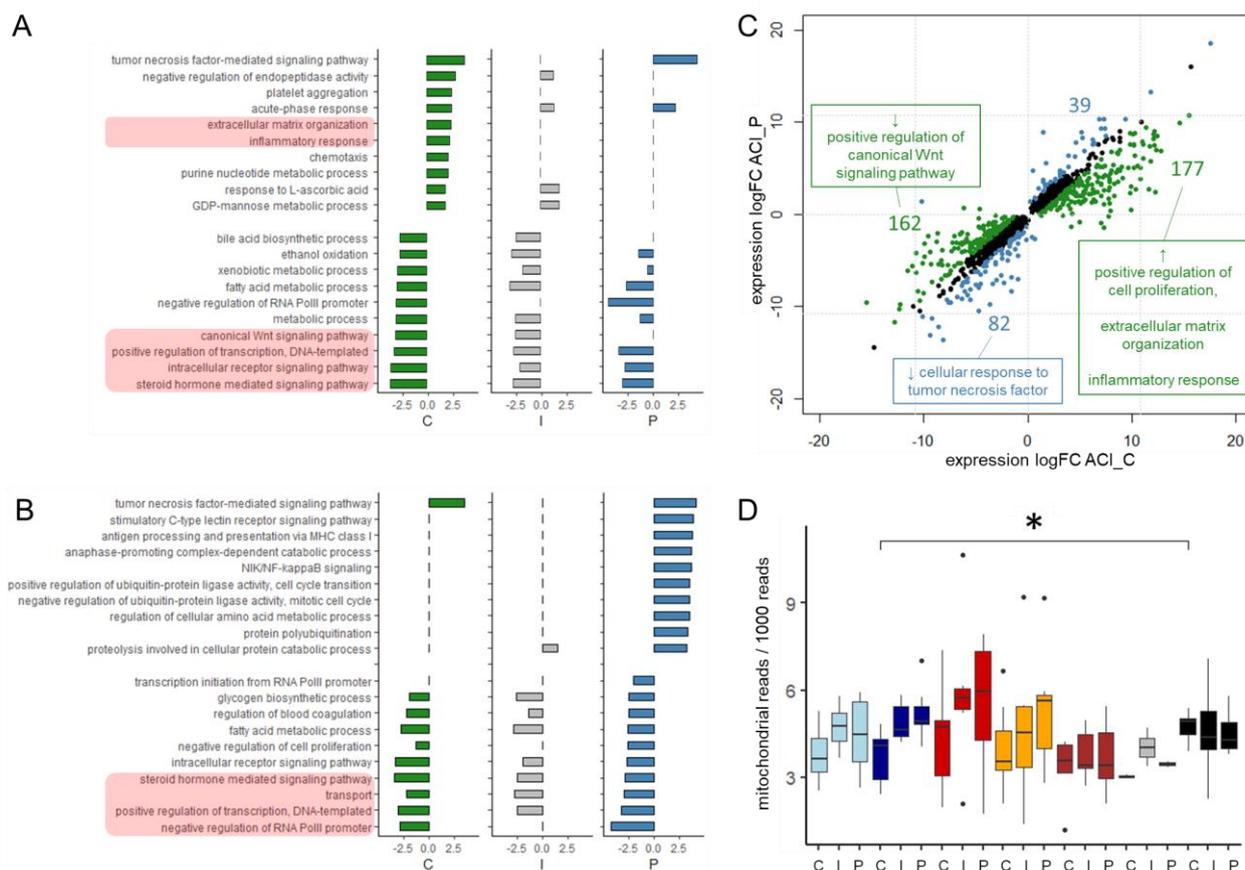
Figure 33: Zone-specific GO enrichment of DEGs detected between ACI and HO. Z-scores represent relative GO enrichment, with negative values corresponding to downregulation and positive values to upregulation in ACI. The 20 GO terms with the strongest enrichment in the pericentral (A) and periportal (B) zone are displayed for all three zones with CV colored in green, IZ in grey and PV in blue. The full table of GO terms is shown in Sup. Table 8. Terms enriched in DMR associated genes as well are highlighted in red. C – Scatterplot of pericentral and periportal expression logFC. Genes with |logFC$_{ACI\_CV}$| more than 1 fold change higher than |logFC$_{ACI\_PV}$| are defined as pericentral specific DEGs, highlighted in green. Accordingly, periportal specific DEGs are marked in blue. GO terms of interest enriched (p < 0.05) in zone-specifically deregulated genes are displayed. D – Fraction of mitochondrial reads per zone (pericentral = C, intermediate = I, periportal = P) and phenotype (NC = light blue, HO = dark blue, STEA = red, EARLY = orange, NASH = dark red, FIB3 = black, ACI = grey). Significant differences detected by Wilcoxon test are marked by asterisk (p < 0.05).

Intriguingly, extracellular matrix organization and inflammatory response were uniquely detected in pericentral hepatocytes. Moreover, the canonical Wnt signalling pathway was downregulated in pericentral and intermediate zones. Even though few enriched GO terms (Sup. Table 8) were shared between DEGs and DMRs (Figure 33), matching terms included top terms such as the pericentral depletion of Wnt signaling pathway and enrichment of extracellular matrix organization (Figure 33 A). Taken together, GO analysis suggests major regulatory and metabolic changes in cirrhotic liver.

In order to substantiate zone-specific differences, only DEGs with $|logFC_{ACI\_CV}|$ more than 1 fold change higher than $|logFC_{ACI\_PV}|$ were considered for pericentral specific GO enrichment analysis, and vice versa for periportal (Figure 33 C). Indeed, this analysis confirmed the pericentral downregulation of canonical Wnt signalling, along with the upregulation of cell proliferation, extracellular matrix organization and inflammatory response. DEGs specific for periportal hepatocytes in ACI were enriched for cellular response to tumour necrosis factor (Figure 33 C), which was however also identified in the pericentral zone (Figure 33 A).

In addition, the analysis of mitochondrial abundance, deduced from RRBS coverage, revealed a significant difference between HO and ACI in the pericentral zone (Wilcox test $p < 0.05$). In line with previous findings in healthy liver (Figure 21 A), HO samples exhibit a mitochondrial gradient along the porto-central axis. In contrast, the ratio of mitochondrial reads was equal across zones in ACI (Figure 33 D), i.e. in cirrhosis pericentral hepatocytes apparently resemble periportal hepatocytes on the level of mitochondrial abundance as well.

The observed deregulation of the Wnt signalling pathway in ACI is of particular interest, as we and others identified Wnt signaling as an important regulator of pericentral hepatic zonation in normal human liver tissue (Brosch et al., 2018; McEnerney et al., 2017). Interestingly, in pericentral hepatocytes derived from cirrhotic tissue, several upstream regulators of Wnt signalling such as LGR5, NOTUM and RSPO2 are strongly downregulated, while further downstream regulators including β-catenin and Frizzled are upregulated, which could potentially be seen as a compensatory effect (Sup. Figure 8 A). Although most DEGs detected in ACI were not associated with differential DNA methylation (Sup. Figure 7 F), the Wnt signaling pathway is highly enriched for DMRs in pericentral hepatocytes, and especially key regulators such as WNT, AXIN2 and LGR5 show pronounced DNA methylation differences in ACI (Sup. Figure 8 A). In periportal hepatocytes, which show only low expression of Wnt signaling pathway members in normal liver, Wnt signaling is even further downregulated, accompanied by differential DNA methylation of several pathway members (Sup. Figure 8 B). These observations highlight the importance of the Wnt signaling pathways in hepatic zonation, and suggest the downregulation of Wnt signaling to play a major role in the loss of hepatic zonation in progressed fatty liver disease.

In contrast, the Notch signaling pathway, which is enriched in periportal hepatocytes of normal human liver tissue (Brosch et al., 2018), and was described to be involved in hepatic regeneration upon partial hepatectomy (Köhler et al., 2004), was not significantly deregulated in ACI. Yet, a trend towards pericentral and intermediate upregulation of NOTCH3 was observed (HO.I vs ACI.I: FDR = 0.014, logFC = - 4.6, HO.C vs ACI.C: FDR = 0.030, logFC = -3.7, Sup. Figure 9). The Notch ligand JAG1, which is periportally zonated in normal tissue, was also not significantly deregulated, but showed higher inter-individual variability (Figure 34 A). SOX9, which interacts with Notch signaling in a positive feedback loop (Leung et al., 2016), was significantly upregulated in pericentral and intermediate hepatocytes of cirrhotic tissue, with expression levels resembling those in periportal hepatocytes (Sup. Figure 9, Figure 34 A). Moreover, the liver progenitor marker EPCAM was strongly upregulated in the pericentral zone as well (logFC = 9.8, FDR = 1.9×10^{-6}, Figure 34 A). Together with the loss of pericentral zonation signatures of Wnt signaling pathway members such as LGR5 and AXIN2, and the overall downregulation of liver progenitor marker TBX3 (Figure 34 A), the pericentral upregulation of several morphogens, which show periportal expression profiles in normal liver tissue, indicates the disruption of regulatory morphogen signaling interactions in cirrhotic liver.
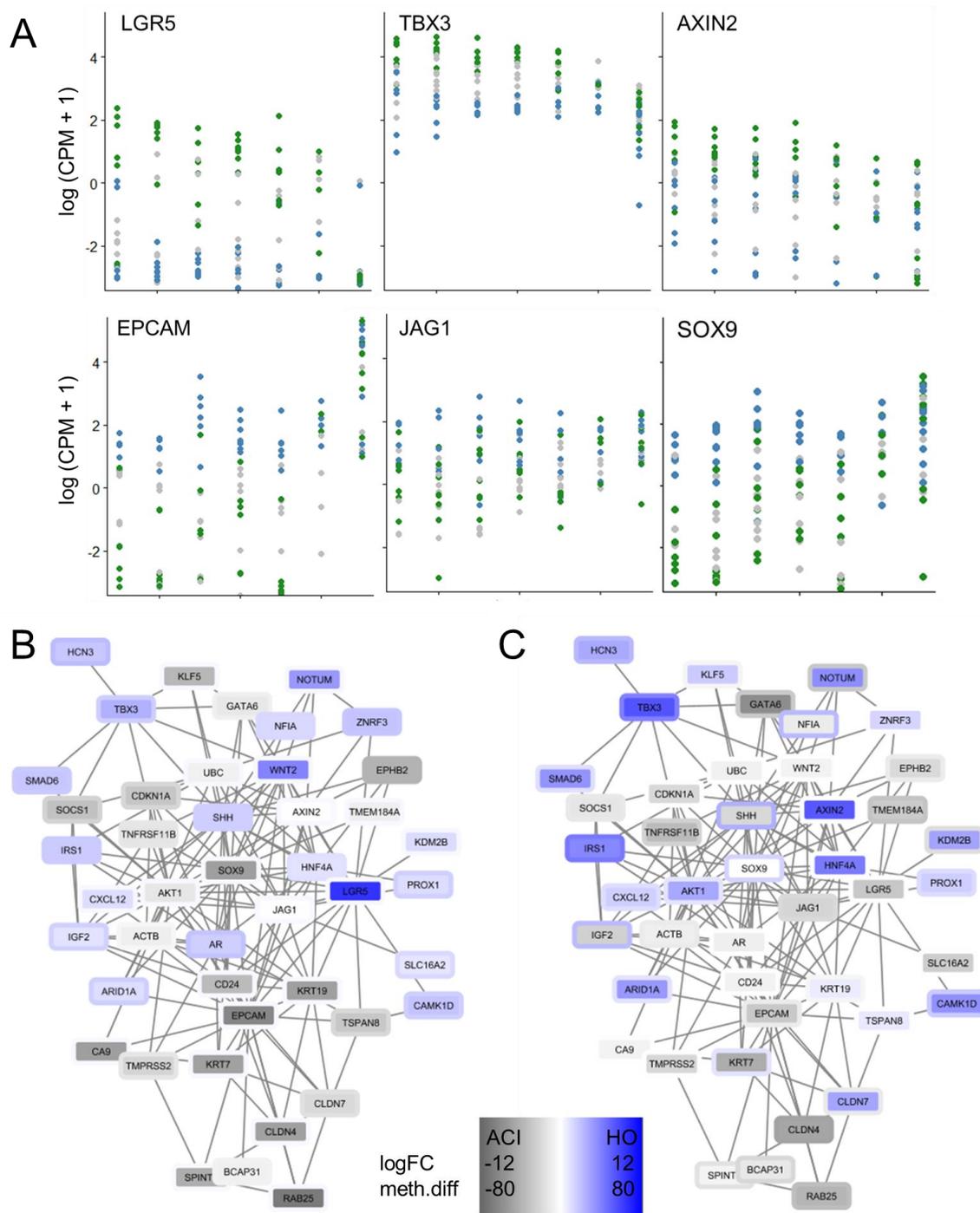
Figure 34: Morphogen gradients and resulting regulatory networks in fatty liver disease. A – Transcriptional profiles of LGR5, TBX3, AXIN2, EPCAM, JAG1 and SOX9. The x axis represents fatty liver disease progression: NC, HO, STEA, EARLY, NASH, FIB3 and ACI. Transcription values are depicted as log (CPM + 1). Green corresponds to pericentral (CV), grey to intermediate (IZ) and blue to periportal (PV) samples. B and C – String interaction networks of LGR5, AXIN2, TBX3, EPCAM, JAG1 and SOX9 with their direct interaction neighbors with deregulated expression in pericentral ACI. Fill colors represent pericentral, border colors periportal hepatocytes. B – Expression illustrated as logFC between pericentral ACI and pericentral HO samples. Blue corresponds to downregulation and grey to upregulated expression in ACI. C – Methylation difference of associated DMRs. Hypomethylation in ACI is depicted in grey, hypermethylation in blue. Genes that do not show any significant differences (FDR < 0.01) are marked in white.

In order to assess this further, morphogens of interest (LGR5, TBX3, AXIN2, EPCAM, JAG1 and SOX9) as well as direct interaction partners which featured both transcriptional (Figure 34 B) and epigenetic (Figure 34 C) deregulation in pericentral hepatocytes of cirrhotic tissue, were visualized as interaction network. This network highlights the intricate interplay between Wnt and Notch signaling, but also includes important factors of other signaling pathways such as Sonic Hedgehog signaling, which is described as a pro-fibrogenic pathway in fatty liver disease (Rangwala et al., 2011).

In summary, the zonal analysis of steatohepatitis, fibrosis and cirrhosis revealed the progressive loss of transcriptional and epigenetic zonation in late stages of fatty liver disease. In particular the portalization of pericentral hepatocytes, which is seen on the level of gene expression, but also in zonal mitochondrial abundance, might present an important feature of disease progression. In addition, on epigenetic level phenotype-specific zonation patterns were detected, which were apparently not directly associated with zonated gene expression. These deregulations of hepatic zonation might be associated with the pericentral disruption of Wnt signaling and corresponding upregulation of periportal morphogens.

3.2 Spatial reconstruction of single cell transcriptomes into tissue context

Multiple observations on hepatocytes microdissected from different zones (3.1) indicate a gradient-like regulation of hepatic zonation along the porto-central axis. In particular differential gene expression between pericentral and periportal hepatocytes revealed an expression gradient (Figure 16 B), contradicting the notion of distinct hepatic zones. This is in line with previous findings from bioinformatical spatial reconstruction of single cell RNA sequencing data of hepatocytes from mouse liver (Halpern et al., 2017). The inherent low resolution of our microdissection approach limits however the more detailed analysis of expression gradients in our human dataset. To this end, spatial reconstruction of high throughput single cell RNA sequencing data of a human liver was pursued as complementary approach.

3.2.1 Pseudospace inference of human hepatocytes

Raw single cell transcriptome data of hepatocytes isolated from human liver were provided by Dr. Mario Brosch and Prof. Dr. Jochen Hampe (TU Dresden). Data were generated using the commercial 10X Genomics Chromium Single Cell Gene Expression Solution, which is a droplet based high-throughput single cell RNA sequencing approach comparable to the Drop-seq method (Macosko et al., 2015). Data were processed by Dr. Karl Nordström and gene-wise UMI counts for 14,949 detected barcodes were provided as sparse matrix for further analysis. In principle, each individual barcode identifies a single cell. However, particularly in high-throughput single cell RNA-seq datasets there is a high amount of technical noise. Thus, rigorous quality filtering has to be applied prior to data analysis. The chosen filtering criteria are based on recommendations by several state of the art scRNA-seq data analysis packages (Macosko et al., 2015; McCarthy et al., 2017; Wolf et al., 2018). During initial filtering, cells with unique UMI count below 2,500 were excluded (Sup. Figure 10 A) to remove cells with low coverage. To reduce drop out rates, cells covering less than 500 genes were excluded as well. Additionally, cells with more than 2,000 genes and or 12,500 UMIs were filtered out to remove potential cell duplets. Cells with more than 15 % of reads mapping to the mitochondrial genome were also excluded, as high amounts of mitochondrial reads indicate dead or damaged cells. Thus, initial filtering yielded in total 8,181 single cells with sufficient data quality.

It should be noted that no filtering of low-abundance genes was performed, as genes such as LGR5 or TBX3, which are only expressed at low levels in a small subset of cells, are of major interest regarding liver zonation. Instead, the average expression of each gene across all cells was taken into account for calculation of highly variable genes (HVGs, Macosko et al., 2015), which were then used for downstream analysis. Principal Component Analysis (PCA) of HVGs (Sup. Figure 10 B) showed a dense cluster of cells and several outliers. Based on expression levels of marker genes, most outliers were identified as non-parenchymal cells (NPCs) such as immune cells and cholangiocytes (Sup. Figure 10 B and C). Additional filtering by marker gene expression as well as based on principal components (Sup. Figure 10 B and C) resulted in a final dataset of 7,912 high confidence primary human hepatocytes, which were then used for spatial reconstruction.
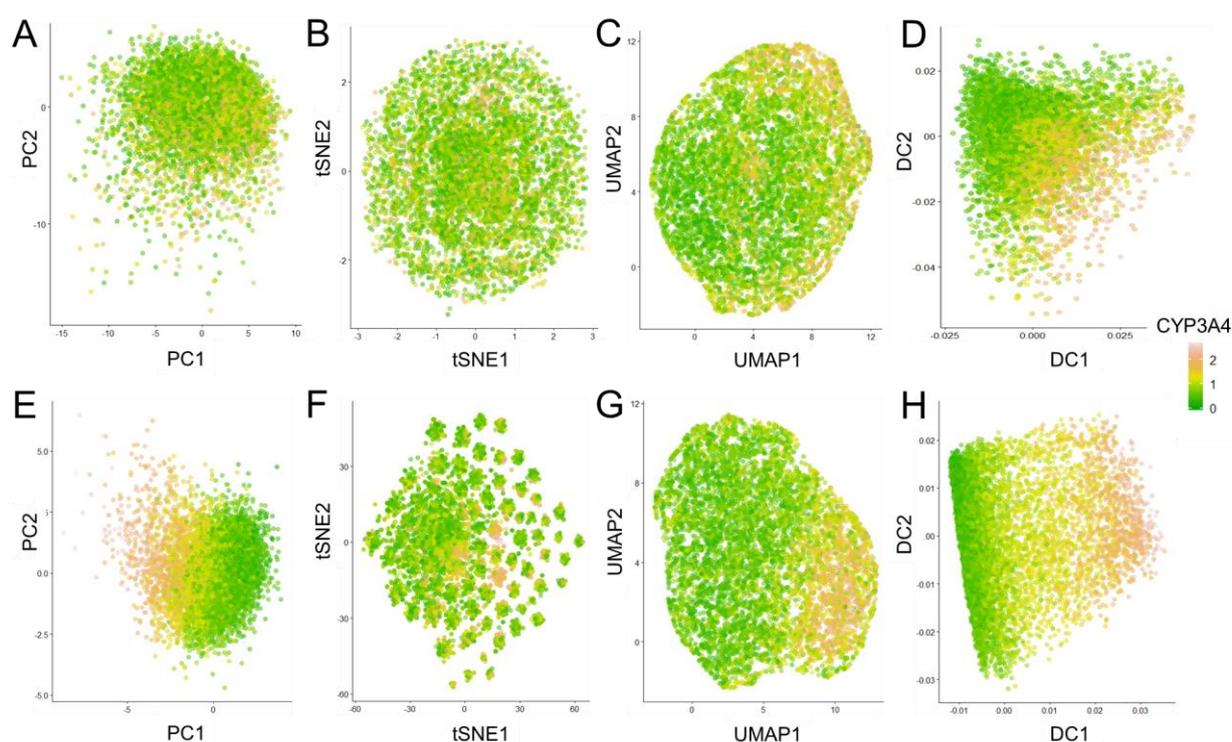
Figure 35: Dimensionality reduction by PCA (A, E), t-SNE (B, F), UMAP (C, G) and diffusion map (D, H), colored by pericentral CYP3A4 expression with yellow representing intermediate and orange high expression levels. The plots are either produced on all HVGs (A – D) in the dataset or on 107 HVGs with known zonal expression patterns in human liver (E – H). For t-SNE plots different perplexities from 1 to 100 were tested. Presented plots (B, F) are based on a representative perplexity of 20 and 11 principle components derived from PCA (A).

Expression analysis of microdissected hepatocytes from three hepatic zones revealed zonation as a major driver of variance in healthy liver (Sup. Figure 2 A). Thus, it was reasoned that HVGs in the single cell expression dataset should contain spatial information. Expression variability of single hepatocytes was calculated based on average expression and corresponding dispersion for each gene as outlined in the SEURAT package (Macosko et al., 2015), yielding 3,633 HVGs. Subsequently, several approaches for dimensionality reduction and visualization were considered for pseudo-spatial clustering of cells. To this end, resulting plots were colored by single cell expression of strong zonation markers such as the pericentral marker CYP3A4 (Figure 35). All findings were validated by using at least 10 pericentral and 10 periportal marker genes, ensuring CYP3A4 profiles to be representative of actual zonated expression (data not shown). Based on all HVGs in the dataset, PCA (Figure 35 A) did not reveal any obvious clustering of single cells by zonation markers. As t-distributed stochastic neighbor embedding (t-SNE) depends heavily on the chosen perplexity, which can be considered as a score for the number of expected direct neighbors, several reasonable perplexities between 2 and 100 were tested without yielding any obvious zonal clustering. Figure 35 B displays a representative t-SNE (perplexity = 2). Uniform Manifold Approximation and Projection (UMAP), a more recent non-linear algorithm featuring more consistent clustering and preservation of continuous expression profiles than t-SNE (McInnes et al., 2018), indicated a slight accumulation of pericentral hepatocytes, although no gradient-like patterns were observable (Figure 35 C). As the expected expression gradient points to continuous dynamics in the single cell dataset, a k nearest neighborhood

diffusion map algorithm, which is a spectral non-linear method relying on continuous diffusion-like dynamics and was originally proposed for pseudotime inference (Haghverdi et al., 2015), might serve as reasonable tool for pseudo-spatial reconstruction along the hepatic porto-central axis. Indeed, a diffusion map, on all HVGs detected in the dataset (Figure 35 D, parameters detailed in chapter 2.2.2.2) showed the best separation of hepatic zones of all tested dimensionality reduction approaches. Yet, the separation appeared to be rather noisy and did not reveal a clear zonation profile along the diffusion component 1 (DC1), which generally represents the strongest source of variation.

Diffusion map geneset	Correlation between DC1 and:	Pearson correlation	P-value
HVGs (3633)	HAL (PV)	-0.558	< 2.2e-16
	CYP3A4 (CV)	0.473	< 2.2e-16
Zonated HVGs (127)	HAL (PV)	**-0.814**	< 2.2e-16
	CYP3A4 (CV)	**0.906**	< 2.2e-16
Higly expressed zonated HVGs (34)	HAL (PV)	-0.811	< 2.2e-16
	CYP3A4 (CV)	0.902	< 2.2e-16
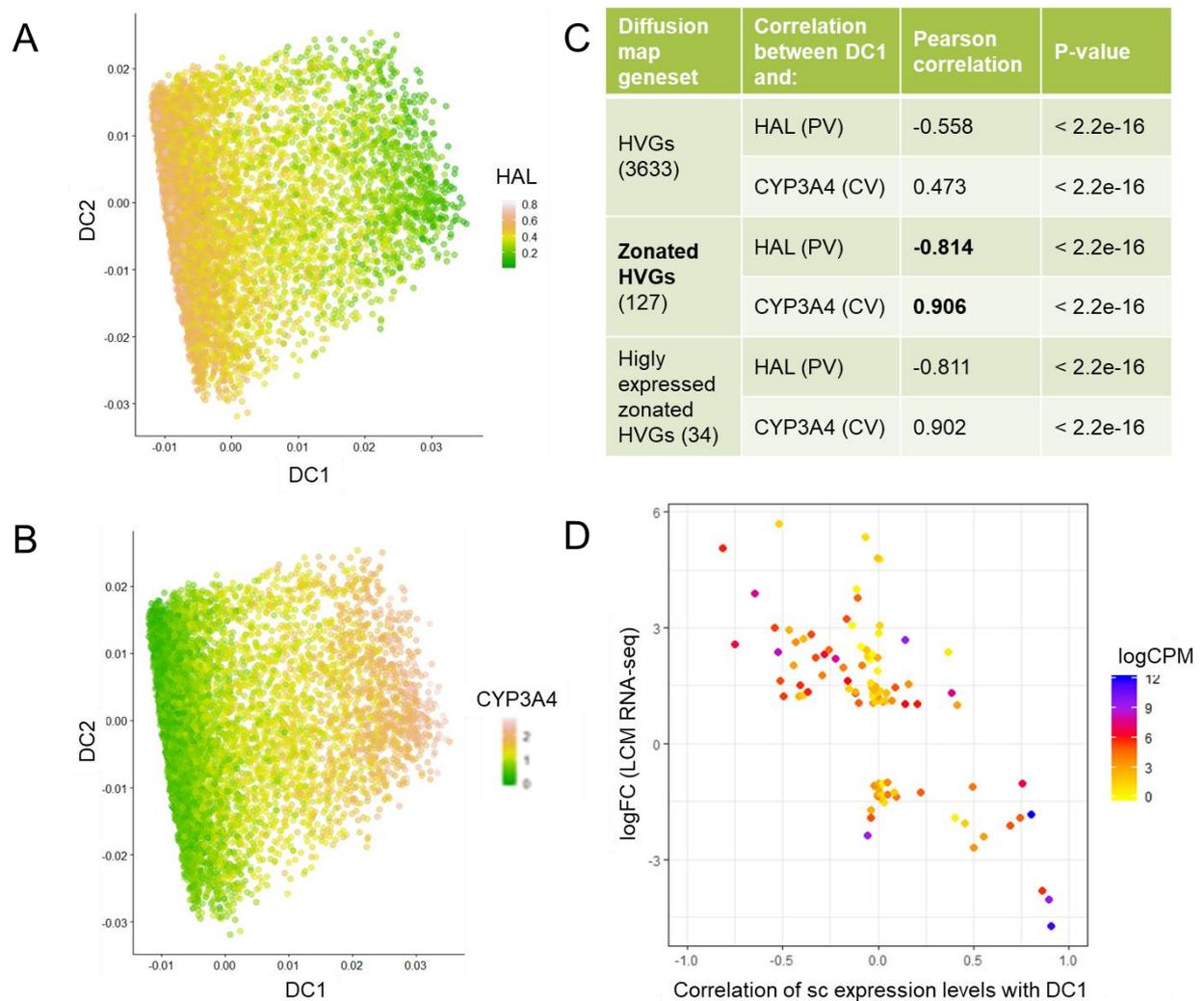
Figure 36: Diffusion map based on 107 HVGs with known zonal expression patterns in human liver colored by periportal marker HAL (A) and pericentral marker CYP3A4 (B, same plot as Figure 35 H). C – Table with Pearson Correlations and corresponding p-values between diffusion component 1 (DC1) and marker genes expression levels for diffusion maps generated from all HVGs, zonated HVGs and the subset of zonated genes with the highest expression level in the single cell RNA-seq dataset. Counts in brackets display the number of genes in the respective analysis. D – Scatterplot of Pearson correlations between DC1 and single cell expression levels of the zonated HVGs used for diffusion map generation versus corresponding $\log_2$ expression fold changes between CV and PV derived by LCM RNA-seq.

Thus, further analysis was limited on HVGs with significant zonation patterns in the laser capture microdissection (LCM) RNA-seq dataset (3.1.1) as reference gene set, thereby reducing other sources of variation and focusing the analysis on spatial variability. Indeed, this approach markedly improved separation of hepatic zones (Figure 35 E – H). While particularly t-SNE, but also UMAP did not yield adequate spatial reconstruction, both PCA and diffusion map resulted in a clear separation of pericentral and periportal hepatocytes, suggesting a gradient-like expression pattern (Figure 35 E and H). However, hepatocytes of the intermediate zone were not well represented in the PCA. Contrary, the diffusion map revealed a striking separation of the three hepatic zones in a gradient along DC1 (Figure 35 H). Thus, further validation of pseudospace inference focused on the diffusion map approach.

For visualization of zonation, a combination of the representative marker genes CYP3A4 (pericentral) and HAL (periportal) was selected, which showed clear separation of hepatic zones along DC1 in the diffusion map generated on HVGs with zonated expression signatures (Figure 36 A and B). Contrary, DC2 showed high Pearson correlations with expression levels of the tumor metastasis markers CXCL1 ($r = -0.830$, $p < 2.2e-16$) and CXCL2 ($r = -0.780$, $p < 2.2e-16$, Acharyya et al., 2012), even though seemingly healthy liver tissue was analyzed. As the sample was derived from a colon cancer patient undergoing screening for liver metastasis with negative metastasis diagnostic, expression of these two markers might be explained by early stage metastasis development, not yet detectable by standard diagnostic liver biopsy assessment. As DC1 already showed a convincing spatial separation, DC2 was not further considered in this context.

To assess the performance of diffusion map based pseudospace inference, Pearson correlations between DC1 of diffusion maps generated on different subsets of HVGs and single cell expression levels of these marker genes were assessed (Figure 36 C). As already suggested by diffusion map plots (Figure 35 D and H), the diffusion map calculated on all 3,633 HVGs resulted in correlations between DC1 and single cell expression levels of HAL ($r = -0.558$) and CYP3A4 ($r = 0.473$) considerably lower than the approach using only zonated HVGs. Highly significant Pearson correlation coefficients ($p < 2.2e-16$) between DC1 of zonated HVGs and single cell expression levels of the marker genes ($r_{HAL} = -0.814$, $r_{CYP3A4} = 0.906$) demonstrate the high correlation of DC1 with expression patterns of these strongly zonated genes (Figure 36 C). Interestingly, limiting the set of HVGs for diffusion map calculation on the 34 genes with the highest average expression level and lowest drop-out rate (average normalized $\log_2$ expression > 0.1) yielded almost the same correlation values as when using all zonated HVGs (Figure 36 C), which indicates that the diffusion map is mostly driven by HVGs with high average expression. Presumably, this can be explained by low-abundance genes being more susceptible to technical noise, thereby limiting their utility in retrieving spatial information from their expression variability across single cells. Further reducing the diffusion map on only two genes with high average expression and strong zonation profiles (e.g. CYP3A4 and HAL) still yielded a separation by spatial localization. Even though correlations between DC1 and expression levels of marker genes were lower than when using all zonated HVGs ($r_{HAL} = -0.636$, $r_{CYP3A4} = 0.685$), the performance was still better than when using all HVGs in the dataset, underlining the robustness of the reference gene set based diffusion map approach. Moreover, a systematic analysis considering correlations between DC1, generated from all 107 zonated HVGs, and the expected zonation patterns, derived from the LCM RNA-seq dataset (chapter 3.1.1), of zonated HVGs confirmed that particularly genes with strong zonation pattern and high average expression correlate well with DC1 (Figure 36 D). The overall correlation

between DC1 and log$_2$ expression fold changes of all zonated HVGs is highly significant (cor = - 0.67, p-value = 3.526e-15) and in a reasonable range, which indicates the validity of this approach. However, including also lowly expressed genes apparently does not compromise diffusion map generation (Figure 36 C). Thus, the diffusion map generated on all zonated HVGs detected in the dataset without filtering of low-abundance genes with high drop-out rates was selected for dimensionality reduction based pseudospace reconstruction of single cell transcriptomes.

To this end, DC1 coordinates were used to relate single cells to their probable spatial origin. In principle, DC1 could be used to assign each cell to a discrete localization in pseudospace. However, to achieve a more robust modelling for subsequent zone-specific data analysis, DC1 values were rather considered as the probability of a cell to originate from a certain hepatic zone. For this purpose, DC1 thresholds for zone definition were selected manually by inspection of the expression patterns of all 107 zonated HVGs. Although the underlying diffusion map was calculated on actual normalized expression values, imputed expression values were used for visualization to obtain a more interpretable result for lowly expressed genes with high drop-out rates. Hence, in total 5,395 cells were assigned to the periportal zone, 1,750 cells to the intermediate zone and 768 cells to the pericentral zone (Figure 37 A). Interestingly, this approach yielded also a more detailed subdivision within zones (Figure 37 B), which might resemble layers of hepatocytes in the anatomy of the hepatic lobule (Figure 9 B). Previously, a study on zonation of the mouse liver lobule, which features up to 15 rows of hepatocytes along the porto-central axis, defined 9 hepatic layers (Halpern et al., 2017). In the human liver, hepatocytes reside in rows of 15 to 25 cells along the porto-central axis (McEnerney et al., 2017). Thus, it was reasoned that subdividing the diffusion map into 10 layers, on average containing each two rows of hepatocytes, might reflect a reasonable additional classification. Visualization of normalized expression levels as scaled average log$_2$ (CPM + 1) values across these pseudo-spatial layers 1 to 10 of genes detected as zonated in LCM RNA-seq indeed suggests that these layers might represent meaningful hepatic subzones (Figure 37 C). Expected expression profiles of strongly zonated genes including GLUL, HAL and H19, as well as genes with less pronounced zonation patterns such as APOLD1 and FNIP2 are accordingly detected in a gradient-like fashion. Contrary, expression patterns of several genes such as ADAMTS1, CMYA5, HMGB3, DMKN and PLIN1 indicate non-monotonic gene expression. However, these genes exhibit low average expression levels (log$_2$ (CPM + 1) < 0.5). Thus, scaled relative expression levels might overestimate actual differences between layers and should be considered with care. For an additional validation of zonal cell assignment, the percentage of reads mapping to mitochondrial genes was assessed between zones. Indeed, single cells mapped to the periportal zone showed a significantly higher percentage (Wilcox test < 0.001) of mitochondrial reads than those assigned to the intermediate or the pericentral zone (Figure 37 D). This finding independently confirms the porto-central mitochondrial gradient detected from the fraction of mitochondrial reads in the LCM-RRBS dataset (Figure 21 A), and indicates an appropriate single cell assignment to hepatic zones.

To assess whether it is also possible to detect accurate zonation signatures of genes not used for diffusion map generation, pericentral CYP2E1 and periportal H19 were removed from the set of 107 zonated HVGs. DC1 of the resulting diffusion map was then used as spatial covariate to determine pseudospatial expression profiles of both genes (Figure 37 E and G), which revealed striking concordance with their corresponding zonation patterns detected by LCM RNA-seq. This observation might be particularly useful in cases when

genome-wide reference maps are not available and additional spatial expression patterns should be inferred from the single cell dataset. Moreover, patterns of genes with uniform expression along the porto-central axis such as VMP1 were confirmed (Figure 37 F), indicating that the diffusion map approach is not prone to overestimation of zonated gene expression.

Finally, diffusion map based pseudotime inference was validated in an independent dataset, comprising a small number of deeply sequenced single cell transcriptomes of cryopreserved primary human hepatocytes from 3 donors (286 cells, in average 1.8 million reads per cell). Using all 444 HVGs in the dataset resulted in a diffusion map capturing mostly inter-individual differences between donors (Sup. Figure 11 A – B). In contrast, the diffusion map on the 38 HVGs with zonated expression profiles in healthy liver tissue, yielded a separation of hepatic zones along DC1 comparable to the previous analysis of sparse data (Figure 36 A). Inter-individual variability was only detected along DC2, allowing the assignment of hepatocytes from different donors into hepatic zones (Sup. Figure 11 C – D). Taken together, these results suggest the suitability of diffusion map based pseudospace inference and embedding of single cells into hepatic zones along the porto-central axis.
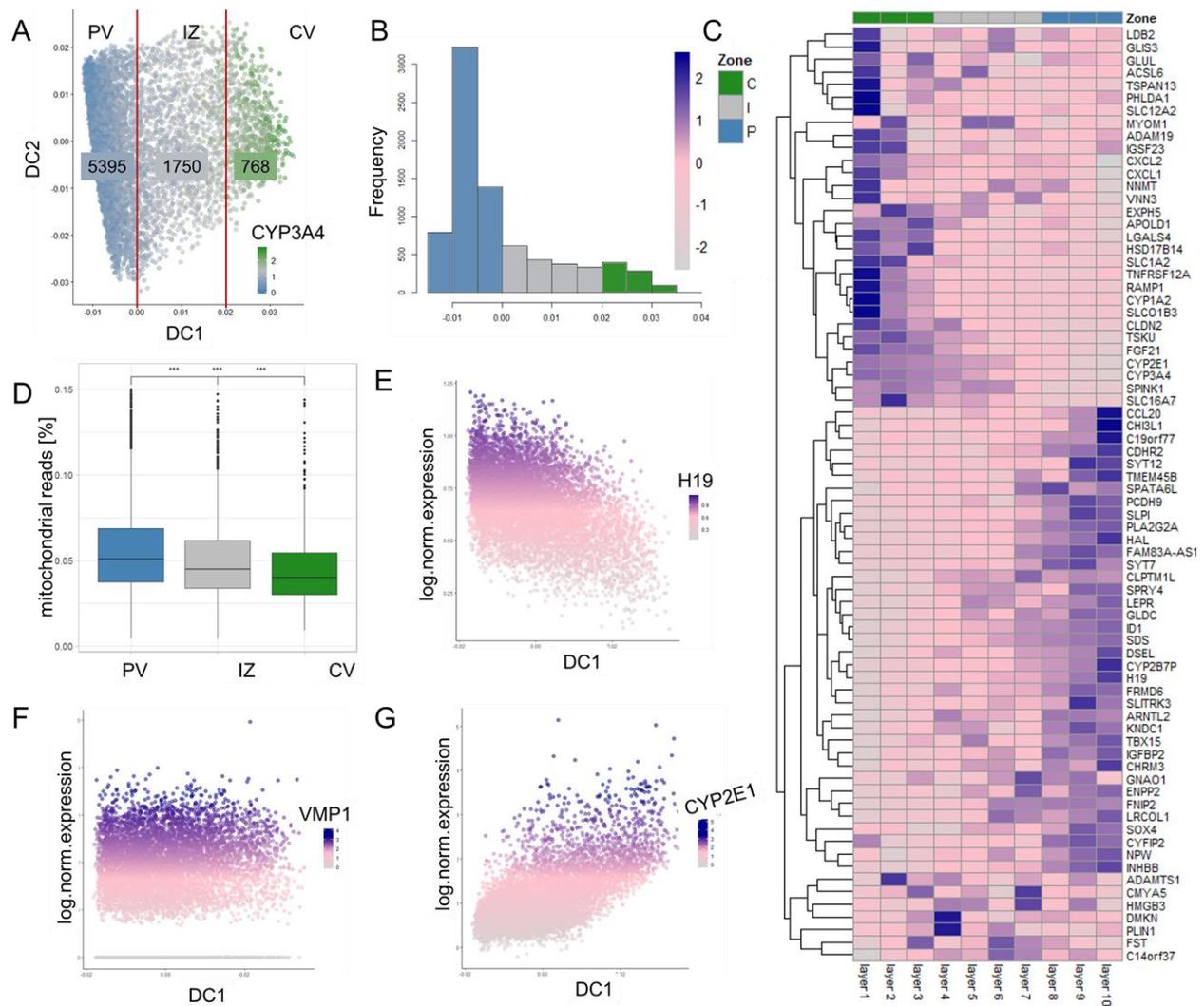
Figure 37: Reconstruction of hepatic zonation signatures from single cell transcriptomes. A – Cells are assigned to the periportal (blue), intermediate (grey) and pericentral (CV) zone based expression of zonation marker genes along DC1. For determination of zonal DC1 thresholds, expression profiles of all 107 HVGs were considered. The displayed plot is colored by exemplary CYP3A4 expression. B – Frequency histogram of cells assigned to the three with further subdivision into smaller layers (1 – 10). C – Heatmap displaying normalized expression levels as scaled average $\log_2$ (CPM + 1) values across pseudo-spatial layers 1 to 10. The heatmap comprises genes detected as zonated in LCM RNA-seq with average $\log_2$ (CPM + 1) > 0.1 in the reconstructed single cell dataset. Blue corresponds to increasing expression in relation to other subzones. Layers are annotated by corresponding hepatic zone and genes are clustered by correlated expression. D – Mitochondrial abundance gradient deduced from the ratio of scRNA-seq reads mapping to mitochondrial genes [%] in cells assigned to PV, IZ and CV. Significant differences between zones detected by Wilcoxon test are marked by asterisks ($p < 0.001$). E, F and G – Proof of principle: Scatterplots of DC1 from a diffusion map generated on zonated HVGs, omitting randomly chosen periportal H19 and pericentral CYP2E1, versus normalized single cell expression levels of H19 (E), equally expressed VMP1 (F) and CYP2E1 (G). Grey represents no or very low gene expression, purple highlights high expression.

3.2.2 Interspecies comparison of liver zonation between human and mouse based on spatial reconstruction of single cells

The transcriptional zonation of the mouse liver lobule has recently been studied systematically by bioinformatical modelling of single cell RNA-seq data (Halpern et al., 2017). This approach combined single cell transcriptome data generated by MARS-seq with sm-FISH of 6 zonation landmark genes to infer spatial information, thereby defining 9 hepatic zones. To enable an interspecies comparison between human and mouse, merged zones 1 to 2 of the mouse data were considered as pericentral, zones 4 to 5 as intermediate, and 8 to 9 as periportal. One-to-one matching orthologues were then compared to our human LCM RNA-seq dataset (3.1.1) and to our human spatially reconstructed scRNA-seq data (3.2.1), yielding 12,179 genes. To obtain a deeper impression on methodical bias and the accuracy of pseudospace inference, the human single cell RNA-seq dataset was compared to the zonal LCM RNA-seq dataset as well.

Expression values of all datasets were normalized as $\log_2$ (CPM + 1). In reconstructed single cell datasets normalized expression values of single cells were averaged across assigned zones. On global transcript level high Spearman correlations between datasets were obtained, indicating an overall comparability of the datasets (Figure 38 A and B). Correlations between methodically different human datasets were higher (0.84 – 0.85) than between mouse and human (LCM: 0.74 – 0.75, sc: 0.69 – 0.70), suggesting a higher impact of species than of technical variability. Out of the 805 zonated genes detected in the human LCM RNA-seq dataset, 575 genes featured one-to-one orthologues in the mouse. Of these, 313 were sufficiently covered in the mouse dataset and including the human scRNA-seq dataset reduced the number of covered zonated genes further to 299. Compared to global transcript levels, these zonated genes yielded lower Spearman correlations (Figure 38 C). The expression of zonated genes still showed considerably high Spearman correlations between the human datasets (0.58 – 0.71), indicating a reasonable spatial reconstruction of the single cell data. Contrary, correlations of zonated genes between human and mouse were heavily reduced (0.29 – 0.40), which suggests differences in hepatic zonation profiles between species.

In this analysis, the LCM RNA-seq approach was considered as the gold standard for the detection of zonated genes expression, as it preserves information regarding the spatial origin of microdissected cells and does not rely on bioinformatical pseudospace inference. However, it is also possible to determine zonated genes in the reconstructed single cell datasets. To this end, Halpern et al. (2017) considered genes with average expression differences between pericentral layer 1 and periportal layer 9 that featured a Kruskal-Wallis q-value < 0.2 as zonated. This approach yielded in total 3,496 zonated genes, representing about 50 % of expressed genes in the dataset.

Adapting this approach, the human single cell dataset yielded 1,479 genes with Kruskal-Wallis q-value < 0.2 of the average expression difference between single cells assigned to the pericentral and periportal zone. Yet, the overlap of zonated genes detected in the individual datasets was rather low (Figure 39 A). Only 21 genes, comprising highly expressed genes with strong zonation profiles such as CYP2E1, CYP1A2, HAL and SDS were detected as zonated in all three datasets. Despite high correlations between expression profiles of zonated genes (Figure 38 C), both human datasets showed only an overlap of 51 zonated genes. The high number of genes only detected as zonated in the single cell dataset (1,472), which did not show any indication for differential gene expression between zones in the LCM RNA-seq dataset (Figure 39 B),

questions the suitability of the method used by Halpern et al. (2017) for differential expression detection in reconstructed single cell data. The 754 zonated genes not detected as zonated by the reconstructed scRNA-seq approach were not sufficiently covered in the single cell dataset (Figure 39 C).
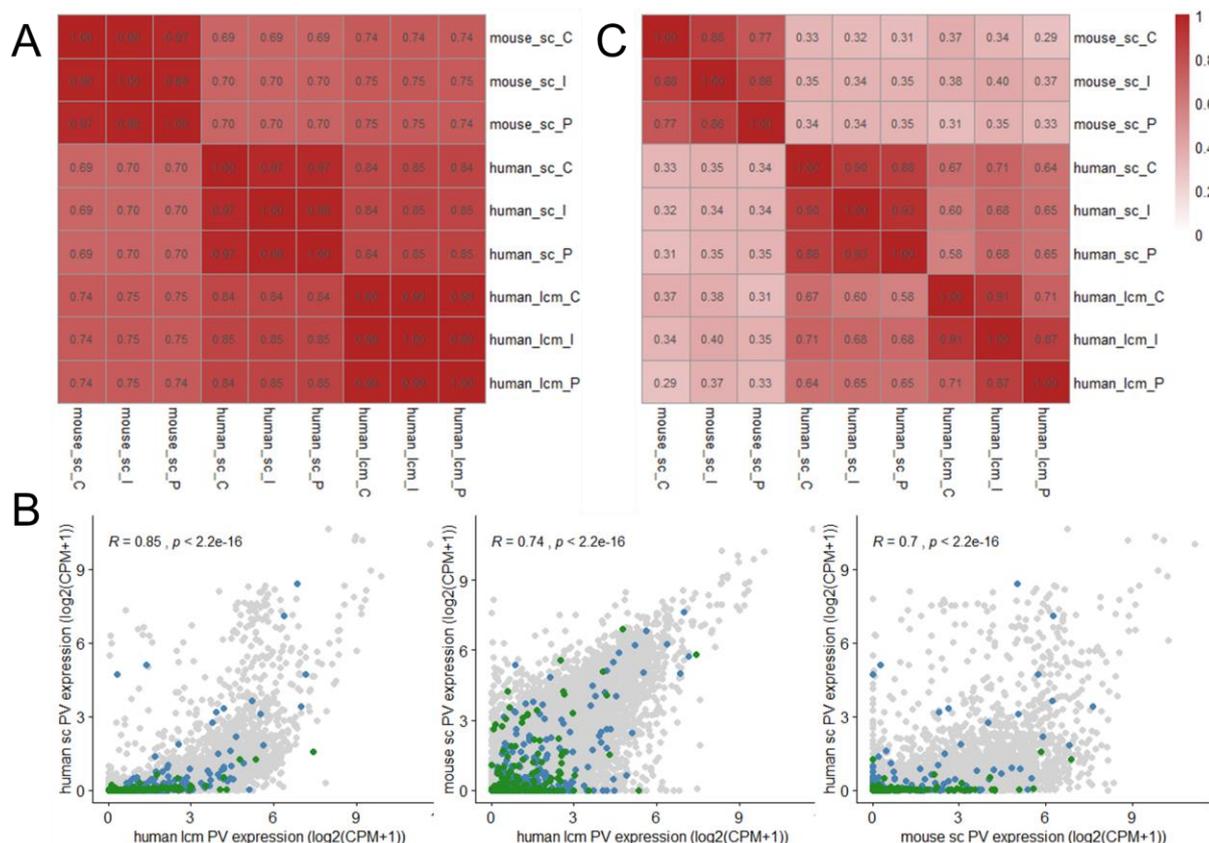
Figure 38: Comparison of zonal expression in human and mouse, detected by LCM (human) or by spatial single cell reconstruction (human and mouse). A – Spearman correlation heatmap of global one-to-one orthologue expression levels for individual zones in human LCM RNA-seq, human reconstructed scRNA-seq and mouse reconstructed scRNA-seq. Dark red corresponds to high correlation. B – Exemplary scatterplots with spearman correlation and corresponding p-values of expression levels normalized as log (CPM + 1) of human LCM RNA-seq versus human scRNA-seq (1), human LCM RNA-seq versus mouse scRNA-seq (2) and mouse scRNA-seq versus human scRNA-seq (3) in the periportal zone. Genes with significant zonation in human detected by LCM RNA-seq (|logFC > 1, FDR < 0.01) are highlighted in green (pericentral) and blue (periportal). C – Spearman correlation heatmap of one-to-one orthologues detected as zonated in the human LCM RNA-seq dataset.

Between mouse and the human LCM RNA-seq dataset 113 genes with conserved zonation signatures, including highly expressed zonation landmark genes, but importantly also morphogens such as LGR5 and SOX9, were observed. The 692 zonated genes uniquely detected in the human comprise 230 genes that do not feature one-to-one matching orthologues in mouse and 262 genes not sufficiently covered in the mouse dataset. The remaining 202 genes zonated in human, comprising important Wnt signaling pathway members like RSPO2, were not detected as zonated in the mouse. However, most non-corresponding genes showed considerably lower expression levels in the mouse single cell dataset as compared to the deeply sequenced human LCM RNA-seq dataset (Figure 39 D). Comparing zonation of spatially reconstructed mouse and human single cell datasets revealed an overlap of 546 genes, whereof 525 were

not detected as significantly zonated by the LCM RNA-seq approach and corresponding $\log_2$ fold changes of these genes did not show any indication for zonal expression differences (Figure 39 B).
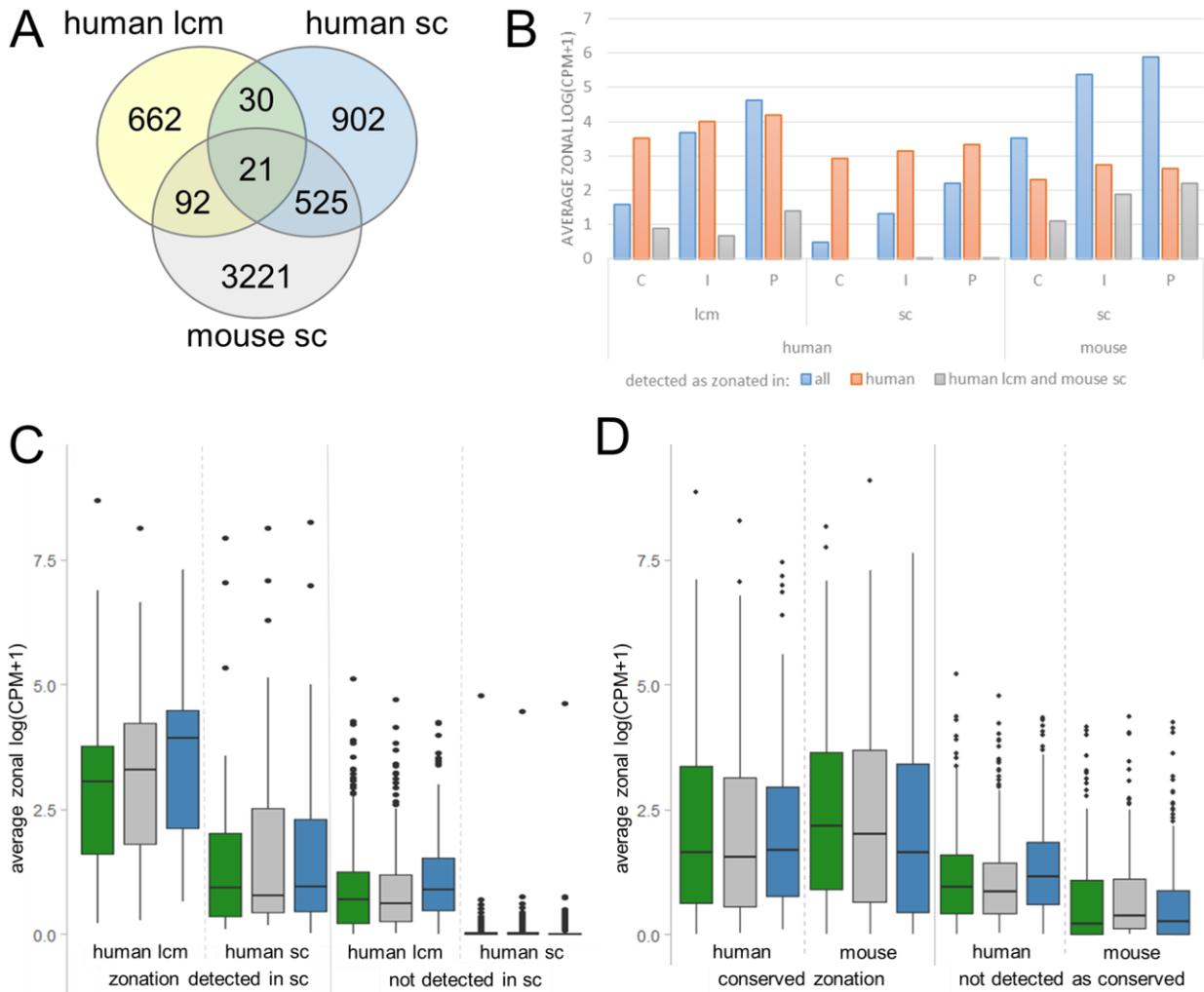
Figure 39: Comparison of zonated genes detected by LCM RNA-seq and by spatially reconstructed scRNA-seq data in human and mouse. In human LCM RNA-seq data, genes with |logFC > 1 between CV and PV (FDR < 0.01) were considered as zonated. In spatially reconstructed scRNA-seq data, genes with Kruskal-Wallis test q < 0.2 between CV and PV (human), respective between layer 1 and 9 (mouse, Halpern et al., 2017) were considered as zonated. A – Venn diagram displaying the overlap of zonated genes. B – $\log_2$ fold changes in the LCM RNA-seq dataset of genes detected in all datasets (all), only in human LCM RNA-seq and mouse scRNA-seq (human lcm & mouse sc), only in human and mouse scRNA-seq (human sc & mouse sc), only in human datasets (human lcm & human sc) or uniquely in human scRNA-seq (human sc only), mouse scRNA-seq (mouse sc only) and human LCM RNA-seq (human lcm only). C – Boxplot of gene expression levels of genes detected in both human datasets and genes zonated in LCM RNA-seq, but not in the human single cell dataset. D – Boxplot of gene expression levels of genes with conserved zonation patterns (113) and of genes detected as zonated in human by LCM RNA-seq (200), but not in the mouse single cell dataset. The pericentral zone is illustrated in green, the intermediate in grey and the periportal in blue.

Interestingly, the 3,221 genes detected uniquely as zonated in the mouse dataset comprised 65 genes that were not found as zonated in the human LCM RNA-seq dataset, even though $\log_2$ fold changes suggested zonated expression patterns (Figure 39 B). These genes were not detected as significantly zonated due to

a corresponding FDR adjusted p-value of more than 0.01, presumably caused by high inter-individual variation. Thus, with higher replicate number, which increases statistical power, these genes might be detected as being zonated in the human as well. However, the vast majority of genes uniquely zonated in the mouse did not feature any zonal expression differences in the human datasets (Figure 39 B).

In summary, these results identify an, at least, partial conservation of hepatic zonation signatures between human and mouse. Additionally, they illustrate differences between the detection of zonated gene expression in LCM RNA-seq data and spatially reconstructed scRNA-seq data, which might include an overestimation of zonated gene expression in single cell datasets.

3.3 Epigenomic and transcriptional characterization of *in vitro* differentiated hepatocyte-like cells

The epigenomic and transcriptional characterization of *in vitro* differentiated hepatocyte-like cells (HLCs) was performed in cooperation with Prof. Dr. Jan Hengstler's group at the IfADo (Dortmund) as part of the BMBF and DLR funded project StemNet. The aim of this project is the improvement of HLC differentiation towards a phenotype more closely resembling adult primary human hepatocytes (PHH), which represent the current gold standard for *in vitro* toxicological and pharmacological experiments. To gain a more profound understanding of *in vitro* differentiation as basis to develop reasonable interventions, the differentiation from iPSCs over definitive endoderm (DE) to HLCs was characterized on transcriptional level by RNA-seq, and on epigenetic level by RRBS for DNA methylation analysis and ATAC-seq for chromatin accessibility profiling (Figure 40). Resulting data were compared to corresponding PHH data. While ATAC-seq was also performed for PHH from three donors, transcriptome and methylome comparisons are based on comparable PHH data, previously generated within the German Epigenome Project (DEEP). Moreover, a complementary analysis of HLCs in comparison to PHH by deep single cell RNA-seq was included.
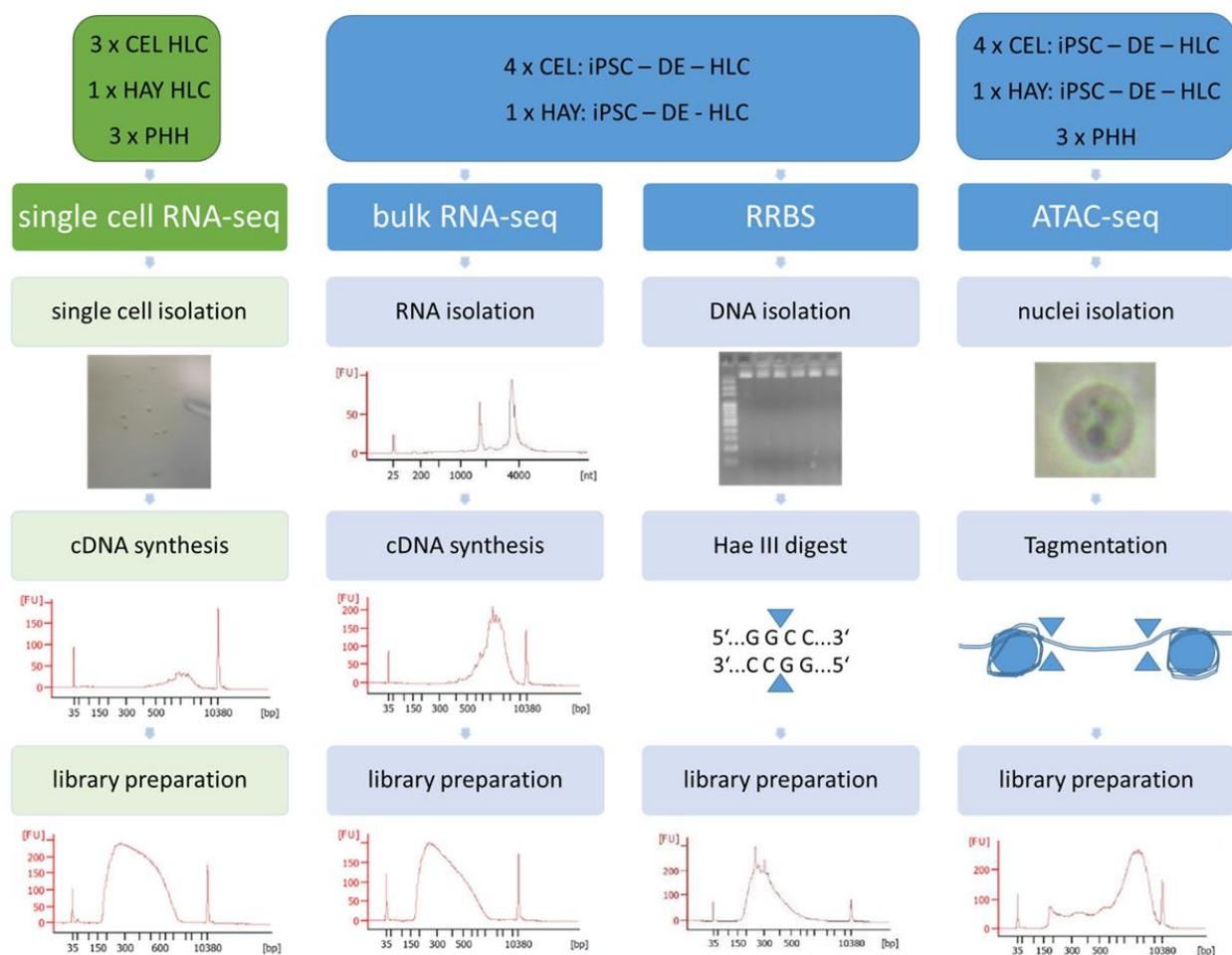
Figure 40: Workflow for the transcriptional and epigenomic characterization of *in vitro* cultivated hepatocyte-like cells in comparison to primary human hepatocytes. Bioanalyzer tracks (red) illustrate typical fragment size distributions of quality control steps for single cell RNA-seq, bulk RNA-seq, RRBS and ATAC-seq.

The project was performed in close cooperation with Patrick Nell and David Feuerborn (IfADo), who performed all cell culture experiments and provided fresh, frozen or fixed frozen cells for the individual assays. Manual isolation of single cells for scRNA-seq was achieved with kind support from Dr. Konstantin Lepikhov. Sequencing was performed by Dr. Gilles Gasparoni and processing of raw data was done by Dr. Karl Nordström (RNA-seq and scRNA-seq) and Abdulrahman Salhab (ATAC-seq and RRBS).
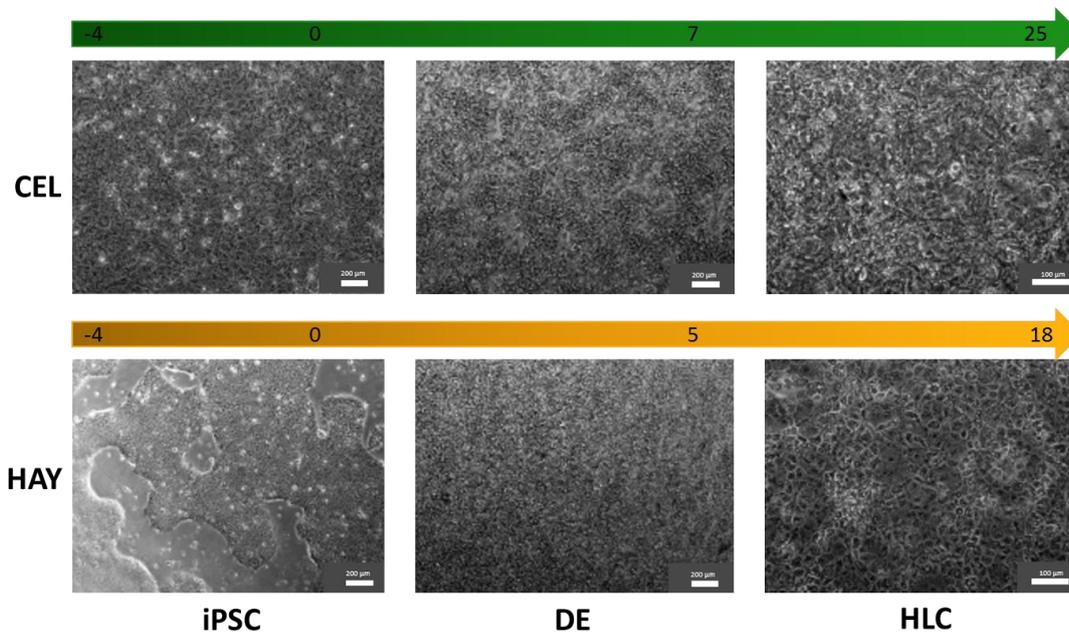
Figure 41: *In vitro* differentiation process from induced pluripotent stem cells (iPSC) to definite endoderm (DE) to hepatocyte-like cells (HLC) following the commercial CEL (Takara Cellartis) protocol and the published HAY protocol from David Hay's lab (Wang et al., 2017). The white bar represents 200 μm. Pictures were provided by Patrick Nell and David Feuerborn (IfADo). Microscopic pictures were provided by Dr. Patricio Godoy (IfADo).

We analyzed two 2D culture systems – the commercially available Cellartis iPS Cell to Hepatocyte Differentiation System (Cellartis, Takara Bio Europe AG, Goteborg, Sweden), further on referred to as CEL protocol, as well as a protocol published by David Hay's group (Wang et al., 2017), referred to as HAY protocol. The detailed components of the CEL protocol are confidential. The HAY protocol relies on human recombinant laminins as extracellular matrix (ECM) in combination with serum-free differentiation (Wang et al., 2017). In both protocols, iPSCs are pre-cultured for 4 days to achieve a pure, pluripotent stem cell population (Figure 41). This is followed by differentiation to DE for 5 (HAY) or 7 days (CEL). For differentiation to HLCs, DE cells are transferred to an ECM and cultured further for 13 (HAY) or 18 days (CEL) to obtain fully differentiated HLCs (Figure 41). Instead of using the proprietary Cellartis ECM, an optimized mixture of Laminin 521 and Laminin 111 was used as HLC differentiation matrix for both protocols. Moreover, it should be noted that CEL and HAY iPSC populations originate from different human donors, which might contribute to variability between both protocols. Based on preliminary results from the first replicate of scRNA-seq, it was decided to focus on the CEL protocol, as HAY cells apparently remained in a less differentiated state than CEL HLCs (Figure 43, Figure 44 and Figure 46). Thus, bulk transcriptomic and epigenomic analysis were performed in 4 replicates of CEL, but only 1 replicate of HAY. The scRNA-seq experiment was done in 3 replicates of CEL HLCs, 3 replicates of PHH and 1 replicate of HAY HLCs, with each replicate comprising up to 96 single cells (Figure 40).

3.3.1 Transcriptional profiling of HLC subpopulations at single cell level

Besides inter-individual and circadian variability, hepatocytes in primary liver tissue exhibit pronounced heterogeneity depending on their localization along the porto-central axis (Brosch et al., 2018). In order to gain deeper insights into the heterogeneity of *in vitro* cultured hepatocyte-like cell (HLC) populations, which was already indicated by immunohistological stainings (Figure 12), scRNA-seq was performed on 3 replicates of CEL HLCs and 1 replicate of HAY HLCs, and compared to primary hepatocytes from 3 human donors (PHH). As dissociation of HLCs did not yield pure single cell suspensions and contained high numbers of duplets, clumps and extracellular matrix particles, theses samples did neither qualify for FACS assisted single cell isolation nor for droplet-based high-throughput scRNA-seq methods. Thus, for each replicate 96 single cells were manually collected, ensuring the proper isolation of single cells. NGS libraries were prepared using a modified version of the Smartseq2 protocol (Picelli et al., 2013), yielding in total 672 single cell RNA-seq libraries. As external RNA control the Ambion® ERCC Spike-In Control, consisting of 92 polyadenylated artificial transcripts with a length of 250 to 2000 nucleotides, was used. The spike-in allows determination of the dynamic range of expression profiling in single cells and can be used to assess the general performance between experiments.
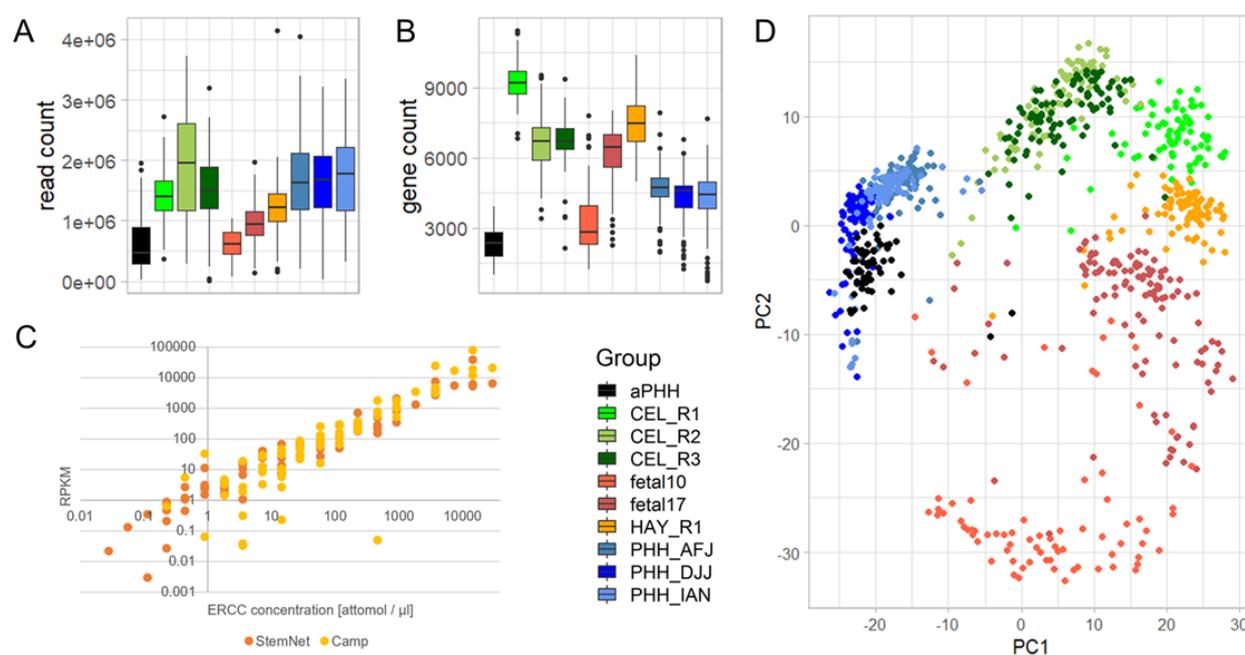
Figure 42: Quality metrics of in-house StemNet and external single cell RNA-seq datasets from Camp et al., 2017. A – Read counts. B – Number of detected genes. C – ERCC spike-in detection rate of StemNet (orange) and Camp (yellow) data, illustrated by concentrations (attomol / µl) versus average RPKM values across single cells of all 92 ERCC spike-in transcripts. D – PCA of the 1000 most variable transcripts. Cells are colored by sample annotation (black – external adult PHH (Camp et al., 2017), external fetal PHH – shades of red, CEL HLCs – shades of green, HAY HLCs – orange, PHH – shades of blue).

A recent single cell RNA-seq study by Camp et al. (2017) generated data on the Fluidigm C1 platform using a comparable Smartseq2 workflow compatible to our single cell dataset. Besides a low number of adult PHH, it comprises also 206 fetal hepatocytes of gestational week 10 and 17 derived from human abortions,

which represent a valuable additional resource for the comprehensive analysis of HLCs at single cell level. In addition, the 82 adult PHH contained in the Camp dataset were used to control for batch effects. Raw external data were downloaded from SRA and processed according to our single cell data workflow (chapter 2.2.2.1) to minimize batch effects.

Our single cell dataset was sequenced to a mean depth of 1.7 million reads per single cell, while the Camp dataset featured a mean coverage of 0.76 million reads per cell (Figure 42 A). The number of detected genes varied between datasets and also between sample groups (Figure 42 B). Internal PHH yielded an average of 4,389 covered genes per cell. Contrary, in Camp PHH a mean of 2,311 genes was detected, which might be explained by the lower sequencing depth of the Camp dataset, but could also reflect the quality of used PHHs. *In vitro* cultured HLCs generally showed higher average gene counts between 6,723 (CEL R3) and 7,991 (CEL R1) detected genes, while gene counts of fetal hepatocytes ranged between HLC and PHH samples. It should be noted that in fetal hepatocytes hardly any reads mapped to ERCC spike-in transcripts, which might, for instance, be attributed to a dilution error during sample preparation. Thus, ERCC spike-in counts were not used for data normalization, but rather as an additional quality parameter to assess comparability between the two datasets. Due to their low ERCC spike-in coverage, fetal hepatocyte samples were not considered for this particular analysis. While the detection of highly abundant spike-in transcripts was achieved with high accuracy in both datasets, the detection of low abundance transcripts was more efficient in our in-house dataset with a lowest detection limit of 0.028 attomol per µl, compared to 0.229 attomol per µl in the Camp dataset. This technical difference might, at least in part, be explained by the lower sequencing depth of the Camp dataset (Figure 42 A). Still, the overall ERCC spike-in detection rates in both datasets suggest the comparability of the expression profiling efficiency of the external and our in-house single cell RNA-seq data (Figure 42 C), therefore allowing the integrated analysis of both datasets.

During initial quality control, cells with less than 300,000 reads and or less than 1,000 detected genes were excluded, resulting in 873 high quality cells (Figure 42 A – B). In order to maintain as much information as possible for downstream analysis, only a very mild gene expression filtering criterion of an average read count above 1 was chosen, yielding 15,020 genes covered across the dataset.

In a PCA of the 1,000 most variable transcripts, single cells from our 3 internal PHH donors clustered together with external adult PHH (Figure 42 D), suggesting no major batch effects between datasets. Fetal hepatocytes of week 10 and 17 formed wide-spread clouds, indicating a rather high variability within these populations. Interestingly, CEL HLC replicates 2 and 3 were well-mixed, while replicate 1 formed an individual cluster closer to HAY HLCs. As PCA visualization would only display batch effects on the first 2 principal components, PCA regression was performed to assess batch effects on additional principal components. This identified potential batch effects on PC5, PC12 and PC15, which were however driven by few outliers.

Although the data cluster by cell-type and not by dataset, a more sensitive batch effect assessment was performed using a k-nearest neighbor batch effect test (kBET, Büttner et al., 2019, Sup. Figure 12 A), which is based on Pearson's chi squared test between replicates. Indeed, this analysis yielded high rejection rates in the combined dataset (Sup Figure 12 A), suggesting the presence of batch effects. However, kBET is very sensitive to any kind of bias, including the variation between CEL HLC replicates, which are most likely not attributed to technical artifacts, but rather to a lesser degree of differentiation of CEL HLC replicate 1

(Figure 43). Thus, the average silhouette width, which estimates the average distance between batches, was calculated in addition. Together with the PCA based visualization of the dataset (Figure 42 D and Sup. Figure 12 B), the low silhouette score of 0.199 suggests, at most, minor batch effects between the external Camp dataset and our StemNet dataset. As even small batch effects might impact downstream data analysis, several batch correction methods were applied, and corrected data were evaluated again by visualization of dimensionality reduction, kBET and silhouette width (Sup. Figure 12). Mutual nearest neighbor correction (MNN) was considered as the most appropriate approach for our single cell dataset, as it does not rely on equal population composition and requires only a subset of phenotypes to be shared between batches (Haghverdi et al., 2018). Thus, MNN correction was anchored on adult PHH, the only phenotype shared between both datasets. It yielded slightly changed clustering with fetal hepatocytes of week 17 cells overlapping the CEL HLC R1 cluster, though clustering of adult PHH appeared to be very similar to clustering without any batch correction (Sup. Figure 12 C). Although average silhouette width was slightly reduced, observed rejection rates were even higher than before batch correction (Sup. Figure 12 A). Hence, MNN correction apparently did not improve batch effects in the dataset. Contrary, batch correction with ComBat, which is based on empirical Bayes frameworks for correction of known covariates, yielded both reduced rejection rates as well as the smallest average silhouette width of all tested normalization methods (Sup. Figure 12 A). While adult PHH still clustered together similarly as before correction, especially fetal hepatocytes of gestational week 10 formed a denser cluster partially overlapping CEL HLC replicate 2 and 3. In addition, CEL HLC replicate 1 and HAY HLC cells formed an overlapping cluster (Sup. Figure 12 D). However, due to the unbalanced distribution of phenotypes within both batches, ComBat correction might be prone to overestimation of batch effects (Nygaard et al., 2015). The used workflows for downstream analysis including pseudotime reconstruction and TF network activity analysis intrinsically handle data normalization (monocle2) or rely on a robust scoring system (SCENIC), which is able to overcome batch effects (Aibar et al., 2017). Thus, it was decided to proceed without batch correction for further analysis.

To obtain a first impression of cellular heterogeneity, normalized expression levels of several marker genes were visualized (Figure 43). The hepatocyte lineage marker HNF4α (Odom et al., 2004) was highly abundant in in-house PHH as well as in both CEL and HAY HLCs. However, it was not detected in a subset of external adult PHH, which, in line with the lower number of detected genes (Figure 42 B), indicates a higher drop-out rate in the Camp dataset. Moreover, HNF4α was not expressed in fetal hepatocytes of gestational week 10, but was detected at low levels in a subpopulation of week 17 fetal hepatocytes (Figure 43 A). The colon marker CDX2 drives intestinal endoderm development, is thus highly expressed in colon (Stringer et al., 2012), and should not be expressed in hepatocytes. Indeed, it was not detected in fetal and adult PHH, but subpopulations of both CEL and HAY HLCs expressed CDX2 in variable ranges (Figure 43 B). Albumin (ALB), the most abundant transcript in mature hepatocytes (Sup. Figure 1 A), was detected at high levels in all PHH cells. Interestingly, ALB was only expressed in a small subpopulation of CEL HLC replicate 1 and HAY HLCs. In contrast, ALB expression was detected in all single cells of CEL HLC replicate 2 and 3, albeit expression levels were mostly not as high as in primary hepatocytes, but rather resembled ALB expression levels in fetal hepatocytes (Figure 43 C). Moreover, the hepatoblast marker α-fetoprotein (AFP) (Elmaouhoub et al., 2007), which was only detected in a small subset of adult PHH, was even more abundant

in HLCs than in fetal hepatocytes (Figure 43 D). Taken together, marker gene expression levels suggest that in some aspects HLCs resemble fetal hepatocytes more closely than mature PHH, and in addition might display gene expression programs not desired in hepatocytes. Besides a considerable intra-sample heterogeneity, there is also variability between HLC replicates that suggest both CEL HLC replicate 1 and HAY HLCs to be less differentiated than CEL HLC replicates 2 and 3.
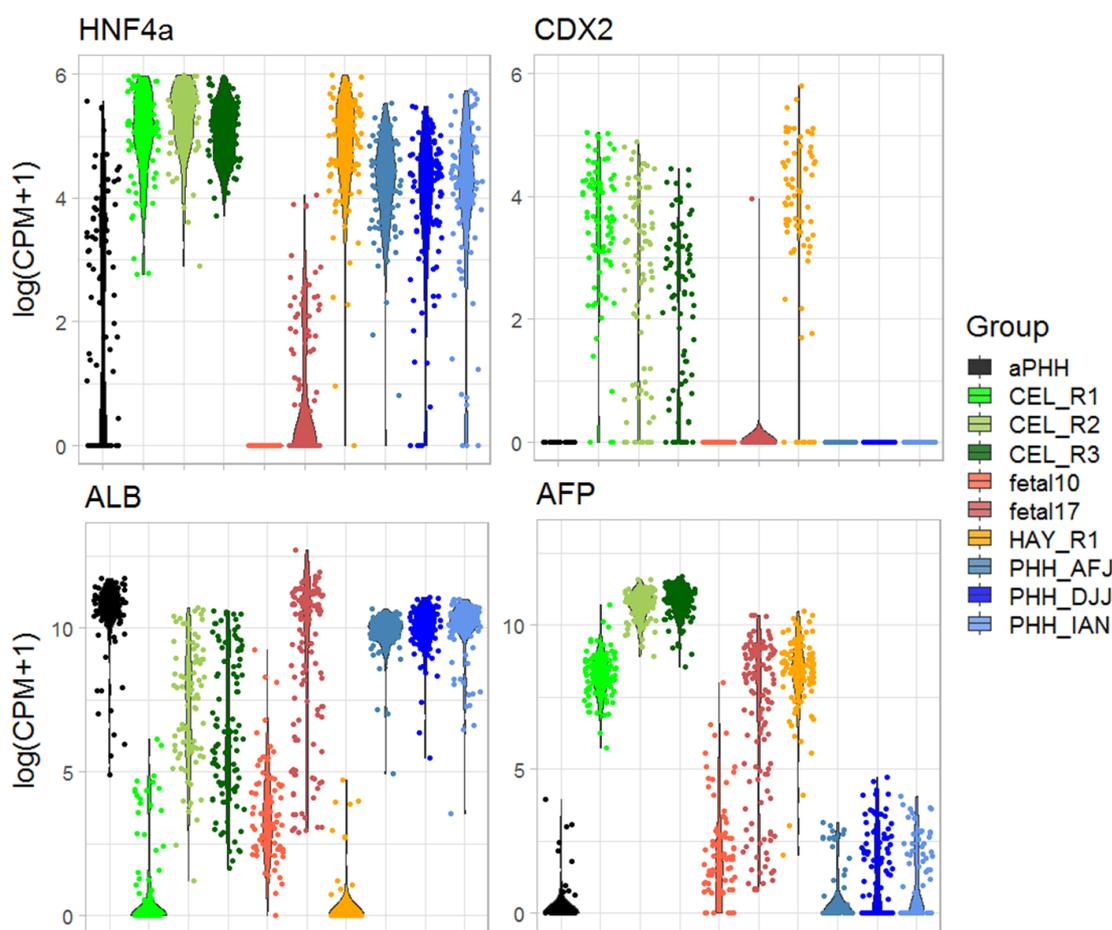
Figure 43: Single cell expression levels (normalized as log (CPM + 1)) of the marker genes HNF4α (A), CDX2 (B), ALB (C), and AFP (D). Cells are colored by sample annotation (black – external adult PHH (Camp et al., 2017), external fetal PHH – shades of red, CEL HLCs – shades of green, HAY HLCs – orange, PHH – shades of blue).

Pseudotime analysis of single cells represents a powerful tool to reconstruct cellular lineages and to track differentiation processes. To this end, pseudotemporal ordering of single cells by monocle2 (Trapnell, 2017) was applied in a semi-supervised mode, in which marker genes are used as anchors for cell clustering. Thus, the clustering is not simply based on highly variable genes, but specifically on genes that co-vary with the marker genes, making the approach more robust against drop-out effects. Using CDX2, ALB and AFP as anchors for pseudotime inference identified 6 individual differentiation states (Figure 44 A). State 1, 4 and 5 are fetal hepatocytes of week 10 and 17. State 2 refers to CEL HLCs (R1, R2 and R3), while state 3 marks HAY HLCs. Both in-house and external PHH form a dense cluster as terminally differentiated state 6 (Figure 44 A – B). State 1 is detected as the root state (Figure 44 A). The first major branching event (1) is the separation between *in vitro* differentiation of HLCs and *in vivo* differentiation of primary hepatocytes. The

in vitro branch segregates then further into CEL and HAY HLCs (2). As already suggested by PCA and marker gene expression, CEL HLC replicate 1 builds a separate cluster at the top of the *in vitro* branch, which suggests a less mature differentiation state. The *in vivo* branch from fetal to adult primary hepatocytes shows a linear differentiation, but also an additional branch (3) with a subset of fetal hepatocytes deviating from the major branch. This subpopulation consists mainly of fetal hepatocytes of gestational week 10, which express considerable levels of fetal liver erythropoiesis markers such as KLF1 and TAL1 (Sup. Figure 13 A – B). Moreover, it should be noted that in pseudotemporal ordering fetal hepatocytes of week 17 appeared to be less mature than fractions of week 10 hepatocytes and showed a considerable heterogeneity. According to Camp et al. (2017), week 17 fetal hepatocytes were rather clumpy and contained many blood cells. Therefore, these cells were cultured for 12 hours to obtain clean fetal hepatocytes (Camp et al., 2017), which might have impacted gene expression profiles and could explain why these cells do not appear more differentiated than week 10 fetal hepatocytes. In summary, pseudotemporal clustering revealed two major branches, segregating *in vivo* and *in vitro* differentiation, which suggests differences in the underlying gene regulatory networks.
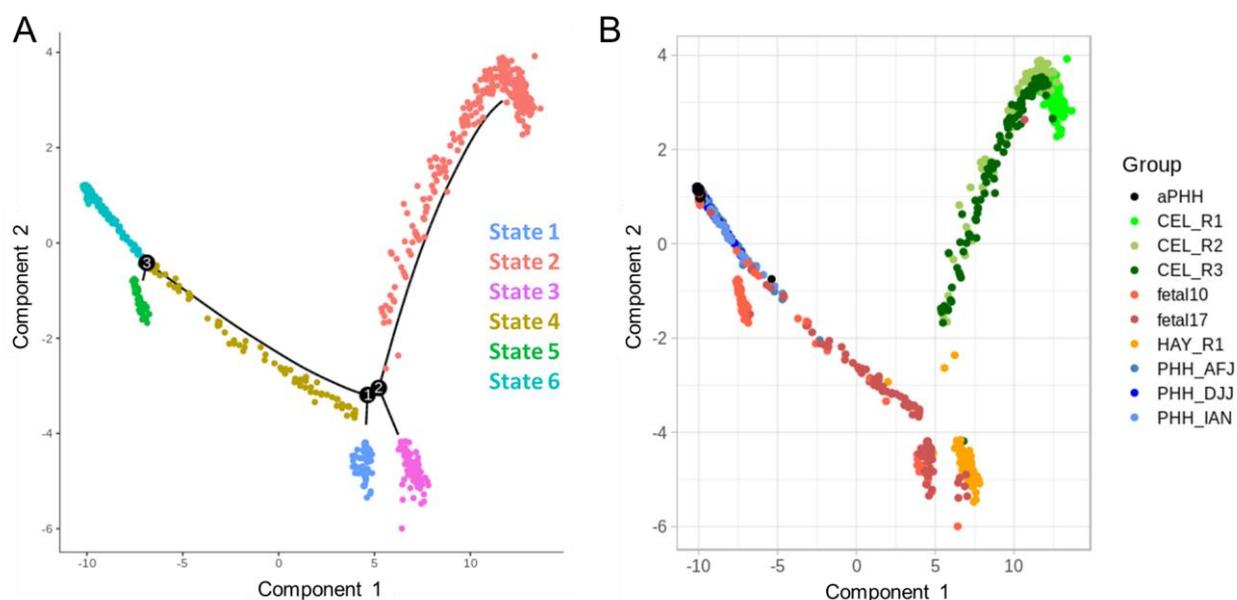
Figure 44: Pseudotemporal ordering of single cells using monocle2 in semi-supervised mode with CDX2, ALB and AFP expression as anchors for cell clustering. A – Reduced dimension map with correlation spanning tree (black) with 3 branching events. In total, 6 different states were identified in the single cell dataset. B – Annotation of pseudotemporal ordering by groups.

In order to pursue this hypothesis further, single cell regulatory network inference and clustering (SCENIC) analysis of HLCs, fetal hepatocytes and adult PHH was carried out as previously described (Aibar et al., 2017). In brief, modules of transcription factors (TFs) and corresponding co-expressed genes were determined to identify potential TF target genes. In order to remove indirect target genes, modules were filtered for those genes that were associated with significant binding motive enrichment of the respective TF. The resulting modules, each consisting of an upstream TF and its putative direct target genes, are further on referred to as regulons. Subsequently, regulon activities for all TFs expressed in the dataset were estimated by integrating expression ranks of all genes in a regulon, yielding single cell activity scores for

each regulon. Binarized activity scores were then used for hierarchical clustering of regulons (Figure 45 A) and correlation based t-SNE (Figure 45 B) to identify regulons which discriminate cellular subpopulations.
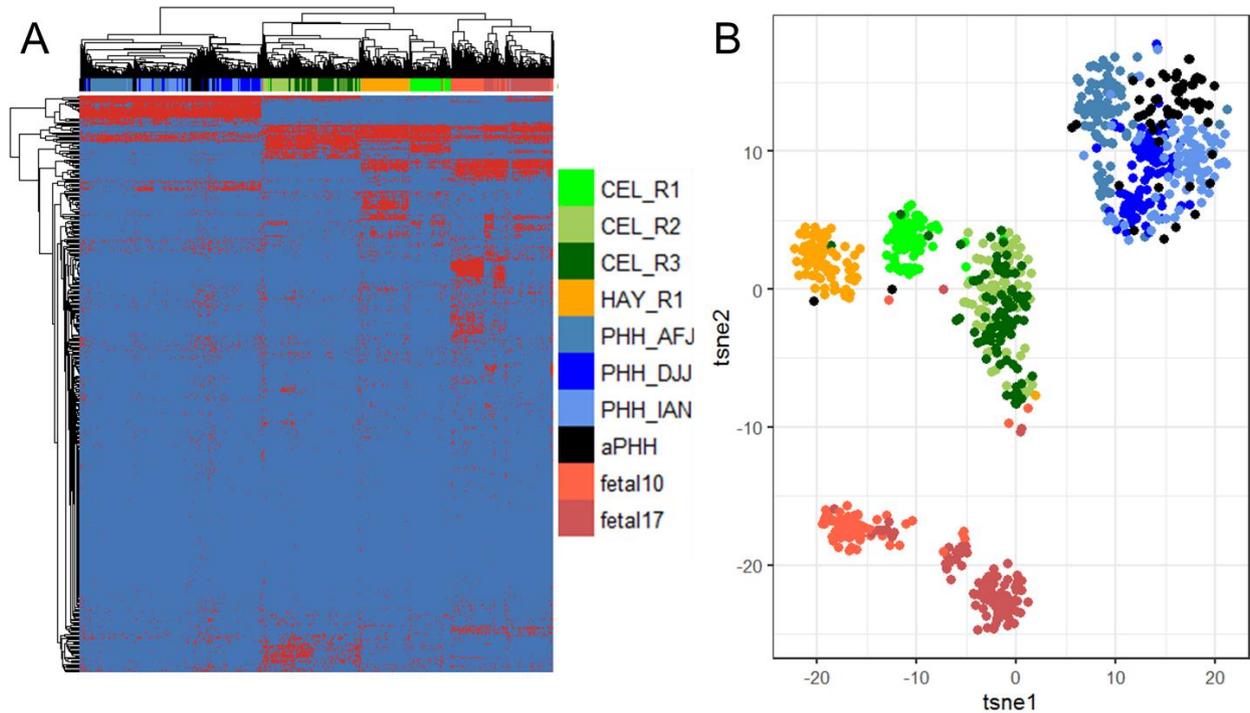
Figure 45: Regulon activity analysis. A regulon is defined as a transcription factor with corresponding genes. A – Binary regulon activity heatmap rows clustered by manhattan distance with active regulons displayed in red. B – Correlation distance based tSNE plot of regulon activity levels, colored by cell annotation as described in Figure 47.

In total, 451 TF regulons with variable activity in the dataset were observed (Figure 45 A). However, it should be noted that the majority of regulons only featured significant activity in few cells, not yielding any relevant information. Thus, only the 176 regulons which showed significant activity in more than 30 % of single cells in at least 1 group were considered. The visualization of t-SNE embeddings of binarized regulon activities shows clear separation of cell-types (Figure 45 B). All adult PHH derived from 4 individual donors cluster together, suggesting no major contribution of batch effects in this analysis. Fetal hepatocytes segregate by gestational week, although a subpopulation, which consists mostly of week 17 hepatocytes, clusters separately. In addition, CEL HLC R1, which is characterized by absent ALB expression in most cells (Figure 46), forms a distinct cluster between HAY HLCs and CEL HLC R2 and R3, confirming the less differentiated state of CEL HLC R1 already suggested by pseudotime analysis (Figure 44). Contrary, CEL HLC R2 and R3 cells are well-mixed, highlighting their similar differentiation state also on basis regulon activity. Interestingly, HLC clusters share more active regulons with fetal hepatocytes (28 regulons) than with adult PHH (14 regulons), which indicates a more fetal-like phenotype in HLCs.

Indeed, 24 regulons, comprising among others PXR (NR1|2), CAR (NR1|3), RXRA, and RXRB, which are known regulators of a multitude of hepatocyte specific functions (Hardwick et al., 2009; Tien and Negishi, 2006), were detected as active uniquely in PHH (Figure 46). Interestingly, FXR (NR1H4) and PPARA were identified to be active both in ALB expressing HLCs and adult PHH, but not in HLC CEL R1 or HAY HLCs. FXR was additionally active in a subpopulation of fetal17 cells, which also expressed ALB. Contrary, CDX2

regulon activity was only detected in *in vitro* cultured cells (Figure 46). As ALB is a major hepatocyte marker, Spearman correlations between ALB expression and regulon activities were calculated in order to identify regulons which might contribute to a more favorable HLC differentiation. In total, the activity of 29 regulons correlated with ALB expression levels in the dataset (r > 0.5, Sup. Table 13), confirming major candidates for the regulation of a hepatocyte-like cell fate such as PXR (r = 0.81), CAR (r = 0.81), FXR (r = 0.67), PPARA (r = 0.61) and RXRA (r = 0.57).
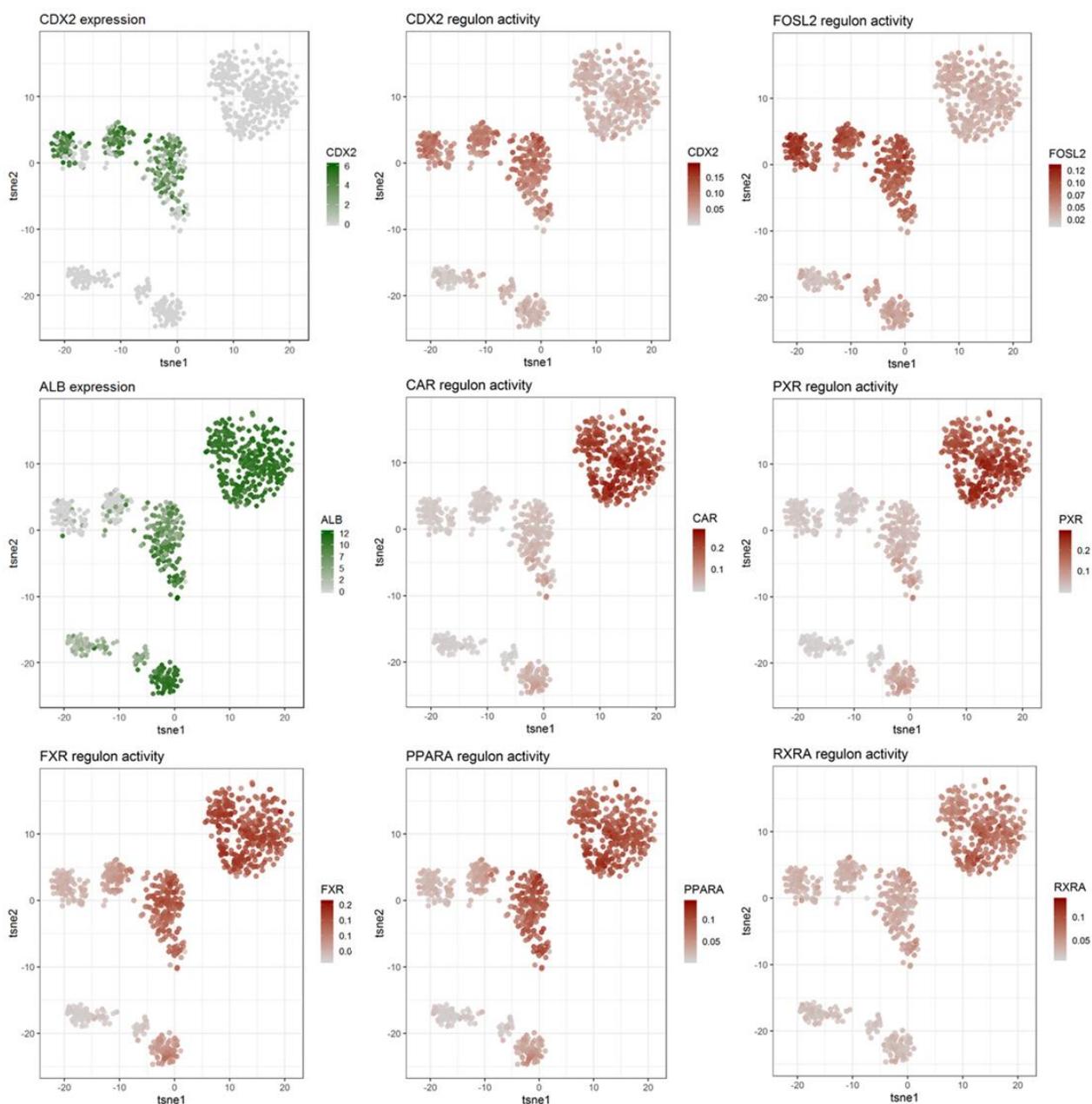
Figure 46: Activity of selected regulons in relation to ALB and CDX2 expression at single cell level. Expression levels are colored in green, regulon activities in red. For cell-type annotation refer to Figure 45 B.

In addition, Spearman correlations between regulon activities and colon marker CDX2 as well as hepatoblast marker AFP revealed major putative regulators of unfavorable expression signatures in HLCs. CDX2 expression correlated strongest with CDX2 regulon activity (r = 0.73, Figure 46), validating the reasonability

of the correlation based approach. While most regulons showed similar correlation with CDX2 and AFP expression (Sup. Table 13), SOX9 regulon activity was exclusively identified as a major regulator of AFP positive hepatoblast signatures (r = 0.64), which is in line with SOX9 being a marker for bipotent liver progenitor cells with the potential to differentiate into hepatocytes or cholangiocytes (Furuyama et al., 2011; Tarlow et al., 2014). Further TFs which might be involved in the regulation of unfavorable expression patterns were among others FOSL2 (Figure 46, r_{AFP} = 0.68, r_{CDX2} = 0.71), KLF5 (r_{AFP} = 0.63, r_{CDX2} = 0.69), KLF6 (r_{AFP} = 0.77, r_{CDX2} = 0.67), and SOX4 (r_{AFP} = 0.67, r_{CDX2} = 0.65). These factors are not described to play a role in hepatocyte differentiation, but seem to be involved in other cellular lineages. For instance, KLF5 is associated with a colon phenotype (Liu et al., 2017), FOSL2 was reported to play a role in colon cancer metastasis (Li et al., 2018), and SOX4 regulates bile duct development (Poncy et al., 2015). Thus, downregulation of these TFs might have a beneficial effect on HLC differentiation. Taken together, this approach revealed several TF regulons that might play a major role in the differentiation process of HLCs, and thus represent interesting targets for interventions.

In summary, the combination of pseudotemporal inference and TF regulon activity analysis of the single cell RNA-seq dataset revealed the less differentiated state of HLCs, rather resembling fetal hepatocytes than adult PHH. Moreover, a set of several major candidates such as the nuclear receptors PXR, CAR and FXR, which might be decisive for proper hepatocyte cell fate specification, was identified. However, the number of candidates for cell culture interventions, such as knock-down of unfavorable TFs or over-expression of PHH specific TFs, is still rather high (Sup. Table 13). Thus, the additional integration of epigenetic data was used to provide more insights into the regulation of *in vitro* differentiation, which would allow a better informed choice of the most promising regulators for cell culture interventions.

3.3.2 Epigenomic characterization of the HLC *in vitro* differentiation process

In addition to single cell RNA-seq of HLCs and PHH, the bulk transcriptome during differentiation from iPSCs to DE to HLCs was assessed by RNA-seq of 4 CEL replicates and 1 HAY replicate. Libraries were generated by a modified Smartseq2 protocol for high input of isolated RNA (chapter 2.2.1.4), and sequenced to a mean depth of 69.2 million reads, covering in total 37,664 genes (Sup. Table 10). In addition, RNA-seq data of 3 male and 3 female primary hepatocyte samples, generated within the DEEP project (http://epigenomesportal.ca/tracks/DEEP, EGA accession: EGAD00001002527), were used for comparison. In a PCA of the 1,000 most variable transcripts in the dataset, PC1 apparently represents the differentiation state, while PC2 might reflect a cultivation effect over time (Figure 47 A). Both iPSC and DE samples cluster together, while CEL HLCs and PHH form separate clusters, indicating major expression differences between DE and HLCs, but also between HLCs and PHH. Moreover, the HAY HLC sample clusters more closely to DE samples than to CEL HLCs, indicating a less mature differentiation state of HAY HLCs, which is in line with pseudotime analysis of the single cell dataset (Figure 44). Hierarchical clustering by euclidean distance of expression z-scores of all expressed genes in the dataset suggests massive expression differences during *in vitro* differentiation from iPSCs to DE to HLCs, between CEL and HAY protocols, and also between terminally differentiated HLCs and PHHs. As already suggested by PCA, HAY HLCs appeared to be more similar to DE than to CEL HLCs. Replicates clustered together, indicating a high reproducibility. However, it

should be noted that for some differentiation marker genes variability between CEL HLC replicates was observed (Sup. Figure 14), which is not reflected on a genome-wide scale (Figure 47 B), but indicates variation in the differentiation stage of individual CEL HLC cultures. As the apparently less promising HAY protocol was performed only in 1 replicate, HAY samples were not considered for more detailed analysis.
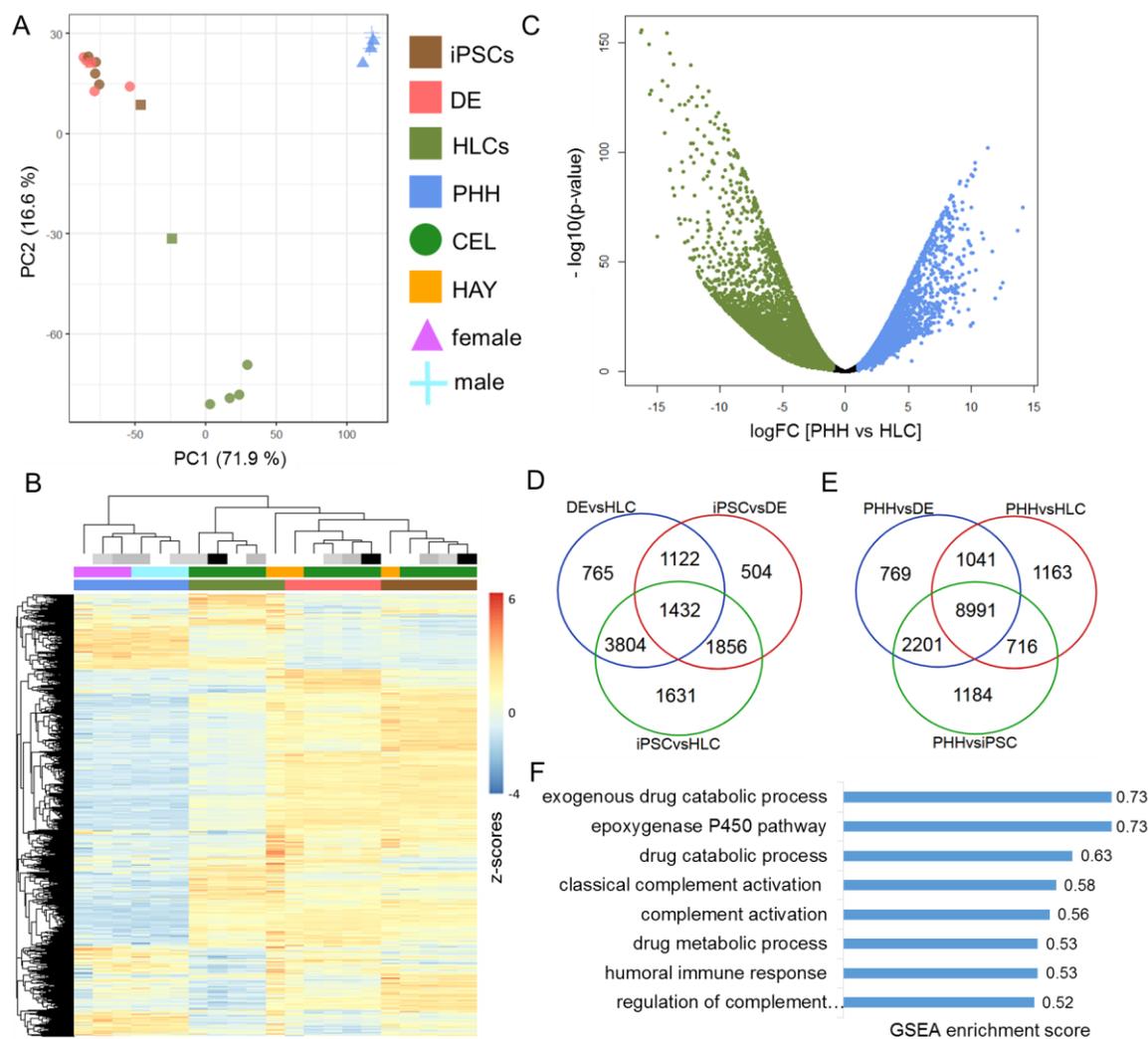
Figure 47: Transcriptome analysis by bulk RNA-seq reveals major expression differences during *in vitro* HLC differentiation and between HLCs and PHH. iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue. The annotation legend with corresponding symbols displays CEL samples (n = 4) in green, HAY samples (n = 1) in orange, female PHH (n = 3) in pink and male PHH (n = 3) in light blue. A – Principal component analysis of the 1000 most variable transcripts. B – Expression z-scores of genes with average CPM > 0.5 clustered by euclidean distance. Replicates are annotated in white, light grey, grey and black. C – Exemplary volcano plot for differential analysis between PHH and CEL HLCs using edgeR. Genes upregulated in PHH (logFC > 1, FDR < 0.01) are displayed in blue, genes upregulated in HLCs (logFC < -1, FDR < 0.01) in green. D and E – Overlap of significant DEGs (|logFC| > 1, FDR < 0.01, average logCPM > -2.2) detected by edgeR during CEL differentiation process from iPSCs to DE to HLCs (D), respective between CEL samples and PHH (E). HAY samples were excluded from differential analysis. F – Gene set enrichment analysis (Kolmogorov-Smirnov test, enrichment score > 0.5, Benjamini-Yekutieli adjusted p-value < 0.01) of genes upregulated in PHH compared to *in vitro* differentiated HLCs.

Differential expression analysis of CEL differentiation confirmed major expression differences (Figure 47 C). Out of 18,799 expressed genes in the dataset, 15,375 were differentially expressed throughout *in vitro* CEL differentiation (Figure 47 D), and 16,152 DEGs were detected when comparing CEL iPSC, DE and HLC samples with PHH (Figure 47 E). Gene set enrichment analysis (GSEA) of genes upregulated in PHH in comparison to CEL HLCs (Figure 47 H) yielded hepatocyte-specific terms including *exogenous drug catabolic process* and *epoxygenase P450 pathway*, but also immune system related processes known to be important in liver such as *complement activation* (Qin and Gao, 2006). Interestingly, GSEA of genes upregulated in HLCs did not result in significant enrichment of any particular biological processes, suggesting a rather unsystematic upregulation of gene expression in HLCs.

Even though CEL HLC expression profiles seem to emulate PHH transcriptomes (Figure 43), they express most hepatocyte-specific genes at levels significantly lower than in PHH (Figure 47 B). Moreover, they partially maintain the expression of stemness markers such as KLF4 and TBX3, and also express aberrant genes such as colon-specific CDX2 and KLF5. Taken together, these findings support the single cell transcriptome based observation that *in vitro* differentiated CEL HLCs remain in a less mature differentiation state than PHH.

Transcriptome analysis of the CEL *in vitro* differentiation process revealed an immense gene expression difference in HLCs compared to PHH (Figure 47 B). Thus, epigenetic profiling of DNA methylation and chromatin accessibility was carried out on iPSCs, DE and HLCs. The obtained data were then compared to the epigenome of PHHs in order to gain a better understanding of the in the *in vitro* differentiation process. DNA methylation analysis of the differentiation process was achieved using HaeIII based RRBS covering in average 7 million CpGs per sample with a mean depth of 11.5 (Sup. Table 11). In addition, PHH DNA methylation data generated by whole genome bisulfite sequencing (DEEP: http://epigenomesportal.ca/tracks/DEEP, EGA accession: EGAD00001002527) were downsampled to the regions sufficiently covered also in the RRBS dataset.

PCA of the 5,000 most variable CpGs (Sup. Figure 15 A) showed a similar clustering as the PCA on the transcriptome data (Figure 47 A), with iPSC and DE clustering closely together and CEL HLCs and PHH forming distinct clusters. The HAY HLC sample clustered between CEL HLCs and DE samples, indicating a less differentiated phenotype also on the level of DNA methylation. Intriguingly, differential analysis using Methylkit revealed major DNA methylation differences in the dataset (meth.diff > 20 %, FDR < 0.01, Figure 48 A). Throughout the CEL *in vitro* differentiation from iPSCs to DE to HLCs, in most DMRs the level of DNA methylation is increasing, with the highest overall DNA methylation level in HLCs. In addition, up to 10 fold more hypomethylated regions were detected in PHH than in cultured cells (Figure 48 A). The majority of DMRs of iPSCs, DE and HLCs versus PHH overlap (Figure 48 B), indicating systematic regulatory differences between *in vivo* derived PHH and the artificial *in vitro* HLC differentiation system, which result in pronounced DNA methylation differences between HLCs and PHH (Figure 48 C). As the restriction enzyme HaeIII was used for RRBS library construction, which leads to an enrichment of CpG-rich sequences associated to putative regulatory regions and gene bodies in the resulting sequencing library (Martinez-Arguelles et al., 2014), the majority of DMRs was annotated to a gene context (Figure 48 D). Interestingly, DMRs hypermethylated in HLCs were more frequently localized in intergenic regions (40%), which might indicate a random hypermethylation effect not related to the regulation of gene expression.

Figure 48: Differential DNA methylation analysis by RRBS reveals pronounced DNA methylation changes during *in vitro* HLC differentiation and between HLCs and PHH. A – Volcano plots of relevant DNA methylation comparisons. The x-axis visualizes the methylation difference (%), while the y-axis depicts negative log10 of FDR adjusted p-values. Significant DMRs (|Δ methylation| > 20%, FDR < 0.01) are colored by the hypomethylated group: iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue. Corresponding DMR counts are depicted as well. B – Overlap of significant DMRs during CEL differentiation process from iPSCs to DE to HLCs and between CEL samples and PHH. C – Heatmap of average CpG methylation levels in DMRs between HLCs and PHH that are also covered in DE and iPSC samples. Red corresponds to high methylation, blue to unmethylated regions. D – Genomic annotation of DMRs between HLCs and PHH.

In addition, a genome segmentation into large domains, defined by different overall DNA methylation levels as previously described by Salhab et al. (2018), revealed also major differences on a genome-wide scale. In this context partially methylated domains (PMDs), which are characterized by reduced average DNA methylation levels, are of particular interest, as changes in PMDs are hallmarks of cell differentiation (Salhab et al., 2018). While iPSCs feature the highest number of PMDs, it drops towards DE and is lowest in HLCs, which show a strong gain in highly methylated domains (Sup. Figure 15 B). It should be noted that the number of PMDs in fully differentiated PHH ranks between DE and HLCs. Moreover, the average DNA methylation level in PMDs decreases from iPSCs to HLCs, and is lowest in PHH (Sup. Figure 15 C). These observations support the pronounced hypermethylation effect during HLC *in vitro* differentiation detected by differential analysis on local level (Figure 48 A), and indicate an undirected deregulation of DNA methylation also on a genome-wide level.

In summary, DNA methylation analysis revealed an increasing hypermethylation from iPSCs to DE to HLCs as well as a pronounced hypermethylation of cultured cells in comparison with PHH, both on local and global level. Intriguingly, *in vitro* cultured cells showed a pronounced transcriptional upregulation of the DNA methyltransferases DNMT1, DNMT3A and DNMT3B as well as of the methyl-cytosine dioxygenase TET1 (Sup. Figure 16). Expression levels were highest in iPSCs and decreased over DE to HLCs (Sup. Figure 16), which is in line with the progressive hypermethylation observed throughout *in vitro* differentiation (Figure 48 A).

Chromatin accessibility profiling by ATAC-seq was performed in 4 CEL replicates and 1 HAY replicate. Within the DEEP project PHH chromatin accessibility was analyzed by DNase-seq, which is a similar approach for chromatin accessibility profiling. However, a detailed comparison between ATAC-seq and DNase-seq revealed assay specific features (Nordström et al., 2019). Thus, ATAC-seq was additionally performed on PHH from 3 male donors (DJJ, AFJ and IAN) to avoid technical bias in differential chromatin accessibility analysis. Libraries were sequenced to a mean depth of 49 million reads (Sup. Table 12). In general, CEL and HAY samples had acceptable ratios of mitochondrial reads (Sup. Table 12), which resulted in 11 to 31 million uniquely mappable reads. Most samples featured rather high fraction of reads in peaks scores (FRiP score, in average 0.17, Sup. Table 12), which indicate low background signals across the genome. Only the first replicate of CEL HLC was excluded due to high background, which impaired proper peak calling (7,344 peaks, 0.016 FRiP). It should be noted that PHH samples exhibit a substantially higher amount of mitochondrial contamination, probably due to less stringent washing of the very fragile hepatocyte nuclei (chapter 2.2.1.5), which resulted in higher duplication rates and ultimately in a low amount of uniquely mappable reads (4 to 7 million, Sup. Table 12). Only up to 45,610 peaks were called using MACS2 in PHH. However, at least for donors DJJ and AFJ the fraction of reads in peaks score (FRiP score, Sup. Table 12) was in an acceptable range, and previous PHH DNase-seq data (DEEP) yielded similar peak counts. Thus, despite their low number of uniquely mappable reads, PHH DJJ and AFJ ATAC-seq data were of sufficient quality. PHH donor IAN only yielded 14,639 peaks and a low FRiP score (0.049), which indicates a rather high background in this sample. However, it was reasoned that for subsequent differential analysis, which includes a background normalization step, using the third replicate would still yield superior results than only using two replicates. Thus, only CEL HLC R1 was excluded from further analysis.
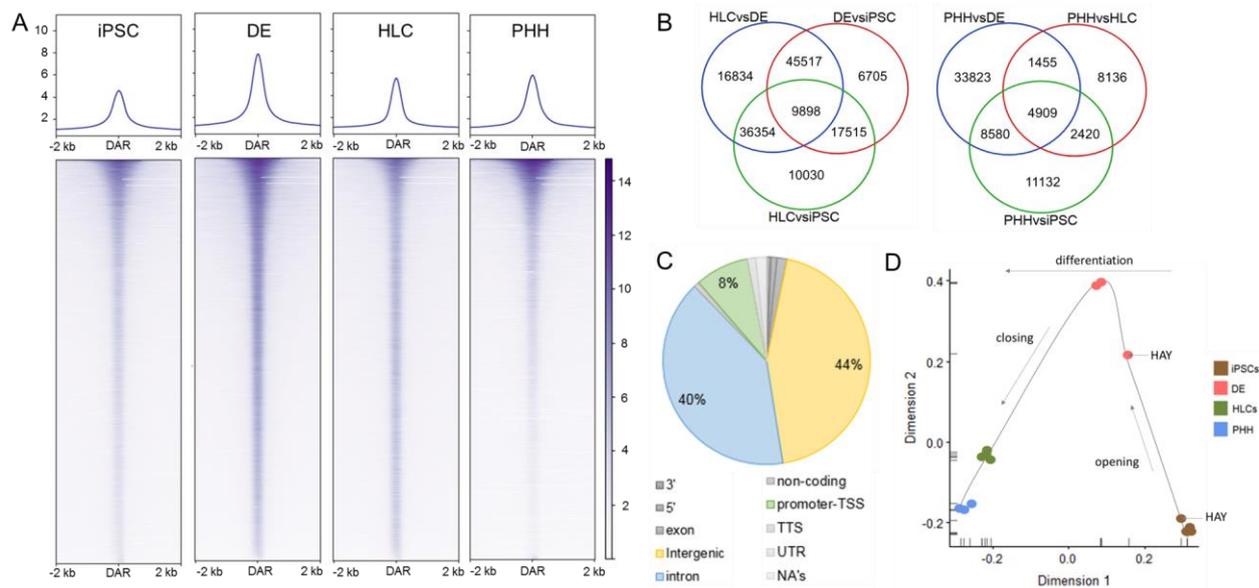
Figure 49: Differential accessibility analysis reveals major changes of chromatin structure during *in vitro* HLC differentiation. A – Heatmap illustrating the sequencing depth normalized ATAC-seq coverage of representative samples in a ± 2 kb window around the centers of all differentially accessible regions (DARs, |logFC| > 1, FDR < 0.01) detected in the dataset using csaw as outlined in chapter 2.2.2.2.3. HAY samples were excluded from differential analysis. Purple implies open chromatin, missing values or zero coverage are displayed in white. B – Overlap of significant DARs (|logFC| > 1, FDR < 0.01, average logCPM > -2.2) during CEL differentiation process from iPSCs to de to HLCs, respective between CEL samples and PHH. C – Genomic annotation of DARs. D – Diffusion map of chromatin accessibility based on log (CPM + 1) values reconstructing the CEL differentiation process. PHH are colored in blue, HLCs in green, DE in red and iPSCs in brown. HAY samples are marked by arrowheads.

PCA of the 50,000 most variable 100 bp bins across the genome showed distinct clusters of iPSC and DE samples, but more widespread clustering of HLCs and PHH, indicating higher variability within these sample groups (Sup. Figure 17 A), which was also observable on the level of detected peak counts (Sup. Figure 17 B). Overall, DE samples featured the highest number of peaks (86,897– 123,399), followed by HLCs (32,561 – 103,601) and iPSCs (44,215 – 64,965), while PHH had the lowest number of peaks. There were no major differences across samples regarding peak length, which averaged around 392 bp (Sup. Figure 17 C).

As can be taken from Sup. Figure 19, mere comparison of called peaks between samples is not conclusive. For example, the chromatin structure upstream and along the gene body of albumin (ALB) is more accessible in PHH, while in HLC there is only little ATAC-seq signal. The MACS2 peak calling identifies only few narrow peaks in the PHH samples, while manual assessment rather suggests a broad accessible region spanning the entire ALB gene. Thus, differential analysis using csaw (Lun and Smyth, 2016) was not performed based on ATAC-seq peaks, but rather on the sum of read counts in 100 bp windows with merging of differentially accessible windows, allowing different sizes of differentially accessible regions (DARs). Moreover, read counts in 10,000 bp windows were used for background normalization, as FRiP scores varied between samples and indicated different amounts of background.

As already in transcriptome and DNA methylation analysis, differential analysis revealed major changes on the level of chromatin accessibility. On a genome-wide level, chromatin accessibility increased strongly from iPSCs to DE, and then became again more closed in HLCs (Figure 49 A). However, PHH featured the lowest

amount of open regions, indicating a more controlled chromatin structure than in *in vitro* cultured cells (Figure 49 A). Most DARs detected in the pairwise comparisons in the dataset were attributed to the high amount of accessible regions in DE, but there were also pronounced differences between HLCs and PHH (Figure 49 B). DAR length was similar between comparisons, with exception of the comparison between PHH and CEL HLC, which frequently tended to be larger (Sup. Figure 15 D). Moreover, the genomic annotation of DARs was similar for all comparisons, with 40 % of DARs located in introns, 44 % in intergenic regions and only 8 % in promoters (Figure 49 C). To allow a functional interpretation of chromatin accessibility changes, DARs were annotated to the closest gene. Geneset enrichment analysis using the same parameters as on genes upregulated in HLCs compared to PHH (Figure 47 F) did not yield any significantly enriched GO terms, indicating random changes in the epigenome during HLC differentiation.

In addition, the diffusion map based pseudotemporal inference of ATAC-seq data comprehensibly modelled the HLC differentiation process, underlining the opening of chromatin from iPSCs to DE, followed by closing towards HLC, which however did not reach the even more closed chromatin state observed in PHH (Figure 49 D).
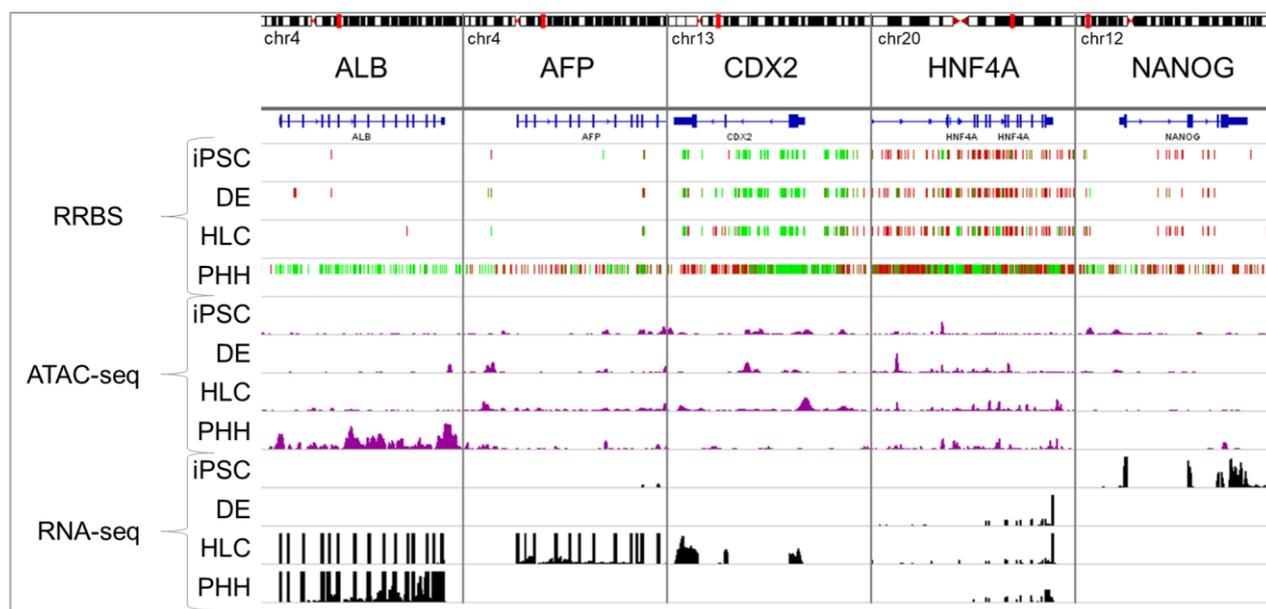
Figure 50: Exemplary genome browser view on the differentiation marker genes ALB, AFP, CDX2, HNF4A and NANOG. Displayed are gene expression (RNA-seq, black), chromatin accessibility (ATAC-seq, purple) and DNA methylation (RRBS) of a representative iPSC, DE, HLC, and PHH sample. Coverage tracks (ATAC-seq and RNA-seq) are normalized by sequencing depth. In RRBS tracks green corresponds to unmethylated CpGs, while red illustrates high methylation levels.

Finally, the integration of single cell data with bulk transcriptome and epigenome data (Figure 50) additionally aimed to gain a better understanding of genome-wide deregulations in *in vitro* cultured HLCs. To this end, gene expression log fold changes between iPSCs and PHH, which represent the desired effect of differentiation, were plotted against log fold changes of HLCs vs. PHH, which reflect the remaining differences between HLCs and the desired differentiation state (Figure 51). Accordingly, positive values correspond to upregulation in iPSCs (x axis) or HLCs (y axis), negative values to upregulated expression in PHH, which is also illustrated by average gene expression levels in HLCs (Figure 51 A) and PHH (Figure

51 B). For instance, genes upregulated in PHH, but not during HLC differentiation, are located in the lower left quadrant of the scatterplot. To obtain a real PHH-like phenotype, these genes would have to be upregulated in HLCs accordingly. Interestingly, genes upregulated in HLCs (upper left quadrant), but not in PHH, are mostly not expressed in PHH (Figure 51 B), underlining the higher number of expressed genes in HLCs (Figure 42 B) and suggesting an aberrant deregulation of gene expression in HLCs (Figure 51 A). This is also reflected on the level of single cell expression variability (Figure 51 C and D), which was highly correlated with average bulk gene expression levels (r_{HLC} = 0.85, r_{PHH} = 0.82).
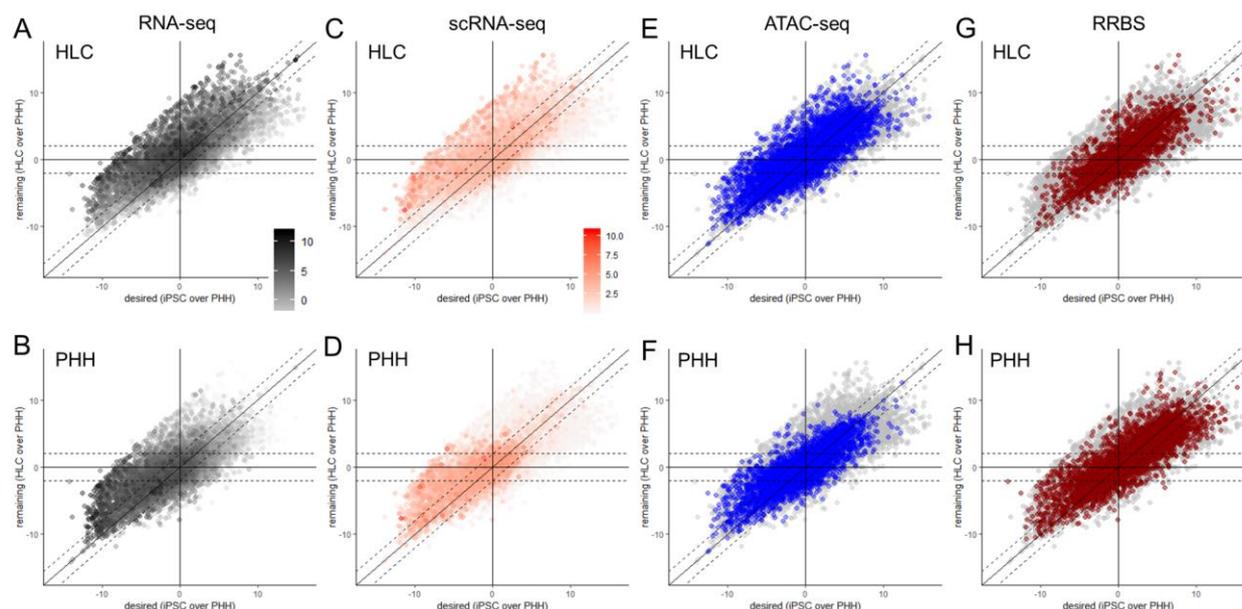
Figure 51: Integrative analysis of bulk RNA-seq data with scRNA-seq, ATAC-seq, and RRBS data. Scatterplots depict bulk RNA-seq log fold changes of iPSC over PHH (x axis: desired differentiation) and HLC over PHH (y axis: remaining differentiation). Accordingly, positive values correspond to upregulation in iPSCs (x) or HLCs (x), negative values to upregulated expression in PHH. Average expression levels, single cell variance and epigenetic data are integrated as color. A and B – Average log (CPM + 1) of bulk RNA-seq in HLCs (A) and PHH (B). Darker grey corresponds to high expression levels. C and D – Variance of single cell expression in HLCs (C) and PHH (D). Darker orange color corresponds to higher variance. To reduce batch effect induced sources of variation, only in-house data were used for calculation of PHH single cell variance. E and F – Genes associated with chromatin regions more accessible in HLC (E) respectively in PHH (F) are highlighted in blue. G and H –Genes associated with hypomethylated regions in HLC (G) and PHH (H) are colored in red. Genes not associated with epigenetic changes are displayed in grey. It should be noted that scatterplots are prone to overplotting. As genes associated with chromatin accessibility (E and F) or DNA methylation (G and H) changes are highlighted on top, genes not associated with epigenetic changes are frequently hidden.

Then, associated epigenetic changes on the level of chromatin accessibility and DNA methylation were integrated by color code, allowing the genome-wide assessment of complex epigenomic changes in relation PHH-like expression signatures and undesired expression patterns in HLCs (Figure 51 E – H). In this context it should be noted that frequently multiple regions with DNA methylation and or chromatin accessibility changes were observed per gene. As both hypomethylated and hypermethylated regions, respective more open and more closed chromatin, can be associated with the same gene, a meaningful statistical evaluation

of epigenetic effects on gene expression remains a major challenge. Mere averaging of observed epigenetic effects per gene would not represent the actual biological meaning, as the effect of epigenetic changes depends also on the genomic localization and changes in both directions associated to the same gene would not be represented properly. Moreover, only considering the region with the largest effect size would also entail a loss of relevant information, as for instance already relatively small changes in chromatin accessibility could impact transcription factor binding dynamics and thereby also gene expression (Sup. Figure 18). To circumvent this problem, genes associated with regions more accessible or hypomethylated in HLCs (Figure 51 E and G) and PHH (Figure 51 F and H) were plotted separately.

The resulting integrative visualization reveals several general features of epigenetic gene expression regulation during *in vitro* HLC differentiation. While increased chromatin accessibility in HLCs can be observed without obvious relation to gene expression deregulation (Figure 51 E), in PHH there is an enrichment of increased accessibility in genes with higher expression in PHH (Figure 51 F, lower left quadrant). Thus, more accessible chromatin is frequently indicative of an upregulated, favorable gene expression of PHH, but not in HLCs. In contrast, DNA methylation changes, which are predominantly characterized by a hypomethylation of PHH (Figure 48 A), do not reflect upregulated gene expression, neither in HLCs (Figure 51 G) nor in PHH (Figure 51 H).

Taken together, this global analysis suggests a well controlled epigenetic landscape in PHH that is involved in the regulation of gene expression (Figure 51 F). In contrast, epigenetic changes in HLCs appear to be more aberrant. On a genome-wide scale, the majority of epigenetic changes, including the increase of chromatin accessibility and in particular the predominant hypermethylation of HLCs, do not reflect gene expression signatures in HLCs.

In order to obtain a functional interpretation of the epigenetic changes that are associated with differentially expressed genes between HLCs and PHH, the overlap of DEGs and genes associated with DARs and or DMRs was calculated, without further regard to a direct overlap of DARs and DMRs. This identified 2,147 genes featuring differential expression associated both with differential DNA methylation and differential chromatin accessibility (Figure 52 A), which were termed epigenetically marked DEGs. Hierarchical clustering of the resulting integrative z-score heatmap (Figure 52 B) revealed 8 gene clusters, characterized by different combinations of transcriptional and epigenetic changes (Figure 52 C). Clusters 1 to 4, which are underrepresented in comparison to the remaining clusters, are characterized by a hypomethylation in HLCs. Thus, the integrative analysis is dominated by a predominant hypermethylation in HLCs as well (clusters 5 to 8, Figure 52 B). Cluster 8, which is characterized by upregulated gene expression associated with hypomethylation and higher chromatin accessibility in PHH, represents the favorable situation that should be achieved also in HLCs (Figure 52 C). Cluster 7 features also upregulated gene expression and DNA hypomethylation in PHH, but also reduced accessibility as compared to HLCs. In contrast, cluster 5 and 6 are upregulated in HLCs, thus representing undesired gene sets. While cluster 5 shows higher chromatin accessibility in HLCs, cluster 6 is associated with a more closed chromatin state (Figure 52 C).
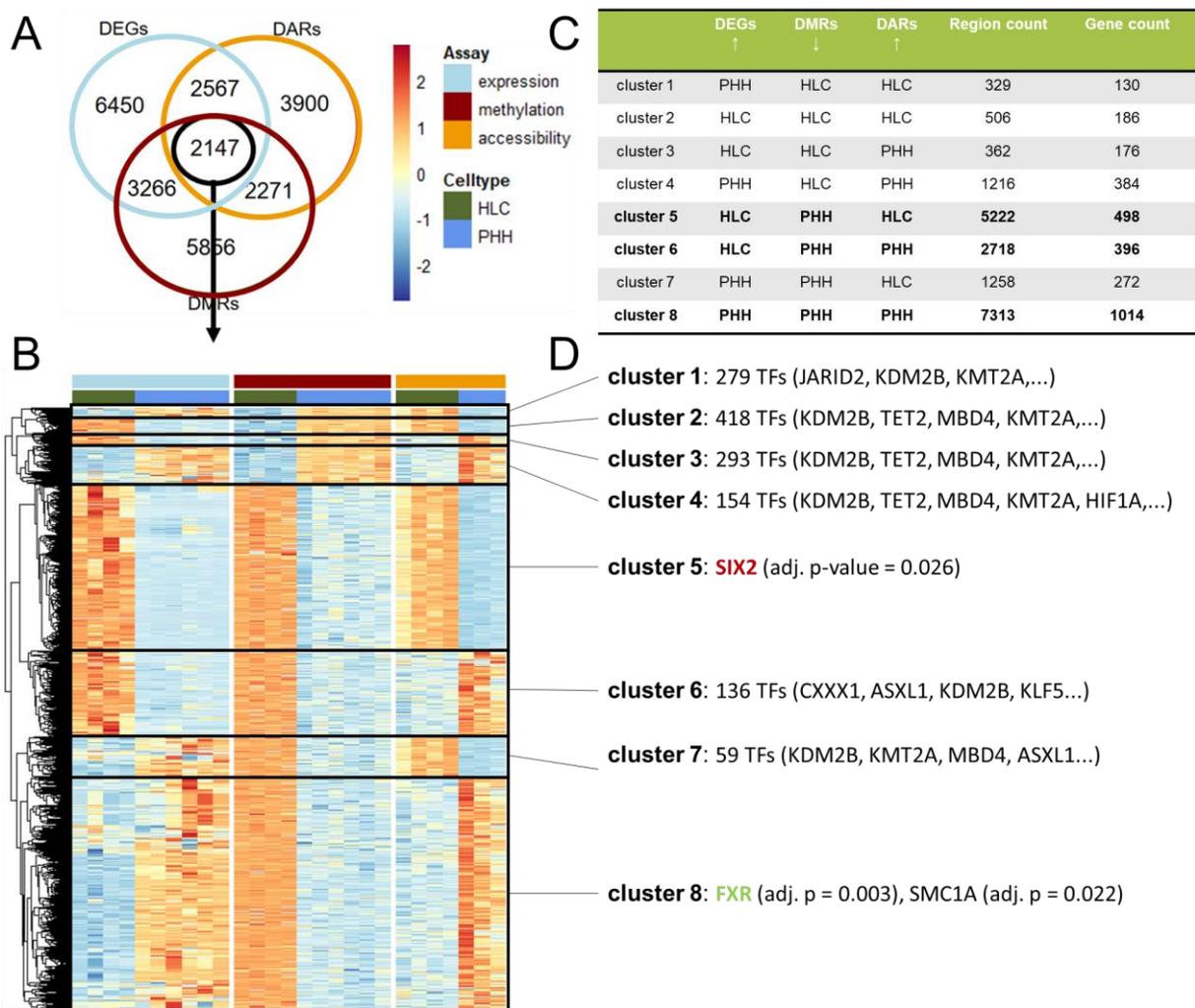
Figure 52: Integrative epigenomic analysis reveals potential key transcriptional regulators of HLC differentiation. A – Overlap of genes featuring differential expression (DEGs, blue), differential DNA methylation (DMRs, red) and differential chromatin accessibility (DARs, orange) between HLCs and PHH. B – Clustered integrated z-score heatmap of expression, methylation and accessibility of DEGs associated with DMRs and DARs. Z-scores were calculated separately for each assay. Red heatmap color represents upregulated expression, hypermethylation and higher accessibility, while blue corresponds to downregulated expression, hypomethylation and closed chromatin. HLC samples are annotated in green and PHH in blue. The heatmap is clustered by manhattan distance, thereby identifying 8 distinct clusters. C – Overview of DEG, DMR and DAR characteristics of the clusters detected in B. The number of differential regions and the count of associated genes are given as well. D – Key transcriptional regulators for each cluster as identified by REGGAE (Kehl et al., 2018) with adjusted p-value less than 0.05. The full list of TFs is provided in Sup. Table 14.

Potential key transcriptional regulators of individual gene clusters were detected by making use of the REGGAE algorithm (Kehl et al., 2018), which is based on a novel overrepresentation analysis of ranked gene sets. Intriguingly, while in most clusters a high number of different transcription factors (59 – 418 TFs) was detected, indicating a broad unspecific deregulation in HLCs, clusters 5 and 8 yielded specific putative key regulators (Figure 52 D). SIX2 (p = 0.026) was detected as the sole master regulator of cluster 5, representing genes which should be downregulated in an improved HLC system. This TF plays a role in the development of several organs including stomach and kidneys, but has not been associated with hepatocyte

differentiation or hepatogenesis so far (Ohto et al., 1998; Self et al., 2006, 2009). In contrast, key regulators of cluster 8, representing the genes that should be further upregulated in HLCs to obtain a PHH-like phenotype, are FXR (p = 0.003) and SMC1A (p = 0.022). While SMC1A was also found in other clusters, FXR is the only top regulator detected exclusively in a single cluster, underlining the specificity of this TF. Moreover, FXR was already identified by previous regulon analysis of single cell transcriptomes (Figure 46), which suggests FXR as a most promising candidate for cell culture interventions.

Interestingly, regulators of clusters with epigenetic patterns that do not directly reflect gene expression, for instance upregulated expression together with hypermethylation and more closed chromatin structure (cluster 1 and cluster 6), were highly enriched for chromatin remodeling (p = 1.3×10^{-9}), and in particular for positive regulation of histone H3-K4 methylation (p = 4.2×10^{-4}), histone H3 acetylation (p = 5.6×10^{-4}), negative regulation of histone H3-K9 methylation (p = 5.9×10^{-3}), and regulation of DNA methylation (p = 2.5×10^{-2}). Finally, it should be noted that CDX2, a marker of undesired colon-like expression patterns in HLC (Sup. Table 13), was identified in cluster 1 to 4 as well (cluster 1: p = 4×10^{-3}, cluster 2: p = 3.4×10^{-6}, cluster 3: p = 4.8×10^{-2}, cluster 4: p = 1.24×10^{-4}), although not ranking in the top 50 displayed in Sup. Table 14.

In summary, the epigenetic analysis revealed pronounced differences on the levels of chromatin accessibility and DNA methylation, both during *in vitro* differentiation of HLCs and in comparison with PHH. Intriguingly, an increasing hypermethylation from iPSCs to HLCs was observed, which is in stark contrast with the massive upregulation of gene expression observed in HLCs, while PHH showed considerably lower DNA methylation levels. Yet, this hypermethylation effect is in line with the transcriptional upregulation of DNMTs in cultured cells. In addition, on a genome-wide scale chromatin accessibility increased in DE samples. HLCs exhibited again a more closed chromatin structure, but not reaching the even more closed state of PHH. While the analysis of the single cell RNA-seq dataset already revealed the less mature state of HLCs in comparison to PHH and yielded a set of putative regulators of hepatocyte differentiation (chapter 3.3.1), the integrative analysis of transcriptomic and epigenomic data narrowed this set of TFs to the most probable key regulators of the desired PHH differentiation state. In this setting, FXR was detected as major putative regulator of desired expression and epigenetic signatures. Moreover, SIX2 was identified as an additional putative regulator of unfavorable gene expression in HLCs, and undesired epigenetic patterns were associated with many epigenetic regulators. The presented integrative analysis represents the basis for the development of reasonable cell culture interventions, including the overexpression and posttranscriptional activation of FXR during HLC differentiation.

4. Discussion

In the current era of single cell sequencing the study of cellular heterogeneity becomes more and more feasible and contributes to a better understanding of numerous intriguing questions of biology. However, there are still major challenges, including the spatial organization of single cells in tissues and integrating bulk epigenomic data into the analysis of single cell datasets. Based on the generation, analysis and integration of different types of genome-wide sequencing data, the present thesis aims to develop strategies to approach these challenges in order to achieve a better transcriptional and epigenomic understanding of hepatocyte heterogeneity in health and complex metabolic diseases as well as during *in vitro* differentiation. Regarding spatial heterogeneity of hepatocytes in the human liver, the first part presents a laser capture microdissection based approach to achieve the comprehensive transcriptional and epigenetic characterization of hepatic zonation in health (Brosch and Kattler et al., 2018) as well as in fatty liver diseases (chapter 4.1). In addition, the resulting transcriptional map of the human liver is used as reference for a novel pseudospace inference approach, which considerably improves spatial reconstruction of single cells into tissue context (chapter 4.2). Furthermore, the final part focuses on the assessment of hepatocyte heterogeneity during *in vitro* differentiation, and in particular demonstrates the integrative analysis of bulk epigenomic datasets with single cell transcriptome data.

4.1 Epigenomic analysis of human liver zonation

In recent years the field of epigenomics, which comprises the genome-wide mapping of different layers of epigenetic modifications by Next Generation Sequencing (NGS) based methods, fundamental achievements have been made. Especially large sequencing consortia like ENCODE and IHEC, including the German contribution DEEP, provide nowadays a plentitude of high quality epigenomic maps for a large number of tissues and cell-types, also related to different diseases for human and in part also for mouse (Stunnenberg et al., 2016). Although databases like the IHEC data portal (https://epigenomesportal.ca/ihec) are already highly valuable resources, it has become more and more apparent that for the proper understanding of tissue regulation it is of utmost importance to study isolated cell-types as well as different subpopulations of the same cell-type that could differ in their functions, which is reflected on the epigenetic level.

Thus, more recent consortia such as the Human Cell Atlas (HCA) or the upcoming LifeTime consortium focus on single cell sequencing technologies. Also, the focus of epigenetic research shifts towards smaller cell numbers – even down to single cell level. While single cell RNA sequencing (scRNA-seq) has become more and more feasible and can be readily applied to profile up to millions of single cells in high throughput assays, the epigenetic analysis of single cells is still a challenge. To date, several protocols for single cell bisulfite sequencing as well as single cell chromatin accessibility analysis are published (Buenrostro et al., 2015; Cusanovich et al., 2015; Guo et al., 2013; Smallwood et al., 2014). However, those methods only capture parts of the genome and require heavy bioinformatical preprocessing, including data aggregation and imputation to allow a meaningful analysis. Hence, the targeted analysis of well defined subpopulations of cells is, at least to date, the most comprehensive study design for many biological questions.

Within the frame of the BMBF funded EpiTriO project, the goal of a joined endeavor between the groups of Prof. Dr. Jochen Hampe (TU Dresden) and Prof. Dr. Jörn Walter was the zone-specific transcriptional and epigenomic characterization of hepatocytes in the human liver. In principle, a spatial analysis could be obtained by recent single cell sequencing technologies, however this would come with two major drawbacks. The preparation of single cell libraries requires tissue dissociation resulting in the loss of spatial information. Indeed, it is possible to bioinformatically infer spatial information from single cell data (chapter 4.2), but published approaches rely on known marker genes (Achim et al., 2015; Satija et al., 2015), for example determined by single molecule FISH (Halpern et al., 2017). In addition, especially methods for single cell DNA methylation analysis yield extremely sparse data (Farlik et al., 2015), which do not allow a comprehensive in-depth methylome analysis. Thus, we developed a complementary approach based on the unique experimental set-up of a pathologically well defined human cohort in combination with sophisticated laser capture microdissection (LCM) coupled to next generation sequencing (NGS), which preserves spatial orientation and still provides deep genome-wide data of approximately 100 to 200 isolated cells. By using adjacent liver cryosections, it was even possible to obtain matching high-quality RNA-seq and RRBS data, thereby allowing the first integrative spatial analysis of transcriptomes and methylomes across three micro-dissected zones along the porto-central axis of human liver (Brosch and Kattler et al., 2018).

4.1.1 Hepatic zonation in healthy liver and early NAFLD

The first integrated spatial epigenomic map of the human liver (chapter 3.1.1) comprises healthy human liver as well as the early spectrum of NAFLD, and can be queried regarding multiple fundamental questions of human liver biology such as the regulation of metabolic zonation and hepatic regeneration (Brosch and Kattler et al., 2018).

Depending on their spatial location within the hepatic lobule, hepatocytes show remarkable metabolic heterogeneity, which is termed metabolic zonation (Jungermann and Sasse, 1978). Previous studies in rodents revealed that zonation is established and maintained by interacting morphogen gradients along the porto-central hepatic axis (Gebhardt, 2014). In particular, a pericentral Wnt-/β-catenin signaling is described as a master regulator of metabolic zonation (Preziosi et al., 2018). Interestingly, the same morphogens were also implied to have major impact on hepatic regeneration (Font-Burgada et al., 2015; Wang et al., 2015). In contrast, hepatic zonation and underlying regulatory mechanisms in human liver are not well understood. Most previous studies relied either on measurement of zonal enzyme activity (Sokal et al., 1989, 1991, 1993) or on immunohistochemical assays (Palmer et al., 1992; Ratanasavanh et al., 1991; Suzuki and Ono, 1987), which only provide information about zonation patterns of few selected metabolic enzymes. A recent study using a similar laser capture microdissection RNA sequencing approach on normal human liver reported 139 transcripts with differential expression across the hepatic lobule (McEnerney et al., 2017). However, the analysis was limited to normal liver tissue from 3 human donors (2 females, 1 male) undergoing curative hepatectomy of liver metastasis. Albeit this study gives a first impression about zonal signatures and indicates an upregulation of Wnt signaling in the pericentral zone as shown in mouse, it lacks the statistical and methodological power to gain deeper insights into the principles of zonated morphogenic and metabolic control in human liver. Thus, while confirming previously detected zonated gene expression (Figure 16 D),

our approach identified 667 additional genes with zonated expression profiles (Figure 16 B). Intriguingly, this pronounced transcriptional gradient was also reflected on functional and epigenetic levels.

The detected mitochondrial gradient (Figure 21 B and C) reflects the known oxygen gradient as well as the resulting oxidative capacity and metabolic functions of each zone (Kietzmann, 2017), and is in line with microscopic findings in rat liver, reporting less mitochondria in pericentral hepatocytes (Loud, 1968).

On the epigenetic level a methylation gradient with a predominant pericentral hypomethylation was observed (Figure 18 B). Differential DNA methylation and expression of associated zonated genes was largely anti-correlated (Figure 19 A), although DMRs correlating with transcription were detected as well. In this context it should be noted that the regulatory effect of DNA methylation depends on its genomic localization. For a long time DNA methylation and gene expression were considered as purely anti-correlated (Razin and Cedar, 1991), with hypomethylation promoting active transcription. Regarding promoter regions and in particular CpG islands, this dogma certainly holds true, but in other regulatory elements such as enhancers and insulators or along the gene body, the situation appears to be much more complex. For example, DNA methylation is often enriched in exons of transcribed genes and apparently has an effect on alternative splicing (Shayevitch et al., 2018) or inhibits unwanted spurious initiation of transcription (Neri et al., 2017). Taken together, also an increase of DNA methylation can exert a regulatory effect on the upregulation of transcription, and hypomethylation can still result in downregulated or deregulated gene expression. Thus, the analysis of genome-wide epigenomic data should not be solely restricted to regions with anti-correlated DNA methylation and gene expression changes, as also anti-dogmatic regions may still hold essential regulatory potential, even though not all mechanistic links are already known.

Furthermore, DNA methylation can influence transcription factor binding dynamics. While most TFs bind to unmethylated DNA sequences (Grossman et al., 2018), there are also TFs that selectively bind methylated CpGs (Yin et al., 2017). Thus, the observed differential methylation of TF binding sites (Figure 20 A) may represent an additional layer of epigenetic zonation control in the human liver. This epigenetic zonation indicates a spatial activity of otherwise uniformly expressed TFs, which constitutes an interesting concept that could also play a role in other tissues.

Both the transcriptomic and epigenomic zonation signatures are preserved in early non-alcoholic fatty liver disease (Figure 17). Donors with bland steatosis and early NASH showed consistent gene expression differences to normal and healthy obese controls corresponding to the severity of the disease across all zones (Sup. Figure 3). Yet, the relative zonated gene expression and DNA methylation patterns remain largely unchanged (Figure 16 B and Figure 18 B), which is compatible with the complete regeneration capacity of human liver in early NAFLD (Cohen et al., 2011).

Our integrated epigenomic map of the human liver also allows to assess existing concepts on liver regeneration, which were largely developed by lineage-tracing experiments in mouse (Font-Burgada et al., 2015; Pu et al., 2016; Wang et al., 2015). The strong transcriptional and epigenetic gradient of the Wnt signaling pathway and modulating factors (Figure 23), together with the pericentral expression of LGR5 and AXIN2 (Figure 24 A) and the corresponding gradient of the liver progenitor marker TBX3 (Lüdtke et al., 2009; Suzuki et al., 2008), indicate a pericentral source of hepatocyte regeneration in homeostatic liver tissue. Conversely, in periportal hepatocytes a JAG1, NOTCH and EPCAM signature (Figure 24) was observed. This might correspond to hepatic progenitor cells bordering the biliary tree, which could represent a relevant

source of regeneration under injury conditions (Cardinale et al., 2011; Font-Burgada et al., 2015). Intriguingly, the interplay of Notch and Wnt signaling was previously suggested to regulate regeneration of hepatocytes in chronic liver disease (Boulter et al., 2012). This is further discussed in chapter 4.1.2, which is based on the zonal analysis of fibrotic and cirrhotic liver tissue, representing the chronic state of fatty liver disease.

Taken together, the integrated spatial map of human liver provides a wealth of new positional insights into zonal networks controlled by epigenetic and transcriptional gradients. In addition, this dataset will serve as a valuable resource for further studies, such as the spatial integration of single cell data as discussed in chapter 4.2.

4.1.2 Hepatic zonation in progressed fatty liver disease

In the human, the onset of non-alcoholic fatty liver disease (NAFLD) as well as alcoholic liver disease (ALD) is characterized by the accumulation of lipid droplets in pericentral hepatocytes (Cohen et al., 2011), which strongly suggests the zone-specific analysis of fatty liver disease progression. Thus, the spatial dataset was extended to more progressed stages of fatty liver disease. The human cohort comprised different stages of NAFLD up to fibrosis state 3 (FIB3). Within the time frame of this thesis it was not possible to obtain fibrosis state 4 samples from donors with a non-alcoholic background. In order to still allow the analysis of the final irreversible stage of fatty liver disease, a set of cirrhosis samples from donors with an alcoholic background was included. In this context, it should be emphasized that even though many features are shared, NAFLD and alcoholic fatty liver disease (ALD) are distinct pathologies (Rowell and Anstee, 2015). Thus, findings obtained from ALD should not be transferred to NAFLD without careful consideration, and need to be confirmed in liver tissue derived from non-alcohol cirrhosis.

It should be noted that the proportions of NPCs in liver tissue can differ between disease stages, as in particular fibrosis and cirrhosis samples are characterized by a pronounced infiltration of immune cells. These shifts in cell-type proportions are major confounders of transcriptional and especially epigenetic changes in whole tissue analysis (Lam et al., 2016). Thus, the LCM based approach, which allows the predominant analysis of hepatocytes, represents a major advance over whole tissue analysis and allows to decipher hepatocyte-specific deregulations in fatty liver disease. On the other hand, different NPC types, as well as their composition, play a major role in the onset and progression of fatty liver disease (Choi et al., 2019; Sato et al., 2019). To understand a tissue during disease development as a whole, a next step would be to integrate information of all involved cell-types. Compared to hepatocytes, NPC types are underrepresented in the liver parenchyma. Thus, it is to date not feasible to analyze them by our LCM based sequencing approach. A possible future alternative might be the integrative analysis of single cell sequencing data coupled to sophisticated imaging technology to gain a better understanding of the interplay of diverse cell-types creating a complex spatial microenvironment.

The present analysis aims to unravel transcriptional and epigenetic signatures of hepatocytes. Low contamination rates of NPCs were verified by the assessment of marker gene expression. Although progressed fatty liver disease is characterized by an increasing inflammatory infiltration (Narayanan et al.,

2016), no major differences between controls, early and progressed disease stages were observed (Sup. Figure 1 A), confirming the applicability of our microdissection approach. Interestingly, cirrhotic samples showed slightly elevated levels of cholangiocyte markers. Indeed, this could be attributed to a higher amount of contamination. However, other NPC markers did not show increased expression levels in cirrhosis, and the amount of cholangiocytes should not differ drastically between phenotypes. Thus, it is rather unlikely to have a high cholangiocyte contamination in ACI samples. The upregulation of expression levels of cholangiocyte-specific genes might be explained by transdifferentiation of hepatocytes towards a cholangiocyte-like phenotype, which was reported to occur during sever liver injury (Michalopoulos et al., 2005).

In progressed disease stages, pronounced deregulations were observed both on the transcriptional and epigenetic level. The highest count of significant DEGs was detected in the pericentral zone. Yet, the number of detected DEGS may not be conclusive, as the number of replicates for phenotype-specific differential analysis (Figure 30) was rather low, limiting statistical power. Moreover, genes just above FDR cut-off (> 0.01), which might still be biologically relevant, would not be included. To prevent a biased data interpretation based on DEG number, zone-specific DEGs were rather compared by correlation analysis of log fold changes, which does not rely on thresholds and gives a better overview on the whole picture. This analysis revealed that the majority of differential gene expression was detected uniformly across all hepatic zones (Figure 33 C). However, distinct zone-specific differences, which intriguingly comprised among others a set of morphogens involved in the regulation of hepatic zonation (Figure 34 A), might have decesive regulatory impact. Thus, within the scope of this thesis, data analysis focused on hepatic zonation signatures, which represent a unique feature of this dataset that is not feasible in published NGS based studies of fatty liver disease (Gerhard et al., 2018; Ryaboshapkina and Hammar, 2017).

A major finding of the zonal analysis of steatohepatitis, fibrosis and cirrhosis is the progressive loss of transcriptional and epigenetic zonation in late stages of fatty liver disease (Figure 29 A and B). Of particular interest is the observed portalization of pericentral hepatocytes, as it might represent a key step of disease progression. This effect was most obvious on the level of gene expression, with expression profiles of zonated genes resembling those in periportal hepatocytes (Figure 28, Figure 32 A and B). Moreover, the zonated gradient of mitochondrial abundance, which is characterized by an increasing amount of mitochondria from pericentral to periportal hepatocytes in normal tissue (Figure 21 A), was disrupted in cirrhotic tissue, as in the pericentral zone the same ratio of mitochondrial reads was detected as in the intermediate and periportal zone (Figure 33 C). In normal liver tissue, the mitochondrial abundance reflects the known oxygen gradient along the porto-central axis and the resulting oxidative capacity (Kietzmann, 2017), which again corresponds to the specialized metabolic functions of each hepatic zone (Gebhardt, 2014). Thus, the disruption of the mitochondrial gradient in cirrhosis might contribute to the portalization of pericentral hepatocytes, which ultimately results in the loss of pericentral liver functions such as xenobiotic metabolism (Talal et al., 2017).

In liver with normal histology, Wnt signaling, which is probably initiated by the endothelial cells of the central vein (Preziosi et al., 2018; Wang et al., 2015) forms a pronounced pericentral gradient (Figure 23), representing a major regulator of hepatic zonation (Figure 53, Birchmeier, 2016). Thus, Wnt signalling is essential for the typical expression profiles of pericentral hepatocytes (McEnerney et al., 2017; Preziosi et

al., 2018). Intriguingly, in progressed stages of fatty liver disease, a considerable downregulation of Wnt signaling was observed in pericentral hepatocytes (Figure 33 A and C, Sup. Figure 8).

Extracellular matrix organization plays a major role in fibrotic and cirrhotic liver tissue (Arriazu et al., 2014), as excessive amounts of type I collagen are deposited around the central vein, thereby forming the so called fibrous septa (Figure 26), which build a kind of physical barrier between the endothelial cells of the central vein and pericentral hepatocytes. Thus, the proper formation of the endothel-initiated Wnt signalling gradient in the pericentral zone might be prevented (Figure 53), which in turn might lead to the observed downregulation of genes that feature pericentral expression profiles in normal liver tissue (Figure 28 A, Figure 32 A).
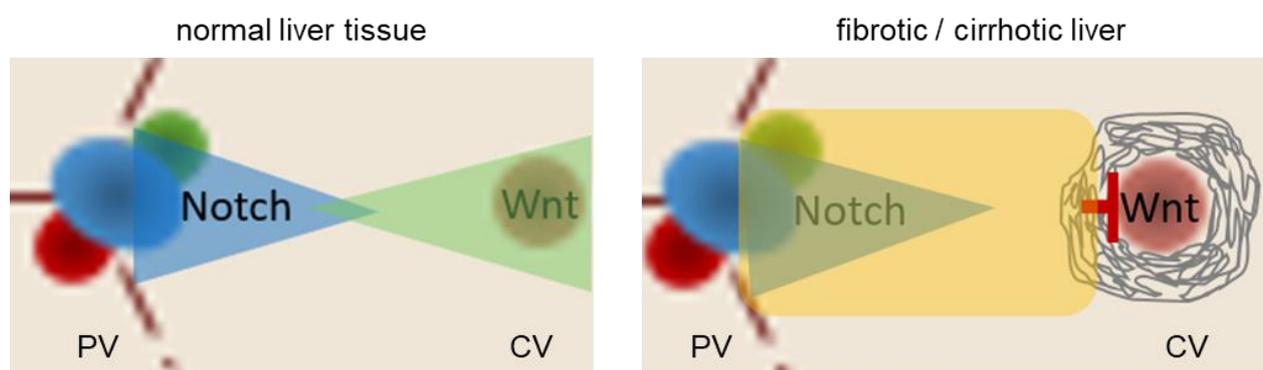
Figure 53: Model of morphogen gradients in normal liver tissue and fibrotic or cirrhotic liver. Fibrotic tissue (grey) forms a barrier that prevents formation of the pericentral Wnt signaling (green), thereby allowing the extension of the periportal Notch (blue) signaling gradients with significant upregulation of SOX9 and EPCAM (orange) towards the pericentral zone.

Moreover, Wnt signaling might be essential to control periportal Notch signaling. There are multiple reports describing a repressive effect of Wnt signaling on the Notch signaling pathway (Boulter et al., 2012; Huang et al., 2014; VanHook, 2012). The underlying molecular mechanisms remain poorly understood, although studies in zebrafish and human cell lines report several proteins such as NUMB or Disheveled to play a role in this regulatory crosstalk. However, neither of them was differentially expressed in our human liver dataset. While Notch signaling itself shows only a trend towards pericentral upregulation in cirrhosis, the periportal morphogen SOX9 is considerably upregulated in pericentral hepatocytes (Sup. Figure 9). Interestingly, SOX9 corresponds to the Notch signaling pathway in a positive feedback loop (Leung et al., 2016; Yin, 2017). However, SOX9 also interacts with the Wnt signaling pathway. While it can apparently directly activate gene expression of several pathway members (Huang et al., 2017; Leung et al., 2016), SOX9 was also reported to inhibit Wnt signaling by promoting β-catenin phosphorylation in the nucleus (Topol et al., 2009). The zone-specific analysis progressed fatty liver disease indicates that morphogens usually expressed in the periportal zone might be upregulated in pericentral hepatocytes (Figure 34 A, Sup. Figure 9) as a consequence of missing Wnt signaling (Figure 53). However, the morphogens deregulated in liver fibrosis and cirrhosis form a complex interaction network (Figure 34 B), whereof underlying molecular mechanisms have to be resolved in future studies.

Taken together, the portalization of pericentral hepatocytes suggests the periportal hepatocyte profile as a kind of ground state, which is only overcome by properly sustained Wnt signaling to achieve a functional

pericentral hepatocyte population. This hypothesis is in line with the presumed regenerative capacity of each zone. While pericentral hepatocytes might play a role in hepatic regeneration during normal liver homeostasis (Wang et al., 2015), periportal hepatocytes apparently represent the major source of hepatic regeneration in chronic liver disease (Font-Burgada et al., 2015), which is characterized by a massive proliferation capacity (Corlu and Loyer, 2012). Thus, in addition to the loss of endothelial Wnt signalling, the periportally driven regeneration of the liver parenchyma could further contribute to the observed portalization of hepatocytes in severe chronic fatty liver disease. As the pericentral zone is crucial for proper metabolic liver functions, in particular regarding xenobiotic metabolism, triglyceride synthesis and glycolysis (Gebhardt, 2014), the portalization of pericentral hepatocytes might be a key factor of liver failure in cirrhosis. If this hypothesis holds true, it may give rise to new research avenues for the development of curative drugs targeting the signaling pathways deregulated in chronic fatty liver disease.

In the human, liver metabolism exhibits a high inter-individual variability, as it depends heavily on individual lifestyle factors such as diet, alcoholic consumption, smoking or drug intake. In addition, the mammalian liver exhibits pronounced sex differences (Saito et al., 2013) and a circadian rhythm (Vollmers et al., 2009) that impact gene expression profiles. To avoid circadian bias in our analysis, all samples were obtained at the same time of day. Moreover, patients were fasted, which should decrease direct dietary effects on gene expression. Still, especially on the epigenetic level a considerable inter-individual heterogeneity was observed, which intriguingly increased in donors with liver fibrosis and cirrhosis (Sup. Figure 7 A). This observation might suggest an undirected deregulation of DNA methylation patterns in late fatty liver disease. Moreover, on the level of DNA methylation, phenotype-specific zonation patterns were detected (Figure 29 B), which were however not reflected on the transcriptional level (Figure 29 A). As replicate numbers per phenotype were rather low (Figure 31), it should be considered that these phenotype-specific zonal DMRs might be of low confidence, detected merely due to statistical artefacts. However, their standard deviations were in the same range as of those DMRs which were shared between phenotypes, suggesting the validity of the phenotype-specific DMRs. It should be noted that with increasing CpG coverage more DMRs per comparisons might be detected, which might also increase the overlap of zonal DMRs between phenotypes. In addition, most DMRs detected in the zone-wise comparison between HO controls and the other phenotypes (Figure 31) were not associated with differential gene expression as well. Thus, only a minor fraction of DNA methylation changes was associated to differential gene expression (Sup. Figure 7 F). In this context it should be noted that the association of DMRs to genes was done in a very simplistic way by annotating the closest gene. Of course, DNA methylation differences in regulatory regions that effect genes over long distance interactions would not be appropriately reflected by this approach. While chromatin conformation assays such as HiC are very useful to unravel long distance interactions between regulatory elements, due to low resolution and extremely high sequencing costs, their applicability in a large set of primary human samples is still not feasible (Díaz et al., 2018). Thus, to date, there is no reasonably improved way to comprehensively connect DMRs with gene expression in a complex human dataset. Yet, this will probably only affect a small fraction of DMRs, as the majority of detected DMRs was localized directly within gene context (Sup. Figure 7 A).

Thus, the vast majority of DMRs is apparently not directly associated with the regulation of gene expression, which could imply several things. These DMRs could be produced as a kind of unspecific bystander effect

by the profound metabolic changes within the cell due to chronic inflammation and the constant production of ECM (Arriazu et al., 2014; Dietrich, 2016). For instance, sustained CYP2E1 mediated ethanol metabolism results in oxidative stress (Bardag-Gorce et al., 2006), and resulting reactive oxygen species might cause aberrant DNA methylation alterations (Weitzman et al., 1994).

Moreover, both NAFLD and ALD patients with liver cirrhosis have a significantly higher risk of developing hepatocellular carcinoma (HCC, Kanwal et al., 2018; Seitz et al., 2018). In line with the higher variability of DNA methylation differences observed in fibrotic and cirrhotic samples (Sup. Figure 7 A), DNA methylation signatures might also be indicative of future disease progression. In cirrhosis, which is described as a gateway for the development of HCC (Ramakrishna et al., 2013), multiple oncogenes such as MYC, CCND1, PAX8, and RAF1 were associated with differential DNA methylation. Thus, it could be possible that a tumorigenic potential is already beginning to manifest on an epigenetic level, while not yet detectable on transcriptional or even morphological levels. To substantiate this hypothesis, fibrotic and cirrhotic DNA methylation profiles would ideally have to be compared with compatible HCC data. However, the increasing cell-type heterogeneity in tumor tissue (Kurebayashi et al., 2018) might interfere with the microdissection of hepatocytes. Thus, additional LCM independent experiments such as DNA methylation profiling of FACS sorted nuclei or even of single cells would be of avail. If the hypothesis holds true, it might open new research directions for DNA methylation signatures as predictive markers for HCC development, which might be more reliable than current serum markers (Zhao et al., 2013). Moreover, DNA methylation based markers that could already be observed in early stages of liver disease would lead to earlier detection and allow the preventive treatment of patients with a high risk for HCC development, which could significantly improve prognosis.

In summary, these results demonstrate the relevance of spatial tissue orientation and the added value gained by integrating positional information into the analysis of tissues in health and disease. Beyond the intriguing observations regarding the regulation of human liver zonation in health and complex metabolic disease, the spatial map of human liver tissue can serve as a valuable reference for data integration, which will be discussed in the following chapter.

4.2 Spatial reconstruction of single cell data

The intriguing observations obtained by integrative analysis of liver zone-specific transcriptome and DNA methylome profiles of human hepatocytes demonstrate the relevance of spatial tissue orientation and the added value gained by integrating positional information into the analysis of tissues in health and disease. Multiple observations indicate a gradient-like regulation of hepatic zonation along the porto-central axis. In line with previous findings in mouse liver (Halpern et al., 2017) and a recent single cell study in human liver (Aizarani et al., 2019), differential gene expression analysis across hepatic zones suggests a monotonic expression gradient (Figure 16 B), instead of distinct hepatic zones with well-defined expression profiles. However, the inherent low resolution of our laser capture microdissection based approach limits the more detailed analysis of expression gradients.

To this end, RNA sequencing of single hepatocytes dissociated from whole liver represents a complementary approach to resolve hepatic zonation at single cell level, even though it entails several issues that have to be taken into account. In particular high-throughput 3'-tag scRNA-seq methods, which can cover more than 10,000 single cells per experiment, result in sparse data, which are prone to increased technical noise, and yield high drop-out rates as well as low numbers of reads (< 50,000) and covered genes per cell (< 2,000, Kolodziejczyk et al., 2015). Moreover, due to single cell resolution, there is also a considerable amount of biological noise, including heterogeneity arising from inter-individual variability, transcription kinetics, biological processes such as cell cycle and apoptosis, but also from cell-types and subpopulations. Therefore, it is of utmost importance to distinguish between technical and biological variability, and to separate biological heterogeneity of interest from biological noise arising from unwanted factors such as cell cycle phase or donor variability. Thus, stringent quality control and proper data normalization are required prior to downstream data analysis and interpretation.

Even though scRNA-seq is a relatively new method, extremely rapid progress is made regarding data analysis methods. Already well established workflows comprise for example marker gene and clustering based characterization of cell-types and subpopulations, the identification of genes, which drive biological processes, as well as differential expression analysis between single cell samples or clustered populations. There are also attempts to perform differential splicing between subpopulations and even to analyze allelic expression patterns (Kolodziejczyk et al., 2015). Moreover, pseudotemporal lineage tracing (Trapnell, 2017), which can be used to track cellular processes over time, and transcription factor regulon activity inference (Aibar et al., 2017) to identify distinct regulatory networks in subpopulations can be applied to answer numerous intriguing questions of single cell biology.

In contrast, spatial analysis of single cell transcriptomes remains, to date, a major challenge and is not yet feasible for most experimental settings. A major drawback of scRNA-seq approaches is that they require tissue dissociation and isolation of single cells, thereby loosing information regarding their spatial origin. There are few published approaches using special experimental settings, which rely on a set of reference genes with known local expression patterns, to achieve spatial reconstruction of single cell transcriptomes into tissue orientation. A study in zebrafish proposed a spatial reconstruction strategy that relies on mapping of imputed single cell expression data to a binary 3 dimensional landmark gene matrix obtained from RNA fluorescence *in situ* hybridization (FISH) of 47 known marker genes (Satija et al., 2015). Similar, spatial reconstruction of single cells in the marine model organism *Platynereis dumerilii* is based on specificity-

weighted mRNA profile mapping to a reference FISH atlas (Achim et al., 2015). Both approaches are based on binarized reference data, which entails the binarization of scRNA-seq data prior to spatial reconstruction, resulting in a loss of quantitative expression information. Thus, they yield convincing results in tissues with clear-cut regional expression profiles of distinct marker genes. In cases of monotonic expression gradients across a tissue, these approaches are however not ideal. For our single hepatocyte transcriptome dataset, the approach by Achim et al. correctly identified hepatocytes close to the central and portal veins, but was not able to assign hepatocytes from the intermediate zone, as these cells do not express zone-specific marker genes (data not shown). More recently, quantitative single-molecule FISH (smFISH), which allows detailed transcript quantification at single cell resolution, albeit only for a small number of selected genes, was used for the generation of a continuous reference map of the hepatic lobule in mouse (Halpern et al., 2017). Single hepatocytes were then assigned into 9 layers along the porto-central axis by probabilistic spatial inference based on 6 reference genes. Halpern et al. correctly identified monotonic expression gradients, as suggested by smFISH data, but also non-monotonic expression profiles with highest expression in intermediate layers. Although this approach evidently represents a valid strategy for spatial tissue reconstruction, it relies only on a small subset of known landmark genes as reference. This can be a major drawback, for instance for spatial reconstruction in liver disease such as cirrhotic tissue with deregulated zonation profiles (Figure 27 A), but also in tissues with no known landmark genes to easily build a smFISH reference. Additionally, single cell data, especially those generated by high-throughput methods such as droplet-based 10X, are inherently noisy and have high drop-out rates. Hence, relying on few marker genes for spatial reconstruction could lead to biased spatial embeddings.

In this context, laser capture microdissection coupled to RNA-seq (LCM RNA-seq) represents a reasonable alternative for the generation of spatial reference maps without prior knowledge of landmark genes (Moor et al., 2018). As we already generated high quality zonal expression reference maps for human liver by LCM, yielding genome-wide, continuous expression values, it was reasoned that these data might serve as a valuable resource to anchor spatial mapping of single cell data. Compared to smFISH, LCM RNA-seq derived data have the advantage of providing spatial information for all expressed genes in the analyzed cell-type. However, a limitation is that our LCM approach averages about 100 cells per zone and does not reach single cell resolution. Thus, probabilistic inference as performed by Halpern et al. would not result in a detailed expression gradient, but rather mirror the resolution of the three microdissected zones. Although this would already add a valuable covariate for single cell data interpretation, a more fine-grained modelling of single cells would be preferable. To overcome this limitation, a complementary approach, which allows the spatial reconstruction at single cell resolution, but is still anchored to low-resolution LCM based reference data, was developed.

The hepatic porto-central axis effectively represents a 1-dimensional structure with radial symmetry that exhibits a functional spatial zonation (Gebhardt, 2014). As in healthy liver this hepatic zonation is a major driver of variance in gene expression data (Sup. Figure 2 A), it was reasoned that it should be possible to capture spatial profiles in multi-dimensional single cell transcriptome data using dimensionality reduction approaches. To this end, several algorithms widely used for scRNA-seq data analysis, including PCA, tSNE, UMAP and diffusion map, were assessed (Figure 35). Based on all highly variable genes (HVGs) in the single cell dataset, the spectral non-linear diffusion map algorithm, which relies on k nearest neighbor

approximation (Haghverdi et al., 2015), showed the most promising performance (Figure 35 D). Diffusion distances, which are the probabilities of a cell to transition into other cells, are calculated for the k nearest neighbors based on a Gaussian kernel with width sigma. In context of spatial reconstruction, a small diffusion distance between two cells can be considered as an indication for neighboring localization in the tissue of origin. After density normalization to incorporate only transition probabilities between different cells, an eigendecomposition of the transition probabilities is performed, which yields diffusion components with decreasing eigenvalues (Haghverdi et al., 2015). Thus, diffusion component 1 (DC1) usually represents the major driver of variance. The reference-free diffusion map based on all HVGs showed separation by pericentral and periportal zones along DC1 (Figure 35 D). However, the separation remained rather noisy, and particularly the intermediate zones was not sufficiently resolved, which can probably be explained by the high amount of variability inherent to single cell data (Kolodziejczyk et al., 2015).

Thus, reducing the gene set to HVGs that show a zonal expression profile in healthy liver should increase the ratio of variability actually caused by spatial origin of single cells. Indeed, anchoring the diffusion map on zonated HVGs, as defined by our LCM based reference dataset, resulted in a strikingly clear separation of hepatic zonation along DC1 (Figure 35 H), which neatly resembled the presumed orientation of hepatocytes along the porto-central axis. Thus, the diffusion map on zonated HVGs was selected as the most promising of the tested approaches for dimensionality reduction based reconstruction of hepatic pseudospace.

In principle, DC1 coordinates could be used to assign each cell a discrete localization in hepatic pseudospace. However, to achieve a more robust modelling, DC1 was rather considered as the probability to belong to a certain hepatic zone (Figure 37 A). To ensure a valid zonal assignment, DC1 thresholds for zone definition were determined by expression patterns of all zonated HVGs (n = 107). For visualization of low abundance genes imputed expression values were used to allow a meaningful assessment of expression gradients. Previous approaches rely also on imputed expression data for the spatial reconstruction itself (Satija et al., 2015), although data imputation might introduce additional variation into the dataset (Andrews and Hemberg, 2019). As the diffusion map was mainly driven by highly expressed genes with pronounced zonation profiles (Figure 36 C – D), which yielded consistent and reasonable spatial embeddings (Figure 36 C – D, Figure 37 C), there was no need for data imputation for diffusion map calculation, thereby avoiding the potential introduction of additional bias.

Although the diffusion map based spatial reconstruction approach was mainly developed for sparse high-throughput scRNA-seq data, it is also feasible in small datasets with higher coverage, such as our Smartseq2 PHH single cell expression dataset (Sup. Figure 11). Interestingly, this dataset comprised 3 individual human donors. Using a deconvolution based normalization method (Lun et al., 2016), it was possible to reduce inter-individual variability in order to allow proper diffusion map based pseudospace inference even across different donors. This further highlights the robustness and applicability of the diffusion map based spatial reconstruction.

Tissue disintegration prior to scRNA-seq library preparation usually results in a loss of cells. Thus, the final dataset does not represent all cells of the tissue sample, and some cellular subpopulations might be preferentially lost during tissue homogenization, which might result in biased cell number proportions between subpopulations. Here, the numbers of assigned cells per hepatic zone reflects mostly the intrinsic

tissue symmetry of the hepatic lobule with few cells neighboring the central vein and more cells localized around the portal field (Figure 37 B), which is in physical space much larger and can thus harbor more cells (Gebhardt, 2014). In addition, spatial reconstruction of single cells yielded subdivision of the three defined hepatic zones, which show meaningful expression gradients and might therefore represent actual hepatic layers (Figure 37 B and C). Interestingly, further analysis of these layers indicated non-monotonic gene expression for a subset of zonated genes (Figure 37 C). This is in line with previous findings in mouse (Halpern et al., 2017). In our LCM RNA-seq dataset there was no indication for non-monotonic genes expression, which could however be attributed to the inherent low cellular resolution of this approach. On the other hand, the expression levels of these genes were rather low, which resulted in an overestimation of expression differences in the scaled heatmap illustration (Figure 37 C). As these expression differences might merely arise from noise in the single cell data, these results should be considered with care. Thus, independent validation of expression profiles in the human hepatic lobule, for example by smFISH, would be necessary to validate non-monotonic gene expression.

A recent pre-print available on bioarxiv suggests to handle spatial reconstruction as a generalized optimal transport problem (Nitzan et al., 2018). This intriguing approach is based on the mathematical transport theory, which was originally proposed in the 18th century by Gaspard Monge and mathematically established in the 1920s to determine optimal routes for cargo transport (Grattan-Guinness, 1994). In a nutshell, a transport problem aims to find minimal physical distances, thus it can be readily applied as a spatial assignment problem of single cells. Nitzan et al. apply this approach amongst others to the previously described mouse hepatocyte dataset (Halpern et al., 2017), robustly identifying hepatic expression gradients. Although conceptually designed as a reference-free pseudospace inference, they also find that anchoring to a reference dataset outperforms the unsupervised reconstruction. Moreover, accuracy apparently increases with the number of reference genes (Nitzan et al., 2018), which again advocates the use of LCM for reference dataset generation.

The transport problem related approach and the diffusion map based spatial reconstruction both effectively model hepatocyte localization along the porto-central axis. In this context, it should also be considered that the hepatic lobule is a 3-dimensional tissue, which also comprises non-parenchymal cell-types (NPCs). The spatial mapping of NPCs, such as endothelial cells of central respective portal veins, is especially challenging, as these cells have a rather low mRNA content and do not feature distinct expression of landmark genes. Recently, this challenge was met by an intriguing approach using paired-cell sequencing based on inefficient digestion of mouse liver, which resulted in a high fraction of duplets in the sequencing library. The spatial information gained by the hepatocyte part of the cell duplet allowed the spatial inference of the attached endothelial cell (Halpern et al., 2018). Even though this approach contributes a first approximation of liver endothelium reconstruction, the data remain sparse and other NPCs such as cholangiocytes or immune cells were not considered. Thus, future work will be needed to improve spatial mapping of NPCs, for instance by highly sophisticated machine learning based combination of single cell RNA-seq with advanced imaging solutions such as sequential FISH (Eng et al., 2019).

Even though the predominant spatial signature of hepatocytes is the localization along the porto-central axis, the ultimate modelling of hepatic tissue would require to additionally infer the second and third dimension of hepatocytes in the hepatic lobule (Figure 54). The second dimension can be seen as a radial

sequence of hepatocytes of the same zonation layer, while the third dimension consists out of stacked, repeating functional units. It is important to note that diffusion component 2 (Figure 37 A) most likely does not represent a spatial dimension of the hepatic lobule, but captures other sources of transcriptional variation such as inter-individual heterogeneity (Sup. Figure 11 D). With the porto-central axis representing the functional spatial separation within the liver lobule, there is only a weak intra-layer cell to cell variability of hepatocytes localized in the same hepatic zone (Halpern et al., 2017). Thus, using existing approaches, modelling of the second and third spatial dimensions based on gene expression variability is probably not feasible. However, spatial reconstruction of hepatocytes into their functional units along the porto-central axis will already allow the elucidation of many intriguing questions of liver biology.
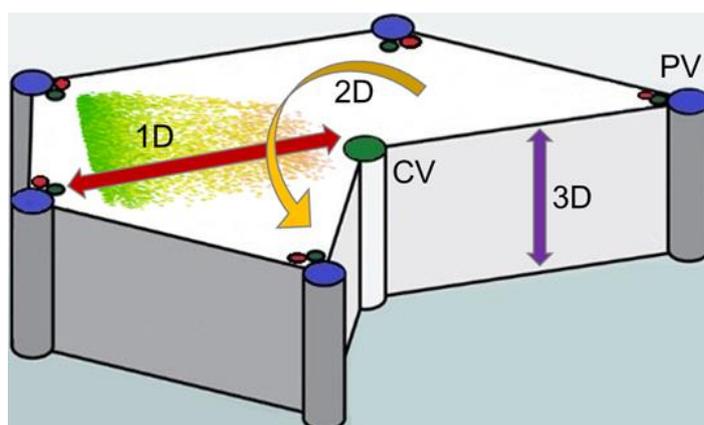
Figure 54: Three-dimensional illustration of the hepatic lobule. The first dimension (1D, red arrow) represents the proto-central axis. The second dimension (2D, orange arrow) is composed of the concentric layers of hepatocytes, while the third dimension (3D, purple arrow) consists out of stacked hexagonal subunits.

Interestingly, an organ scale, zone-specific quantification of the onset of steatosis, which is in contrast to human localized in the periportal zone, using sequential slices for hematoxylin eosin and Glutamine Synthetase (GLUL) stainings in mouse liver, showed distinct zonal onset of lipid droplet accumulation, but also revealed a considerable heterogeneity regarding the extent of steatosis across the organ (Schwen et al., 2016). This finding raises the question whether underlying expression programs and in particular zonated expression signatures show this heterogeneity as well. This might not only be important for the proper spatial reconstruction of steatotic liver tissue, but could in principle also play a role in healthy liver. To date, systematic studies regarding the variability of liver zonation programs in the mammalian liver are lacking. An interesting approach that might contribute to the better understanding of hepatocyte variability in 3 dimensional liver tissue could be based on microdissection of hepatocytes of the second and third hepatic dimensions. Capturing hepatocytes along the porto-central axis in one microdissection at multiple locations in the hepatic lobule would allow the assessment of hepatocyte variability along the second radial dimension (Figure 54). Microdissection of the third dimension would probably be more challenging, as it might be rather difficult to visually identify on the cryosection. In addition to a better understanding of intra-zonal hepatocyte variability allowing more accurate spatial reconstruction, resulting data might serve as a valuable reference to design methods for the spatial deconvolution of bulk RNA-seq data, which are typically generated from hepatocytes of all zones.

Although spatial reconstruction of single cells along the porto-central axis appears to be rather robust, the identification of zonated gene expression from single cells assigned to hepatic zones remains challenging (Figure 39). Based on Kruskal-Wallis statistics, the study by Halpern et al. (2017) on mouse liver detected about 50 % of expressed genes to be significantly zonated. This is several orders of magnitude higher than was previously reported for mouse (Saito et al., 2013). The analysis of zonated gene expression of the human single cell dataset was adapted from the strategy by Halpern et al., with the exception of using all cells in the respective zone instead of only the most extreme layers (cells directly neighboring to portal field or central vein), and yielded almost twice as many zonated genes as the laser capture microdissection based approach (Figure 39 A). On the one hand this might be attributed to the higher resolution of reconstructed single cell data. However, LCM can be considered as a kind of gold standard for spatial analysis, as it retains spatial information of samples and does not rely on bioinformatical pseudospace inference. In addition, the low overlap of detected zonated genes derived from the spatially reconstructed dataset and the LCM dataset strongly suggests an overestimation of zonated gene expression in the single cell dataset. The approach by Halpern et al., which relies on Kruskal-Wallis test between the expression levels of pericentral and periportal cells, might not be the most appropriate method for differential expression analysis between hepatic zones at single cell resolution. Thus, future work is needed to develop more advanced algorithms for differential expression analysis of spatially reconstructed single cell data, which should ideally consider reconstructed pseudospace as single cell localization probabilities, thereby taking uncertainties and inaccuracies of spatial mapping into account to obtain more robust results.

Still, interspecies analysis of one-to-one matching orthologues identified 113 genes with conserved zonation signatures (Figure 39 A), comprising highly expressed landmark genes, such as GLUL, CYP2E1, SDS and APOF, but also morphogens like LGR5 and SOX9 with low transcript levels. The remaining 202 one-to-one matching orthologues zonated in human were not detected as zonated in the mouse. Most of these non-corresponding genes showed rather low expression level in the mouse single cell dataset as compared to our deeply sequenced human data. Thus, differences in zonated gene expression between human and mouse might be attributed to technical factors, such as the low expression profiling efficiency of single cell datasets. Therefore, a more effective interspecies analysis should rely on the same method, ideally on the more robust LCM based approach, which would preclude technical artifacts and allow a more comprehensive data interpretation. As the genes with conserved zonation patterns comprise enzymes with known metabolic zonation and importantly morphogens associated to Wnt and Notch signaling, the interspecies analysis indicates similarities between the regulation of hepatic zonation in human and mouse. This is crucial for the translational interpretation of findings derived from mouse experiments, such as the lineage tracing based identification of a TBX3 expressing pericentral hepatocyte subpopulation as potential hepatic stem cell niche under steady state conditions (Wang et al., 2015), which could be relevant in the human as well (Figure 24).

Spatial reconstruction of high-throughput single cell RNA sequencing data and LCM based RNA-seq can be considered as complementary approaches. In a more complex biological setting, containing additional sources of transcriptional variation, spatial reconstruction may be more challenging. The proper identification of genes driving individual sources of variation, such as spatial, temporal, inter-individual differences, as well as corresponding genes need to be identified. Based on our previous study (Brosch et al., 2018), the

genes that are major contributors to spatial variance in a liver single cell dataset can be isolated and used for spatial reconstruction of the hepatic lobule. Nevertheless, this approach has certain limitations. For instance, it was shown that in liver cirrhosis hepatic zonation profiles are disturbed (Figure 27 A) and pericentral hepatocytes express periportal-like profiles (Figure 28 A). Thus, zonation markers defined by the analysis of healthy tissue (Figure 16 B) are not suited for proper reconstruction of cirrhotic tissue. A possible solution could be the complementary analysis of scRNA-seq data with matching LCM-RNA-seq data from the same sample, which might yield genes with intact zonation profiles that could be used to anchor diffusion map based tissue reconstruction as described for healthy liver tissue (Figure 37 A). Thus, in an ideal experimental set-up, single cell RNA-seq and LCM RNA-seq would be performed on the same sample, thereby decreasing unwanted variability and enhancing accuracy of spatial reconstruction. However, since in most cases this will not be feasible due to cost restraints, using our phenotype-specific LCM based reference dataset as anchor for spatial reconstruction of new single cell RNA-seq datasets represents a readily applicable alternative. Finally it should be noted that the application of LCM anchored diffusion map based pseudospace inference is not restricted to liver tissue, but could be used to capture spatial covariates in other tissues with spatial expression gradients, such as in intestinal epithelium (Moor et al., 2018). In addition, this approach would also be applicable to pseudospace inference of epigenetic single cell datasets.

4.3 Epigenomic characterization of *in vitro* differentiated HLCs

The *in vitro* differentiation of induced pluripotent stem cells (iPSCs) into hepatocyte-like cells (HLCs) shows promise as a future alternative to cryopreserved primary human hepatocytes (PHH, Gebhardt et al., 2003; Godoy et al., 2013; Wang et al., 2017), which represent the current gold standard for toxicological and pharmacological *in vitro* studies (Hengstler et al., 2000). Although there is room for improvement (Godoy et al., 2018), *in vitro* differentiated HLCs have already proven useful. For instance, a patient-specific HLC system was recently used in a drug screen for the treatment of homozygous familial hypercholesterolemia, a heritable metabolic liver disease (Cayo et al., 2017). However, many studies aiming to further optimize differentiation protocols are based on few marker genes (Asumda et al., 2018; Wang et al., 2017), which does not allow a comprehensive interpretation of the HLC differentiation state. Indeed, microarray and NGS based studies concluded that current differentiation protocols yield HLCs that start to emulate a PHH phenotype, but still show massive expression differences compared to PHH (Cahan et al., 2014; Camp et al., 2017; Godoy et al., 2015, 2018). Interestingly, HLCs apparently rather resemble fetal hepatocytes (Camp et al., 2017), and also show mixed characteristics of other cell-types, such as the expression of colon markers (Godoy et al., 2015).

In principle, the presented analysis builds on a previous publication based on gene expression microarray analysis (Godoy et al., 2015), which discovered a set of transcription factors, including CAR, FXR and PXR, as putative key TFs to improve HLC differentiation (Figure 55 A). Moreover, markers of unfavorable expression signatures like CDX2 and KLF5 were identified (Godoy et al., 2015). Hepatic differentiation is a complex process, regulated by interacting signaling pathways and stage-specific cascades of transcription factors. Thus, it seems unlikely that the massive gene expression differences observed between HLCs and PHH (Figure 47 B) would be attributed only to the activity of few TFs. Up to 85.9 % of expressed genes in the dataset showed significant differences, which suggests fundamental disparity also on the level of epigenetic mechanisms, which play a crucial role in the regulation of the TFs involved in hepatic differentiation (Snykers et al., 2009). For instance, promoter regions of certain stem cell marker genes such as OCT4 and NANOG are methylated with the onset of differentiation (Yeo et al., 2007). Moreover, it has been shown that chromatin remodeling agents such as HDAC inhibitors can improve hepatic differentiation (Ye et al., 2016). Thus, we aimed to achieve a comprehensive transcriptional and epigenomic characterization of the *in vitro* differentiation process. To this end, we analyzed besides iPSCs and HLCs, as start and end point of differentiation, also definitive endoderm as an intermediate point of cell fate decision (Figure 40). In addition, a detailed comparison of HLCs and PHH was performed on single cell level to gain a better understanding of underlying regulatory networks (Figure 41). Instead of using a high throughput method to profile expression signatures of thousands of single cells, we manually isolated single cells into 96 well plates and performed a Smartseq2 (Picelli et al., 2013) scRNA-seq approach, which allowed the highly sensitive analysis of full-length transcriptomes with saturated per cell sequencing coverage. Moreover, compatible external single fetal hepatocyte transcriptome data (Camp et al., 2017) provided the means to better characterize the differentiation state of HLCs.
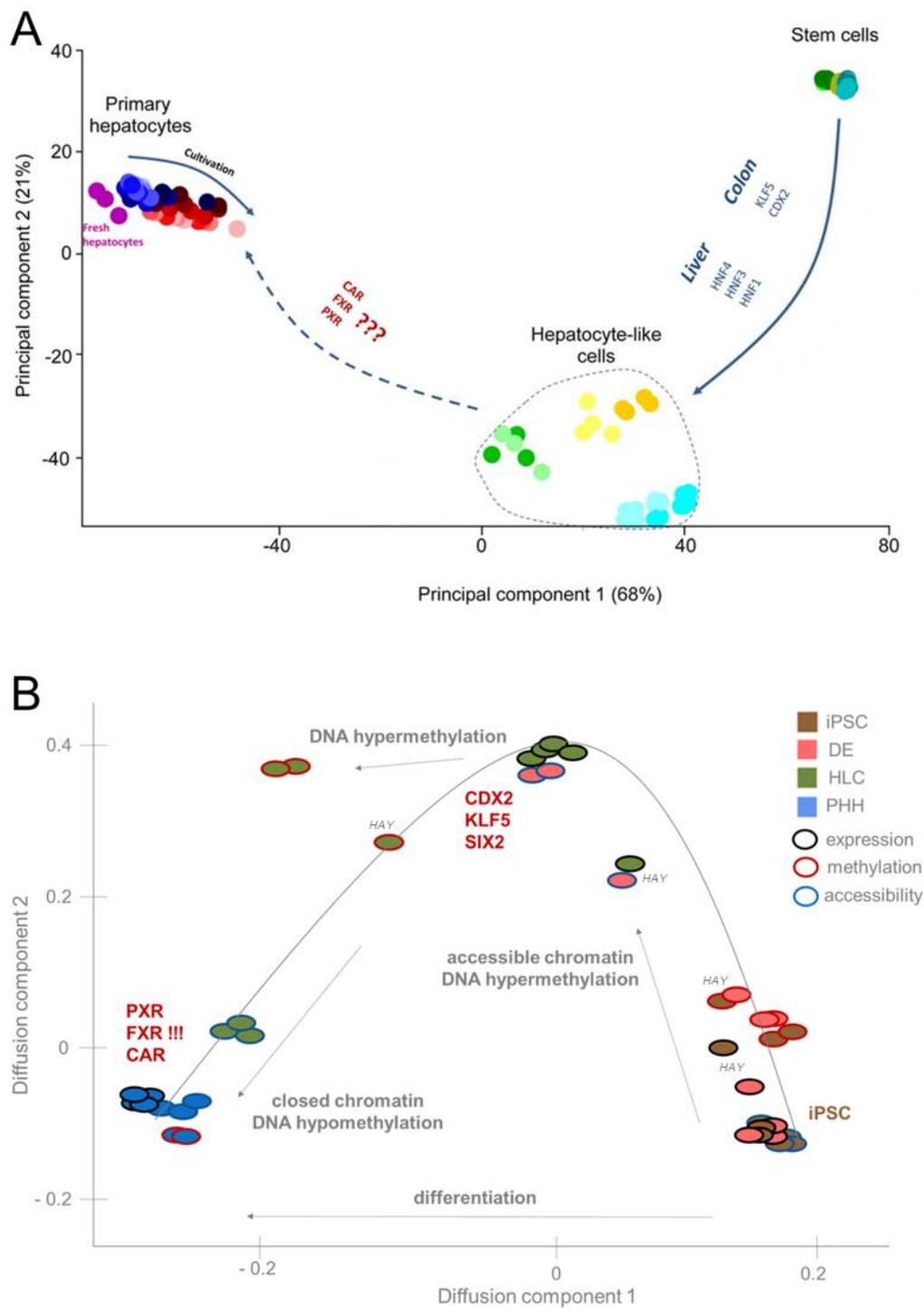
Figure 55: *In vitro* differentiation of HLCs. A – PCA of the 1,000 most variable transcript identified by gene expression microarray analysis of stem cells, HLCs and human hepatocytes (Godoy et al., 2015, modified). B – Diffusion map integrating gene expression, DNA methylation and chromatin accessibility throughout *in vitro* differentiation of iPSCs over DE to HLCs in comparison with PHH. Diffusion components were calculated separately for each assay based on the 1,000 most variable transcripts (RNA-seq), the 5,000 most variable CpGs with coverage above 5 in all samples (DNA methylation), and the 50,000 most variable 100 bp regions (ATAC-seq). Resulting coordinates were plotted into one panel. Diffusion components of DNA methylation were inverted prior to plotting. iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue. Border colors reflect the depicted assay with gene expression in black, DNA methylation in red and chromatin accessibility in blue. HAY samples are labelled, all other points represent CEL samples.

Two different protocols for *in vitro* HLC differentiation in 2-dimensional cell culture were assessed – the Cellartis (CEL) protocol, which is already commercially available at Takara Bio, and a protocol developed by the group of David Hay (University of Edinburgh, Wang et al., 2017). Prior to differentiation, iPSCs were cultivated for 4 days to obtain clean stem cell populations. Then, differentiation over DE to HLCs was completed within 25 days using the CEL protocol, while the HAY protocol achieved differentiation after only 18 days (Figure 40). For both protocols, a mixture of Laminin 521 and Laminin 111, which was optimized in preceding experiments by Patrick Nell and David Feuerborn (IfADo), was used to mimic the extracellular matrix (ECM) of liver tissue. However, further comparison of the two protocols regarding differentiation inducing compounds is not possible, as the detailed components of the CEL protocol remain confidential. Moreover, both on transcriptional and epigenetic level HAY HLCs appeared to be less differentiated than CEL HLCs (Figure 55 B), and also featured higher expression levels of undesired genes such as CDX2 (Figure 43). Thus, rather than comparing both protocols, more detailed analysis focused on the apparently more promising CEL *in vitro* differentiation system.

Interestingly, pseudotime analysis of single cell transcriptomes indicates separate differentiation trajectories for *in vivo* and *in vitro* hepatogenesis (Figure 44), which is in line with distinct underlying regulatory networks (Figure 45). Moreover, TF network inference confirmed that HLCs resemble fetal hepatocytes more closely than adult PHH (Camp et al., 2017). It should also be noted that a considerable batch to batch variability of HLCs was observed. While replicates 2 and 3 showed a very similar differentiation state, replicate 1 appeared to be less differentiated (Figure 43), although all replicates, which were generated from the same iPSC line, should reach a similar differentiation state. Thus, optimized differentiation protocols should also aim to eliminate batch to batch variability to ensure the applicability of HLCs as a reliable tool for *in vitro* studies. In addition, putative regulators of a hepatocyte-like cell fate were identified, comprising among others FXR, PXR and CAR (Figure 46), which were already described as potential key factors by Godoy et al. (2015). Interestingly, PXR and CAR were exclusively active in PHH, while FXR activity was also observed in the subpopulation of HLCs that expressed ALB. Thus, FXR, which encodes the nuclear farnesoid X receptor that can be activated by bile acids (Lefebvre et al., 2009), might be a direct regulator of ALB expression in HLCs, while PXR and CAR are probably rather involved in the regulation of other hepatocyte-specific programs. On the other hand, previously reported potential regulators of undesired gene expression profiles such as CDX2 and KLF5 were confirmed, while an additional set of unfavorable regulons including FOSL2 and GATA5 were discovered as well (Sup. Table 13). It should be noted that activity of both hepatocyte and intestine specific regulons could be detected within the same single hepatocyte-like cell (Figure 46). Thus, instead of a clear transition from DE to a hepatocyte-like phenotype, HLCs apparently rather represent a heterogeneous phenotype showing some characteristics of hepatocytes, but also expression signatures of other DE derived cell-types.

This mixed phenotype of HLCs might be explained by the lack of a properly regulated microenvironment in cell culture. Definitive endoderm represents a major point of cell fate decision that still harbors the capacity to differentiate into multiple cell-types. During *in vivo* development these cell fate specifications are regulated by a complex interplay of temporal and spatial TF cascades from surrounding tissues and cell-types. For instance, a Wnt signaling controlled posterior subpopulation of endoderm expresses CDX2 that promotes

the formation of the gastrointestinal tract (Stringer et al., 2012), while upon inhibition of Wnt signaling the ventral forgut endoderm gives rise to a population of bipotent hepato-pancreatic progenitors (Russell and Monga, 2018). As *in vitro* differentiation seems to recapitulate several aspects of *in vivo* hepatogenesis, the use of Wnt antagonists might lead to a more targeted differentiation of HLCs towards a hepatocyte-like phenotype.

Intriguingly, gene expression, DNA methylation and chromatin accessibility followed similar trajectories modelling the HLC *in vitro* differentiation process (Figure 55 B). Bulk RNA-seq confirmed and extended the previously reported massive gene expression differences between HLCs and PHH (Godoy et al., 2015). During differentiation of iPSCs to DE a general increase of chromatin accessibility was observed, while the chromatin structure became again more closed in HLCs (Figure 49 A). Yet, PHH showed an even lower number of accessible chromatin regions, which is in line with the reduced number of expressed genes. Overall, differential chromatin accessibility was considerable more indicative of gene expression changes than differential DNA methylation (Figure 51 E – H, Figure 52 B). Moreover, a pronounced DNA hypermethylation was observed in cultured cells, which increased with cultivation time (Figure 48 A, Sup. Figure 15 B), while *in vivo* derived PHH showed considerably lower DNA methylation levels (Figure 48 C). Nevertheless, it should be noted that in comparison with PHH cultured cells also featured some hypomethylated regions. In HLCs, hypermethylated regions were more frequently localized in intergenic regions (Figure 48 D), and an increase of highly methylated domains was obvious on a genome-wide scale (Sup. Figure 15 B). Moreover, while differential chromatin accessibility frequently correlated with gene expression changes, DNA methylation patterns were rarely indicative of gene expression (Figure 52 B). Taken together, these observations suggest an undirected hypermethylation of HLCs, which might even persist when differentiated HLCs are cultured for longer periods. This is in line with a previous study reporting aberrant DNA hypermethylation during the differentiation of embryonic stem cells, which apparently resembled DNA methylation signatures frequently observed in primary tumors (Meissner et al., 2008). As differentiated HLCs can be maintained in cell culture for more than 30 days (Carpentier et al., 2016), this is of great interest regarding their applicability for long-time studies. Ideally, differentiated HLCs should show a stable phenotype during maintenance cell culture, as cultivation induced epigenetic and transcriptional changes might heavily bias experiments such as drug screenings. Yet, it should be noted that the cultivation of PHH also results in transcriptional differences (Godoy et al., 2015). Interestingly, cultured PHH resemble gene expression profiles of certain liver disease phenotypes such as NAFLD and cirrhosis (Godoy et al., 2016), which highlights the overall need for improved *in vitro* systems. Thus, further epigenetic profiling of differentiated HLCs maintained for an extended period in culture should unravel the stability of the HLC differentiation state over time, which might suggest additional cultivation measures to further optimized the HLC phenotype.

Furthermore, the pronounced hypermethylation of cultured cells can likely be explained by the strong upregulation of DNMTs during *in vitro* differentiation (Sup. Figure 16). The highest expression level of DNMTs is observed in iPSCs and decreases over DE to HLCs. Interestingly, the de novo DNA methyltransferases DNMT3a and DNMT3b are upregulated in all stages of *in vitro* cultivation, while maintenance DNMT1 is only upregulated in iPSCs and DE, which are still in a proliferative phase. It should be noted that the methyl-cytosine dioxygenase TET1 is also upregulated in cultured cells. TET enzymes

mediate active DNA de-methylation by iterative oxidation of 5-methyl-cytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formyl-cytosine (5fC) and 5-carboxy-cytosine (5cC, Ito, 2011; Tahiliani et al., 2009; Wossidlo, 2011). The upregulation of TET1 expression might be reflected in the hypomethylated regions detected in cultured cells (Figure 48 A). However, the bisulfite conversion method used for preparation of RRBS libraries (chapter 2.2.1.2) does not allow to disentangle these oxidized derivates. Actually, 5hmC is not distinguished from 5mC (Huang et al., 2010), while 5fC and 5cC are mostly detected as unmethylated CpGs (Tierling et al., 2018). Thus, part of the observed DNA hypermethylation might be attributed to 5hmC, which would suggest high DNA methylation dynamics during *in vitro* differentiation. Hence, the HLC system would also represent an in interesting model to assess DNA methylation and de-methylation dynamics by using specialized sequencing methods that allow the analysis of oxidative 5mC derivates at base resolution (Skvortsova et al., 2017).

The integration of gene expression, chromatin accessibility and DNA methylation allowed the definition of gene sets with distinct epigenetic patterns. Remarkably, only for genes that showed a correlation of upregulated gene expression and increasing chromatin accessibility, but were also associated with DNA hypermethylation in HLCs, specific key regulators could be identified (Figure 52 D, cluster 5 and 8). For differentially expressed genes that were associated with other epigenetic patterns, such as upregulated gene expression, less accessible chromatin and DNA hypermethylation in HLCs (Figure 52 D, cluster 6), high numbers of individual TFs were detected as putative regulators. Interestingly, these TFs were highly enriched for proteins involved in chromatin remodeling, which suggests a high epigenetic dynamic associated with these genes. Moreover, SIX2 was detected as a novel putative master regulator of genes hypermethylated, more accessible and strongly upregulated in HLCs. This TF does not seem to play a role during *in vivo* hepatogenesis, but rather during the development of other organs such as stomach and kidney (Ohto et al., 1998; Self et al., 2006, 2009). During kidney development SIX2 seems to keep multipotent progenitor cells in an undifferentiated state (Kobayashi et al., 2008), which suggests that SIX2 might be involved in retaining *in vitro* differentiated HLCs in a hepatoblast-like state. Thus, the downregulation of SIX2 during *in vitro* differentiation might benefit HLCs towards a more adult hepatocyte-like phenotype. In addition, for genes hypomethylated, more accessible and upregulated in PHH FXR was confirmed as the key regulator, which highlights FXR as the most promising target for cell culture interventions in order to upregulated PHH expression patterns in HLCs.

Taken together, the observed epigenetic patterns during *in vitro* differentiation (Figure 55 B) suggest the following model: The starting population of induced pluripotent stem cells already features a quite accessible chromatin structure. During early events of differentiation from iPSCs to DE the overall chromatin structure becomes even more accessible in order to allow a wide-spread transcriptional activation of a multitude of genes. This would also explain the predominant DNA hypermethylation in cultured cells, as well as its lack of association with differential gene expression, as a kind of unspecific bystander effect due to the upregulation of DNMT expression. Then, during further differentiation from DE to HLCs the chromatin structure becomes more closed again. Actually, the progressive restriction of chromatin accessibility might be a key feature of cellular differentiation (Atlasi and Stunnenberg, 2017), defining the increasingly specialized and tightly regulated expression profiles of terminally differentiated cell-types. This is in line with the strong upregulation of non-hepatocyte specific genes in HLCs, as they still feature a more accessible

chromatin structure than PHH. Hence, while on the level of chromatin accessibility HLCs would need to be more restricted, e.g. more tightly controlled, on the level of DNA methylation the persisting DNA hypermethylation of HLCs indicates some kind of overshooting effect. Thus, possible cell culture interventions also should entail the manipulation of epigenetic modifiers to reach a similar epigenetic state as PHH (Figure 55 B).

Furthermore, it is important to note the complex interplay of hepatocytes with non-parenchymal cell-types and extracellular matrix components, as well as the intricate network of oxygen gradient and morphogen interactions (Kietzmann, 2017) that shape the specialized phenotype of spatially distinct hepatocytes in liver tissue (chapter 4.1). This tightly controlled 3-dimensional microenvironment is most probably key to obtain fully differentiated adult hepatocytes. However, modelling this complex microenvironment in cell culture remains a major challenge, which is not feasible in conventional 2-dimensional cell culture. Several attempts including co-culture with other cell-types as well as 3-dimensional organoid cultures have been made (Prior et al., 2019). For instance, Camp et al. (2017) report liver bud organoids differentiated from a co-culture of hepatic endoderm, endothelial and mesenchymal cell, which model human liver bud development *in vitro*. Although organoid-derived hepatocytes exhibited only an immature hepatoblast-like phenotype (Camp et al., 2017), organoids certainly represent a promising approach to achieve a more reliable liver-like *in vitro* system. In this context, it would be interesting to study the epigenome of organoids during differentiation and compare it to the epigenetic profiles of 2-dimensional HLC cultures. Then, the cell culture interventions developed from the characterization of 2-dimensional HLC differentiation presented in this thesis might be of use in the optimization of liver organoid cultures as well.

In summary, the presented integrative analysis suggests an epigenetic model of *in vitro* differentiation (Figure 55 B), and furthermore represents the basis for the development of reasonable cell culture interventions. It should be noted that sole over-expression of TFs such as FXR might indeed result in a slightly improved HLC phenotype, but the massive gene expression differences between HLCs and PHH will most probably not be overcome. A more promising – though challenging – approach would be a combination of several interventions, such as overexpression of key TFs together with downregulation of unfavorable colon-specific TFs and additional interventions on the epigenetic level. For instance, HDAC inhibitors such as TSA and valproic acid have been shown to exhibit a positive effect on hepatic *in vitro* differentiation, although they do not exhibit targeted effects (Dong et al., 2013; Karantzali et al., 2008; Kondo et al., 2014). However, these compounds exhibit an undirected effect on chromatin accessibility (Li and Sun, 2019). In the case of HLCs a targeted restriction of chromatin accessibility would be favorable. To this end, the increasingly popular CRISPR/Cas9 technology might offer possibilities to perform targeted epigenome editing (Xie et al., 2018). In addition, the observed DNA hypermethylation in cultured cells could be prevented by directly targeting the expression of DNMTs. Moreover, supplementing the differentiation medium with certain modulating compounds might have a beneficial effect as well. For instance, the activity of FXR could be increased by bile acids, while Wnt signaling induced CDX2 expression might be reduced by Wnt antagonists in the medium. In order to quickly assess the success of these interventions the use of differentiation marker genes such as HNF4α, ALB, AFP and CDX2 by quantitative PCR would be a reasonable first step. Nevertheless, in order to obtain a conclusive view on the state of HLC differentiation bulk RNA-seq should be the method of choice.

In conclusion, the comprehensive integrated analysis of transcriptomic and epigenomic data together with an in-depth analysis of single cell transcriptomes of selected cells represents a workflow that might be applicable to other experimental settings and could provide deeper insights into many biological questions beyond the *in vitro* differentiation of HLCs. In addition, the putative key regulators identified by this analysis provide now the means to develop and assess reasonable cell culture interventions that could readily contribute to the improvement of HLC differentiation and maintenance in cell culture.

5. References

Abarca-Gómez, L., Abdeen, Z.A., Hamid, Z.A., Abu-Rmeileh, N.M., Acosta-Cazares, B., Acuin, C., Adams, R.J., Aekplakorn, W., Afsana, K., Aguilar-Salinas, C.A., et al. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. The Lancet *390*, 2627–2642.

Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P.G., Manova-Todorova, K., Leversha, M., Hogg, N., Seshan, V.E., et al. (2012). A CXCL1 Paracrine Network Links Cancer Chemoresistance and Metastasis. Cell *150*, 165–178.

Achim, K., Pettit, J.-B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat. Biotechnol. *33*, 503–509.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F., and Grün, D. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87.

Akhtar, A. (2015). The Flaws and Human Harms of Animal Experimentation. Camb. Q. Healthc. Ethics *24*, 407–419.

Alison, M.R., and Lin, W.-R. (2011). Hepatocyte turnover and regeneration: Virtually a virtuoso performance. Hepatology *53*, 1393–1396.

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., and Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. Indian J. Microbiol. *56*, 394–404.

Andrews, T.S., and Hemberg, M. (2019). False signals induced by single-cell imputation. F1000Research *7*, 1740.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

Anstee, Q.M., and Goldin, R.D. (2006). Mouse models in non-alcoholic fatty liver disease and steatohepatitis research: Mouse models in NAFLD and NASH research. Int. J. Exp. Pathol. *87*, 1–16.

Arriazu, E., Ruiz de Galarreta, M., Cubero, F.J., Varela-Rey, M., Pérez de Obanos, M.P., Leung, T.M., Lopategi, A., Benedicto, A., Abraham-Enachescu, I., and Nieto, N. (2014). Extracellular Matrix and Liver Disease. Antioxid. Redox Signal. *21*, 1078–1097.

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat. Methods *11*, 1138–1140.

Asumda, F.Z., Hatzistergos, K.E., Dykxhoorn, D.M., Jakubski, S., Edwards, J., Thomas, E., and Schiff, E.R. (2018). Differentiation of hepatocyte-like cells from human pluripotent stem cells using small molecules. Differentiation *101*, 16–24.

Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. Nat. Rev. Genet. *18*, 643–658.

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. Nat. Chem. Biol. *11*, 555–557.

Bardag-Gorce, F., French, B.A., Nan, L., Song, H., Nguyen, S.K., Yong, H., Dede, J., and French, S.W. (2006). CYP2E1 induced by ethanol causes oxidative stress, proteasome inhibition and cytokeratin aggresome (Mallory body-like) formation. Exp. Mol. Pathol. *81*, 191–201.

Barthel, A., Schmoll, D., and Unterman, T.G. (2005). FoxO proteins in insulin action and metabolism. Trends Endocrinol. Metab. TEM *16*, 183–189.

Baubec, T., and Schübeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. Curr. Opin. Genet. Dev. *25*, 85–92.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

Behjati Ardakani, F., Kattler, K., Nordström, K., Gasparoni, N., Gasparoni, G., Fuchs, S., Sinha, A., Barann, M., Ebert, P., Fischer, J., et al. (2018). Integrative analysis of single-cell expression data reveals distinct regulatory states in bidirectional promoters. Epigenetics Chromatin *11*.

Bell, C.C., Hendriks, D.F.G., Moro, S.M.L., Ellis, E., Walsh, J., Renblom, A., Fredriksson Puigvert, L., Dankers, A.C.A., Jacobs, F., Snoeys, J., et al. (2016). Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. Sci. Rep. *6*.

Bell, C.C., Lauschke, V.M., Vorrink, S.U., Palmgren, H., Duffin, R., Andersson, T.B., and Ingelman-Sundberg, M. (2017). Transcriptional, Functional, and Mechanistic Comparisons of Stem Cell–Derived Hepatocytes, HepaRG Cells, and Three-Dimensional Human Hepatocyte Spheroids as Predictive In Vitro Systems for Drug-Induced Liver Injury. Drug Metab. Dispos. *45*, 419–429.

Bell, C.C., Dankers, A.C.A., Lauschke, V.M., Sison-Young, R., Jenkins, R., Rowe, C., Goldring, C.E., Park, K., Regan, S.L., Walker, T., et al. (2018). Comparison of Hepatic 2D Sandwich Cultures and 3D Spheroids for Long-term Toxicity Applications: A Multicenter Study. Toxicol. Sci. *162*, 655–666.

Benedict, M., and Zhang, X. (2017). Non-alcoholic fatty liver disease: An expanded review. World J. Hepatol. *9*, 715.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bertola, A., Bonnafous, S., Anty, R., Patouraux, S., Saint-Paul, M.-C., Iannelli, A., Gugenheim, J., Barr, J., Mato, J.M., Le Marchand-Brustel, Y., et al. (2010). Hepatic Expression Patterns of Inflammatory and Immune Response Genes Associated with Obesity and NASH in Morbidly Obese Patients. PLoS ONE *5*, e13577.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat. Struct. Mol. Biol. *14*, 1008–1016.

Biesinger, J., Wang, Y., and Xie, X. (2013). Discovering and mapping chromatin states using a tree hidden Markov model. BMC Bioinformatics *14 Suppl 5*, S4.

Birchmeier, W. (2016). Orchestrating Wnt signalling for metabolic liver zonation. Nat. Cell Biol. *18*, 463–465.

Blanco, A., and Blanco, G. (2017). Medical biochemistry (London, United Kingdom ; San Diego, CA: Academic Press, an imprint of Elsevier).

Bouazoune, K., Miranda, T.B., Jones, P.A., and Kingston, R.E. (2009). Analysis of individual remodeled nucleosomes reveals decreased histone-DNA contacts created by hSWI/SNF. Nucleic Acids Res. *37*, 5279–5294.

Boulter, L., Govaere, O., Bird, T.G., Radulescu, S., Ramachandran, P., Pellicoro, A., Ridgway, R.A., Seo, S.S., Spee, B., Van Rooijen, N., et al. (2012). Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. Nat. Med. *18*, 572–579.

Bralet, M.P., Branchereau, S., Brechot, C., and Ferry, N. (1994). Cell lineage study in the liver using retroviral mediated gene transfer. Evidence against the streaming of hepatocytes in normal liver. Am. J. Pathol. *144*, 896–905.

Bronner, C., Krifa, M., and Mousli, M. (2013). Increasing role of UHRF1 in the reading and inheritance of the epigenetic code as well as in tumorogenesis. Biochem. Pharmacol. *86*, 1643–1649.

Brosch, M., Kattler, K., Herrmann, A., von Schönfels, W., Nordström, K., Seehofer, D., Damm, G., Becker, T., Zeissig, S., Nehring, S., et al. (2018). Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. Nat. Commun. *9*.

Brunt, E.M., Janney, C.G., Di Bisceglie, A.M., Neuschwander-Tetri, B.A., and Bacon, B.R. (1999). Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am. J. Gastroenterol. *94*, 2467–2474.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Burger, L., Gaidatzis, D., Schübeler, D., and Stadler, M.B. (2013). Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. *41*, e155–e155.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Büttner, M., Miao, Z., Wolf, F.A., Teichmann, S.A., and Theis, F.J. (2019). A test metric for assessing single-cell RNA-seq batch correction. Nat. Methods *16*, 43–49.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: Network Biology Applied to Stem Cell Engineering. Cell *158*, 903–915.

Caldwell, S., Ikura, Y., Dias, D., Isomoto, K., Yabu, A., Moskaluk, C., Pramoonjago, P., Simmons, W., Scruggs, H., Rosenbaum, N., et al. (2010). Hepatocellular ballooning in NASH. J. Hepatol. *53*, 719–723.

Camp, J.G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., Kanton, S., Kageyama, J., Damm, G., Seehofer, D., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. Nature.

Cardinale, V., Wang, Y., Carpino, G., Cui, C.-B., Gatto, M., Rossi, M., Bartolomeo Berloco, P., Cantafora, A., Wauthier, E., Furth, M.E., et al. (2011). Multipotent stem/progenitor cells in human biliary tree give rise to hepatocytes, cholangiocytes, and pancreatic islets. Hepatology *54*, 2159–2172.

Carpentier, A., Nimgaonkar, I., Chu, V., Xia, Y., Hu, Z., and Liang, T.J. (2016). Hepatic differentiation of human pluripotent stem cells in miniaturized format suitable for high-throughput screen. Stem Cell Res. *16*, 640–650.

Cayo, M.A., Mallanna, S.K., Di Furio, F., Jing, R., Tolliver, L.B., Bures, M., Urick, A., Noto, F.K., Pashos, E.E., Greseth, M.D., et al. (2017). A Drug Screen using Human iPSC-Derived Hepatocyte-like Cells Reveals Cardiac Glycosides as a Potential Treatment for Hypercholesterolemia. Cell Stem Cell *20*, 478-489.e5.

Cernea, S., Cahn, A., and Raz, I. (2017). Pharmacological management of nonalcoholic fatty liver disease in type 2 diabetes. Expert Rev. Clin. Pharmacol. *10*, 535–547.

Chen, P., Wang, R., Yue, Q., and Hao, M. (2018). Long non-coding RNA TTN-AS1 promotes cell growth and metastasis in cervical cancer via miR-573/E2F3. Biochem. Biophys. Res. Commun. *503*, 2956–2962.

Chen, T.-W., Li, H.-P., Lee, C.-C., Gan, R.-C., Huang, P.-J., Wu, T.H., Lee, C.-Y., Chang, Y.-F., and Tang, P. (2014). ChIPseek, a web-based analysis tool for ChIP data. BMC Genomics *15*, 539.

Choi, W.-M., Kim, M.-H., and Jeong, W.-I. (2019). Functions of hepatic non-parenchymal cells in alcoholic liver disease. Liver Res. *3*, 80–87.

Cohen, J.C., Horton, J.D., and Hobbs, H.H. (2011). Human Fatty Liver Disease: Old Questions and New Insights. Science *332*, 1519–1523.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Corlu, A., and Loyer, P. (2012). Regulation of the g1/s transition in hepatocytes: involvement of the cyclin-dependent kinase cdk1 in the DNA replication. Int. J. Hepatol. *2012*, 689324.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Dalton, S.R., and Bellacosa, A. (2012). DNA demethylation by TDG. Epigenomics *4*, 459–467.

Dan, Y.Y., Riehle, K.J., Lazaro, C., Teoh, N., Haque, J., Campbell, J.S., and Fausto, N. (2006). Isolation of multipotent progenitor cells from human fetal liver capable of differentiating into liver and mesenchymal lineages. Proc. Natl. Acad. Sci. U. S. A. *103*, 9912–9917.

Díaz, N., Kruse, K., Erdmann, T., Staiger, A.M., Ott, G., Lenz, G., and Vaquerizas, J.M. (2018). Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. Nat. Commun. *9*.

Dietrich, C.G. (2016). Molecular changes in hepatic metabolism and transport in cirrhosis and their functional importance. World J. Gastroenterol. *22*, 72.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dong, X., Pan, R., Zhang, H., Yang, C., Shao, J., and Xiang, L. (2013). Modification of histone acetylation facilitates hepatic differentiation of human bone marrow mesenchymal stem cells. PloS One *8*, e63405.

Doskočil, J., and Šorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochim. Biophys. Acta *55*, 953–959.

Drel, V.R., Mashtalir, N., Ilnytska, O., Shin, J., Li, F., Lyzogubov, V.V., and Obrosova, I.G. (2006). The leptin-deficient (ob/ob) mouse: a new animal model of peripheral neuropathy of type 2 diabetes and obesity. Diabetes *55*, 3335–3343.

Dupuis-Sandoval, F., Poirier, M., and Scott, M.S. (2015). The emerging landscape of small nucleolar RNAs in cell biology: Emerging landscape of small nucleolar RNAs. Wiley Interdiscip. Rev. RNA *6*, 381–397.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. *10*, 2709–2721.

Eissenberg, J.C., and Elgin, S.C.R. (2014). HP1a: a structural chromosomal protein regulating transcription. Trends Genet. *30*, 103–110.

Elmaouhoub, A., Dudas, J., and Ramadori, G. (2007). Kinetics of albumin- and alpha-fetoprotein-production during rat liver development. Histochem. Cell Biol. *128*, 431–443.

Eng, F.J., and Friedman, S.L. (2000). Fibrogenesis I. New insights into hepatic stellate cell activation: the simple becomes complex. Am. J. Physiol. Gastrointest. Liver Physiol. *279*, G7–G11.

Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature *568*, 235–239.

Fan, J.Y., Gordon, F., Luger, K., Hansen, J.C., and Tremethick, D.J. (2002). The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat. Struct. Biol. *9*, 172–176.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. Cell Rep. *10*, 1386–1397.

Farrelly, L.A., Thompson, R.E., Zhao, S., Lepack, A.E., Lyu, Y., Bhanu, N.V., Zhang, B., Loh, Y.-H.E., Ramakrishnan, A., Vadodaria, K.C., et al. (2019). Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. Nature.

Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. *4*, e1000039.

Fletcher, T.M., and Hansen, J.C. (1996). The nucleosomal array: structure/function relationships. Crit. Rev. Eukaryot. Gene Expr. *6*, 149–188.

Font-Burgada, J., Shalapour, S., Ramaswamy, S., Hsueh, B., Rossell, D., Umemura, A., Taniguchi, K., Nakagawa, H., Valasek, M.A., Ye, L., et al. (2015). Hybrid Periportal Hepatocytes Regenerate the Injured Liver without Giving Rise to Cancer. Cell *162*, 766–779.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U. S. A. *89*, 1827–1831.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003). Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1 Heterochromatic Complex for DNA Methylation-based Transcriptional Repression. J. Biol. Chem. *278*, 24132–24138.

Furuyama, K., Kawaguchi, Y., Akiyama, H., Horiguchi, M., Kodama, S., Kuhara, T., Hosokawa, S., Elbahrawy, A., Soeda, T., Koizumi, M., et al. (2011). Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. Nat. Genet. *43*, 34–41.

Gebhardt, R. (2014). Liver zonation: Novel aspects of its regulation and its impact on homeostasis. World J. Gastroenterol. *20*, 8491.

Gebhardt, R., Hengstler, J.G., Müller, D., Glöckner, R., Buenning, P., Laube, B., Schmelzer, E., Ullrich, M., Utesch, D., Nicola Hewitt, et al. (2003). New Hepatocyte In Vitro Systems for Drug Metabolism: Metabolic Capacity and Recommendations for Application in Basic Research and Drug Development, Standard Operation Procedures. Drug Metab. Rev. *35*, 145–213.

Gerets, H.H.J., Tilmant, K., Gerin, B., Chanteux, H., Depelchin, B.O., Dhalluin, S., and Atienzar, F.A. (2012). Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. Cell Biol. Toxicol. *28*, 69–87.

Gerhard, G.S., Legendre, C., Still, C.D., Chu, X., Petrick, A., and DiStefano, J.K. (2018). Transcriptomic Profiling of Obesity-Related Nonalcoholic Steatohepatitis Reveals a Core Set of Fibrosis-Specific Genes. J. Endocr. Soc. *2*, 710–726.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Dong, X., Parthun, M., and Jacob, S.T. (2002). Inhibitors of Histone Deacetylase and DNA Methyltransferase Synergistically Activate the Methylated Metallothionein I Promoter by Activating the Transcription Factor MTF-1 and Forming an Open Chromatin Structure. Mol. Cell. Biol. *22*, 8302–8319.

Gilgenkrantz, H., and Collin de l'Hortet, A. (2018). Understanding Liver Regeneration. Am. J. Pathol. *188*, 1316–1327.

Godoy, P., Hewitt, N.J., Albrecht, U., Andersen, M.E., Ansari, N., Bhattacharya, S., Bode, J.G., Bolleyn, J., Borner, C., Böttger, J., et al. (2013). 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. *87*, 1315–1530.

Godoy, P., Schmidt-Heck, W., Natarajan, K., Lucendo-Villarin, B., Szkolnicka, D., Asplund, A., Björquist, P., Widera, A., Stöber, R., Campos, G., et al. (2015). Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. J. Hepatol. *63*, 934–942.

Godoy, P., Widera, A., Schmidt-Heck, W., Campos, G., Meyer, C., Cadenas, C., Reif, R., Stöber, R., Hammad, S., Pütter, L., et al. (2016). Gene network activity in cultivated primary hepatocytes is highly similar to diseased mammalian liver tissue. Arch. Toxicol. *90*, 2513–2529.

Godoy, P., Schmidt-Heck, W., Hellwig, B., Nell, P., Feuerborn, D., Rahnenführer, J., Kattler, K., Walter, J., Blüthgen, N., and Hengstler, J.G. (2018). Assessment of stem cell differentiation based on genome-wide expression profiles. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *373*.

Gordillo, M., Evans, T., and Gouon-Evans, V. (2015). Orchestrating liver development. Development *142*, 2094–2108.

Grattan-Guinness. (1994). Companion encyclopedia of the history and philosophy of the mathematical sciences (London ; New York: Routledge).

Grossman, S.R., Engreitz, J., Ray, J.P., Nguyen, T.H., Hacohen, N., and Lander, E.S. (2018). Positional specificity of different transcription factor classes within enhancers. Proc. Natl. Acad. Sci. *115*, E7222–E7230.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat. Protoc. *6*, 468–481.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guillemette, B., Bataille, A.R., Gévry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005). Variant Histone H2A.Z Is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning. PLoS Biol. *3*, e384.

Guillouzo, A., Corlu, A., Aninat, C., Glaise, D., Morel, F., and Guguen-Guillouzo, C. (2007). The human hepatoma HepaRG cells: a highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. Chem. Biol. Interact. *168*, 66–73.

Guiro, J., and Murphy, S. (2017). Regulation of expression of human RNA polymerase II-transcribed snRNA genes. Open Biol. *7*, 170073.

Gunasekara, C.J., Scott, C.A., Laritsky, E., Baker, M.S., MacKay, H., Duryea, J.D., Kessler, N.J., Hellenthal, G., Wood, A.C., Hodges, K.R., et al. (2019). A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol. *20*.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. *23*, 2126–2135.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics *31*, 2989–2998.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352–356.

Halpern, K.B., Shenhav, R., Massalha, H., Toth, B., Egozi, A., Massasa, E.E., Medgalia, C., David, E., Giladi, A., Moor, A.E., et al. (2018). Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. Nat. Biotechnol.

Hardwick, J.P., Osei-Hyiaman, D., Wiland, H., Abdelmegeed, M.A., and Song, B.-J. (2009). PPAR/RXR Regulation of Fatty Acid Metabolism and Fatty Acid omega-Hydroxylase (CYP4) Isozymes: Implications for Prevention of Lipotoxicity in Fatty Liver Disease. PPAR Res. *2009*, 952734.

Hardy, S., Jacques, P.-E., Gévry, N., Forest, A., Fortin, M.-E., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet. *5*, e1000687.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hart, S.N., Li, Y., Nakamoto, K., Subileau, E., Steen, D., and Zhong, X. (2010). A comparison of whole genome gene expression profiles of HepaRG cells and HepG2 cells to primary human hepatocytes and human liver tissues. Drug Metab. Dispos. Biol. Fate Chem. *38*, 988–994.

Heard, E., Clerc, P., and Avner, P. (1997). X-chromosome inactivation in mammals. Annu. Rev. Genet. *31*, 571–610.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hengstler, J.G., Utesch, D., Steinberg, P., Platt, K.L., Diener, B., Ringel, M., Swales, N., Fischer, T., Biefang, K., Gerl, M., et al. (2000). CRYOPRESERVED PRIMARY HEPATOCYTES AS A CONSTANTLY AVAILABLE IN VITRO MODEL FOR THE EVALUATION OF HUMAN AND ANIMAL DRUG METABOLISM AND ENZYME INDUCTION*. Drug Metab. Rev. *32*, 81–118.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet. TIG *20*, 320–326.

Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. Cell. Mol. Life Sci. CMLS *61*, 2571–2587.

Higgins, G.M., and Anderson, R.M. (1931). Restoration of the liver of the white rat following partial surgical removal. Arch Pathol *12*, 186–202.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226–232.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. *16*, 71–84.

Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J. Biol. Chem. *175*, 315–332.

Hovestadt, V., Jones, D.T.W., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., et al. (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature *510*, 537–541.

Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. Dev. Camb. Engl. *113*, 119–127.

Huang, C.-Z., Xu, J.-H., Zhong, W., Xia, Z.-S., Wang, S.-Y., Cheng, D., Li, J.-Y., Wu, T.-F., Chen, Q.-K., and Yu, T. (2017). Sox9 transcriptionally regulates Wnt signaling in intestinal epithelial stem cells in hypomethylated crypts in the diabetic state. Stem Cell Res. Ther. *8*.

Huang, M., Chang, A., Choi, M., Zhou, D., Anania, F.A., and Shin, C.H. (2014). Antagonistic interaction between Wnt and Notch activity modulates the regenerative capacity of a zebrafish fibrotic liver model. Hepatology *60*, 1753–1766.

Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. PLoS ONE *5*, e8888.

Ioshikhes, I.P., and Zhang, M.Q. (2000). Large-scale human promoter mapping using CpG islands. Nat. Genet. *26*, 61–63.

Ito, S. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Jakobsen, J.S., Waage, J., Rapin, N., Bisgaard, H.C., Larsen, F.S., and Porse, B.T. (2013). Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. Genome Res. *23*, 592–603.

Jeffrey T. Leek <Jtleek@Gmail.Com>, W. Evan Johnson<Wej@Bu.Edu>, Hilary S. Parker <Hiparker@Jhsph.Edu>, Elana J.Fertig <Ejfertig@Jhmi.Edu>, Andrew E. Jaffe <Ajaffe@Jhsph.Edu>, John D. Storey <Jstorey@Princeton.Edu>, Yuqing Zhang<Zhangyuqing.Pkusms@Gmail.Com>, Leonardo Collado Torres<Lcollado@Jhu.Edu> (2017). sva.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics *28*, 1805–1806.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jungermann, K., and Sasse, D. (1978). Heterogeneity of liver parenchymal cells. Trends Biochem. Sci. *3*, 198–202.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. Chembiochem Eur. J. Chem. Biol. *12*, 206–222.

Kanwal, F., Kramer, J.R., Mapakshi, S., Natarajan, Y., Chayanupatkul, M., Richardson, P.A., Li, L., Desiderio, R., Thrift, A.P., Asch, S.M., et al. (2018). Risk of Hepatocellular Cancer in Patients With Non-Alcoholic Fatty Liver Disease. Gastroenterology *155*, 1828-1837.e2.

Karantzali, E., Schulz, H., Hummel, O., Hubner, N., Hatzopoulos, A., and Kretsovali, A. (2008). Histone deacetylase inhibition accelerates the early events of stem cell differentiation: transcriptomic and epigenetic analysis. Genome Biol. *9*, R65.

Kato, S., Ishii, T., and Kouzmenko, A. (2015). Point mutations in an epigenetic factor lead to multiple types of bone tumors: role of H3.3 histone variant in bone development and disease. BoneKEy Rep. *4*, 715.

Kattler, K., Bachelor thesis (2013). Creation of histone modification reference maps in the HepaRG cell line using ChIP-sequencing.

Kattler, K., Master thesis (2015). Genome-wide chromatin accessibility studies in mammalian cells.

Kehl, T., Schneider, L., Kattler, K., Stöckel, D., Wegert, J., Gerstner, N., Ludwig, N., Distler, U., Schick, M., Keller, U., et al. (2018). REGGAE: a novel approach for the identification of key transcriptional regulators. Bioinforma. Oxf. Engl.

Kennedy, S., Rettinger, S., Flye, M.W., and Ponder, K.P. (1995). Experiments in transgenic mice show that hepatocytes are the source for postnatal liver growth and do not stream. Hepatol. Baltim. Md *22*, 160–168.

Kietzmann, T. (2017). Metabolic zonation of the liver: The oxygen gradient revisited. Redox Biol. *11*, 622–630.

Kim, C.H., and Younossi, Z.M. (2008). Nonalcoholic fatty liver disease: a manifestation of the metabolic syndrome. Cleve. Clin. J. Med. *75*, 721–728.

Kim, D., and Kim, W.R. (2017). Nonobese Fatty Liver Disease. Clin. Gastroenterol. Hepatol. *15*, 474–485.

Kim, J.K., Estève, P.-O., Jacobsen, S.E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. Nucleic Acids Res. *37*, 493–505.

Kinkley, S., Helmuth, J., Polansky, J.K., Dunkel, I., Gasparoni, G., Fröhler, S., Chen, W., Walter, J., Hamann, A., and Chung, H.-R. (2016). reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells. Nat. Commun. *7*, 12514.

Kleiner, D.E., Brunt, E.M., Van Natta, M., Behling, C., Contos, M.J., Cummings, O.W., Ferrell, L.D., Liu, Y.-C., Torbenson, M.S., Unalp-Arida, A., et al. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology *41*, 1313–1321.

Kobayashi, A., Valerius, M.T., Mugford, J.W., Carroll, T.J., Self, M., Oliver, G., and McMahon, A.P. (2008). Six2 Defines and Regulates a Multipotent Self-Renewing Nephron Progenitor Population throughout Mammalian Kidney Development. Cell Stem Cell *3*, 169–181.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. *17*, 691–707.

Köhler, C., Bell, A.W., Bowen, W.C., Monga, S.P., Fleig, W., and Michalopoulos, G.K. (2004). Expression of Notch-1 and its ligand Jagged-1 in rat liver during liver regeneration. Hepatol. Baltim. Md *39*, 1056–1065.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. *41*, 376–381.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. Mol. Cell *58*, 610–620.

Kondo, Y., Iwao, T., Yoshihashi, S., Mimori, K., Ogihara, R., Nagata, K., Kurose, K., Saito, M., Niwa, T., Suzuki, T., et al. (2014). Histone deacetylase inhibitor valproic acid promotes the differentiation of human induced pluripotent stem cells into hepatocyte-like cells. PloS One *9*, e104010.

Kordes, C., and Häussinger, D. (2013). Hepatic stem cell niches. J. Clin. Invest. *123*, 1874–1880.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865–868.

Koteish, A., and Diehl, A.M. (2001). Animal models of steatosis. Semin. Liver Dis. *21*, 89–104.

Kumar, R., and Mohan, S. (2017). Non-alcoholic Fatty Liver Disease in Lean Subjects: Characteristics and Implications. J. Clin. Transl. Hepatol. *XX*, 1–8.

Kumar, R., Rastogi, A., Sharma, M., Bhatia, V., Garg, H., Bihari, C., and Sarin, S. (2013). Clinicopathological characteristics and metabolic profiles of non-alcoholic fatty liver disease in Indian patients with normal body mass index: Do they differ from obese or overweight non-alcoholic fatty liver disease? Indian J. Endocrinol. Metab. *17*, 665.

Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., Kitago, M., Shinoda, M., Kitagawa, Y., and Sakamoto, M. (2018). Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification: KUREBAYASHI ET AL. Hepatology *68*, 1025–1041.

Lam, K., Pan, K., Linnekamp, J.F., Medema, J.P., and Kandimalla, R. (2016). DNA methylation based biomarkers in colorectal cancer: A systematic review. Biochim. Biophys. Acta BBA - Rev. Cancer *1866*, 106–120.

Lamarche, B.J., Orazio, N.I., and Weitzman, M.D. (2010). The MRN complex in double-strand break repair and telomere maintenance. FEBS Lett. *584*, 3682–3695.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. Genome Res. *20*, 320–331.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*, e1003118.

Lee, J.J., Murphy, G.F., and Lian, C.G. (2014). Melanoma epigenetics: novel mechanisms, markers, and medicines. Lab. Investig. J. Tech. Methods Pathol. *94*, 822–838.

Lee, S.-A., Kim, T.Y., Kwak, T.K., Kim, H., Kim, S., Lee, H.J., Kim, S.-H., Park, K.H., Kim, H.J., Cho, M., et al. (2010). Transmembrane 4 L six family member 5 (TM4SF5) enhances migration and invasion of hepatocytes for effective metastasis. J. Cell. Biochem. *111*, 59–66.

Lefebvre, P., Cariou, B., Lien, F., Kuipers, F., and Staels, B. (2009). Role of Bile Acids and Bile Acid Receptors in Metabolic Regulation. Physiol. Rev. *89*, 147–191.

Lemire, J.M., Shiojiri, N., and Fausto, N. (1991). Oval cell proliferation and the origin of small hepatocytes in liver injury induced by D-galactosamine. Am. J. Pathol. *139*, 535–552.

Leung, C.O.-N., Mak, W.-N., Kai, A.K.-L., Chan, K.-S., Lee, T.K.-W., Ng, I.O.-L., and Lo, R.C.-L. (2016). Sox9 confers stemness properties in hepatocellular carcinoma through Frizzled-7 mediated Wnt/β-catenin signaling. Oncotarget *7*.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harb. Perspect. Biol. *6*, a019133.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, M., and Fang, Y. (2015). Histone variants: the artists of eukaryotic chromatin. Sci. China Life Sci. *58*, 232–239.

Li, W., and Sun, Z. (2019). Mechanism of Action for HDAC Inhibitors—Insights from Omics Approaches. Int. J. Mol. Sci. *20*, 1616.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Li, S., Fang, X.-D., Wang, X.-Y., and Fei, B.-Y. (2018). Fos-like antigen 2 (FOSL2) promotes metastasis in colon cancer. Exp. Cell Res. *373*, 57–61.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. Science *341*, 1237905–1237905.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat. Commun. *4*, 1563.

Liu, Y., Chidgey, M., Yang, V.W., and Bialkowska, A.B. (2017). Krüppel-like factor 5 is essential for maintenance of barrier function in mouse colon. Am. J. Physiol. Gastrointest. Liver Physiol. *313*, G478–G491.

Loud, A.V. (1968). A quantitative stereological description of the ultrastructure of normal rat liver parenchymal cells. J. Cell Biol. *37*, 27–46.

Lüdtke, T.H.-W., Christoffels, V.M., Petry, M., and Kispert, A. (2009). Tbx3 promotes liver bud expansion during mouse development by suppression of cholangiocyte differentiation. Hepatology *49*, 969–978.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Lun, A.T.L., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. *44*, e45–e45.

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research *5*, 2122.

Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet. *32*, 225–237.

Macdonald, R.A. (1961). "Lifespan" of liver cells. Autoradio-graphic study using tritiated thymidine in normal, cirrhotic, and partially hepatectomized rats. Arch. Intern. Med. *107*, 335–343.

Machado, M., Marques-Vidal, P., and Cortez-Pinto, H. (2006). Hepatic histology in obese patients undergoing bariatric surgery. J. Hepatol. *45*, 600–606.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

MacParland, S.A., Liu, J.C., Ma, X.-Z., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., et al. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat. Commun. *9*.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods *9*, 1185–1188.

Marinov, G.K. (2017). On the design and prospects of direct RNA sequencing. Brief. Funct. Genomics *16*, 326–335.

Martinez-Arguelles, D.B., Lee, S., and Papadopoulos, V. (2014). In silico analysis identifies novel restriction enzyme combinations that expand reduced representation bisulfite sequencing CpG coverage. BMC Res. Notes *7*.

Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487–498.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics btw777.

McEnerney, L., Duncan, K., Bang, B.-R., Elmasry, S., Li, M., Miki, T., Ramakrishnan, S.K., Shah, Y.M., and Saito, T. (2017). Dual modulation of human hepatic zonation via canonical and non-canonical Wnt pathways. Exp. Mol. Med. *49*, e413.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv180203426 Cs Stat.

van Meer, P.J.K., Kooijman, M., Gispen-de Wied, C.C., Moors, E.H.M., and Schellekens, H. (2012). The ability of animal studies to detect serious post marketing adverse events is limited. Regul. Toxicol. Pharmacol. *64*, 345–349.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Meng, H., Cao, Y., Qin, J., Song, X., Zhang, Q., Shi, Y., and Cao, L. (2015). DNA Methylation, Its Mediators and Genome Integrity. Int. J. Biol. Sci. *11*, 604–617.

Michalopoulos, G.K., Barua, L., and Bowen, W.C. (2005). Transdifferentiation of rat hepatocytes into biliary cells after bile duct ligation and toxic biliary injury. Hepatology *41*, 535–544.

Moor, A.E., Harnik, Y., Ben-Moshe, S., Massasa, E.E., Rozenberg, M., Eilam, R., Bahar Halpern, K., and Itzkovitz, S. (2018). Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. Cell *175*, 1156-1167.e15.

Morey, C., Navarro, P., Debrand, E., Avner, P., Rougeulle, C., and Clerc, P. (2004). The region 3' to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. EMBO J. *23*, 594–604.

Naito, M., Hasegawa, G., and Takahashi, K. (1997). Development, differentiation, and maturation of Kupffer cells. Microsc. Res. Tech. *39*, 350–364.

Narayanan, S., Surette, F.A., and Hahn, Y.S. (2016). The Immune Landscape in Nonalcoholic Steatohepatitis. Immune Netw. *16*, 147.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77.

Ng, H.H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol. Cell. Biol. *20*, 1394–1406.

Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2018). Charting a tissue from single-cell transcriptomes. BioRxiv.

Nordström, K.J., Schmidt, F., Gasparoni, N., Salhab, A., Gasparoni, G., Kattler, K., Müller, F., Ebert, P., Costa, I.G., Pfeifer, N., et al. (2019). Unique and assay specific features of NOMe-, ATAC- and DNase I-seq data. BioRxiv.

Norton, L., Chen, X., Fourcaudot, M., Acharya, N.K., DeFronzo, R.A., and Heikkinen, S. (2014). The mechanisms of genome-wide target gene regulation by TCF7L2 in liver cells. Nucleic Acids Res. *42*, 13646–13661.

Nygaard, V., Rødland, E.A., and Hovig, E. (2015). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. Biostatistics kxv027.

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front. Endocrinol. *9*.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378–1381.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953–959.

Ohto, H., Takizawa, T., Saito, T., Kobayashi, M., Ikeda, K., and Kawakami, K. (1998). Tissue and developmental distribution of Six family gene products. Int. J. Dev. Biol. *42*, 141–148.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Ortica, S., Tarantino, N., Aulner, N., Israël, A., and Gupta-Rossi, N. (2014). The 4 Notch receptors play distinct and antagonistic roles in the proliferation and hepatocytic differentiation of liver progenitors. FASEB J. *28*, 603–614.

Osna, N.A., Donohue, T.M., and Kharbanda, K.K. (2017). Alcoholic Liver Disease: Pathogenesis and Current Management. Alcohol Res. Curr. Rev. *38*, 147–161.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr. Biol. CB *10*, 475–478.

Pais, R., Barritt, A.S., Calmus, Y., Scatton, O., Runge, T., Lebray, P., Poynard, T., Ratziu, V., and Conti, F. (2016). NAFLD and liver transplantation: Current burden and expected challenges. J. Hepatol. *65*, 1245–1257.

Palmer, C.N.A., Coates, P.J., Davies, S.E., Shephard, E.A., and Phillips, I.R. (1992). Localization of cytochrome P-450 gene expression in normal and diseased human liver byin situ hybridization of wax-embedded archival material. Hepatology *16*, 682–687.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. GigaScience *7*.

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I, Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D.P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. *19*, 221–233.

Peng, H., and Sun, R. (2017). Liver-resident NK cells and their potential functions. Cell. Mol. Immunol.

Petersen, B.E., Goff, J.P., Greenberger, J.S., and Michalopoulos, G.K. (1998). Hepatic oval cells express the hematopoietic stem cell marker Thy-1 in the rat. Hepatol. Baltim. Md *27*, 433–445.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Pollard, M.O., Gurdasani, D., Mentzer, A.J., Porter, T., and Sandhu, M.S. (2018). Long reads: their purpose and place. Hum. Mol. Genet. *27*, R234–R241.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat. Rev. Mol. Cell Biol. *16*, 245–257.

Poncy, A., Antoniou, A., Cordi, S., Pierreux, C.E., Jacquemin, P., and Lemaigre, F.P. (2015). Transcription factors SOX4 and SOX9 cooperatively control development of bile ducts. Dev. Biol. *404*, 136–148.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat. Biotechnol. *28*, 1057–1068.

Pott, S., and Lieb, J.D. (2015). Single-cell ATAC-seq: strength in numbers. Genome Biol. *16*.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C.A., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet. *48*, 681–686.

Preziosi, M., Okabe, H., Poddar, M., Singh, S., and Monga, S.P. (2018). Endothelial Wnts regulate β-catenin signaling in murine liver zonation and regeneration: A sequel to the Wnt-Wnt situation: PREZIOSI ET AL. Hepatol. Commun. *2*, 845–860.

Prior, N., Inacio, P., and Huch, M. (2019). Liver organoids: from basic research to therapeutic applications. Gut gutjnl-2019-319256.

Pu, W., Zhang, H., Huang, X., Tian, X., He, L., Wang, Y., Zhang, L., Liu, Q., Li, Y., Li, Y., et al. (2016). Mfsd2a+ hepatocytes repopulate the liver during injury and regeneration. Nat. Commun. *7*, 13369.

Qin, X., and Gao, B. (2006). The complement system in liver diseases. Cell. Mol. Immunol. *3*, 333–340.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Racine-Samson, L., Scoazec, J., D'Errico, A., Fiorentino, M., Christa, L., Moreau, A., Roda, C., Grigioni, W.F., and Feldmann, G. (1996). The metabolic organization of the adult human liver: A comparative study of normal, fibrotic, and cirrhotic liver tissue. Hepatology *24*, 104–113.

Raiber, E.-A., Hardisty, R., van Delft, P., and Balasubramanian, S. (2017). Mapping and elucidating the function of modified bases in DNA. Nat. Rev. Chem. *1*, 0069.

Ramadoss, P., Marcus, C., and Perdew, G.H. (2005). Role of the aryl hydrocarbon receptor in drug metabolism. Expert Opin. Drug Metab. Toxicol. *1*, 9–21.

Ramakrishna, G., Rastogi, A., Trehanpati, N., Sen, B., Khosla, R., and Sarin, S.K. (2013). From Cirrhosis to Hepatocellular Carcinoma: New Molecular Insights on Inflammation and Cellular Senescence. Liver Cancer *2*, 367–383.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rangwala, F., Guy, C.D., Lu, J., Suzuki, A., Burchette, J.L., Abdelmalek, M.F., Chen, W., and Diehl, A.M. (2011). Increased production of sonic hedgehog by ballooned hepatocytes. J. Pathol. *224*, 401–410.

Rappaport, A.M. (1977). Microcirculatory units in the mammalian liver. Their arterial and portal components. Bibl. Anat. 116–120.

Rashidi, H., Luu, N.-T., Alwahsh, S.M., Ginai, M., Alhaque, S., Dong, H., Tomaz, R.A., Vernay, B., Vigneswara, V., Hallett, J.M., et al. (2018). 3D human liver tissue from pluripotent stem cells displays stable phenotype in vitro and supports compromised liver function in vivo. Arch. Toxicol. *92*, 3117–3129.

Ratanasavanh, D., Beaune, P., Morel, F., Flinois, J.-P., Guengerich, F.P., and Guillouzo, A. (1991). Intralobular distribution and quantitation of cytochrome P-450 enzymes in human liver as a function of age. Hepatology *13*, 1142–1151.

Raven, A., Lu, W.-Y., Man, T.Y., Ferreira-Gonzalez, S., O'Duibhir, E., Dwyer, B.J., Thomson, J.P., Meehan, R.R., Bogorad, R., Koteliansky, V., et al. (2017). Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. Nature *547*, 350–354.

Razin, A., and Cedar, H. (1991). DNA methylation and gene expression. Microbiol. Rev. *55*, 451–458.

Reddington, J.P., Pennings, S., and Meehan, R.R. (2013). Non-canonical functions of the DNA methylome in gene regulation. Biochem. J. *451*, 13–23.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. *14*, 9–25.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. *26*, 139–140.

Rodriguez-Seguel, E., Mah, N., Naumann, H., Pongrac, I.M., Cerda-Esteban, N., Fontaine, J.-F., Wang, Y., Chen, W., Andrade-Navarro, M.A., and Spagnoli, F.M. (2013). Mutually exclusive signaling signatures define the hepatic and pancreatic progenitor cell lineage divergence. Genes Dev. *27*, 1932–1946.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858–5868.

Roskams, T., De Vos, R., Van Eyken, P., Myazaki, H., Van Damme, B., and Desmet, V. (1998). Hepatic OV-6 expression in human liver disease and rat experiments: evidence for hepatic progenitor cells in man. J. Hepatol. *29*, 455–463.

Rowell, R.J., and Anstee, Q.M. (2015). An overview of the genetics, mechanisms and management of NAFLD and ALD. Clin. Med. *15*, s77–s82.

Ruan, J., Ouyang, H., Amaya, M.F., Ravichandran, M., Loppnau, P., Min, J., and Zang, J. (2012). Structural basis of the chromodomain of Cbx3 bound to methylated peptides from histone h1 and G9a. PloS One *7*, e35376.

Russell, J.O., and Monga, S.P. (2018). Wnt/β-Catenin Signaling in Liver Development, Homeostasis, and Pathobiology. Annu. Rev. Pathol. Mech. Dis. *13*, 351–378.

Ryaboshapkina, M., and Hammar, M. (2017). Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. Sci. Rep. *7*.

Saito, K., Negishi, M., and James Squires, E. (2013). Sexual dimorphisms in zonal gene expression in mouse liver. Biochem. Biophys. Res. Commun. *436*, 730–735.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*.

Salhab, A., Nordström, K., Gasparoni, G., Kattler, K., Ebert, P., Ramirez, F., Arrigoni, L., Müller, F., Polansky, J.K., Cadenas, C., et al. (2018). A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. Genome Biol. *19*, 150.

Sano, K., Fujioka, Y., Nagashima, K., Hashizume, T., Komori, M., Kitada, M., Kamataki, T., and Miyazaki, T. (1989). Distributional variation of P-450 immunoreactive hepatocytes in human liver disorders. Hum. Pathol. *20*, 1015–1020.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Sassa, S., Sugita, O., Galbraith, R.A., and Kappas, A. (1987). Drug metabolism by the human hepatoma cell, Hep G2. Biochem. Biophys. Res. Commun. *143*, 52–57.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Sato, K., Kennedy, L., Liangpunsakul, S., Kusumanchi, P., Yang, Z., Meng, F., Glaser, S., Francis, H., and Alpini, G. (2019). Intercellular Communication between Hepatic Cells in Liver Diseases. Int. J. Mol. Sci. *20*, 2180.

Schaub, J.R., Malato, Y., Gormond, C., and Willenbring, H. (2014). Evidence against a Stem Cell Origin of New Hepatocytes in a Common Mouse Model of Chronic Liver Injury. Cell Rep. *8*, 933–939.

Schillebeeckx, M., Schrade, A., Löbs, A.-K., Pihlajoki, M., Wilson, D.B., and Mitra, R.D. (2013). Laser capture microdissection–reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced adrenocortical neoplasia in the mouse. Nucleic Acids Res. *41*, e116–e116.

von Schönfels, W., Beckmann, J.H., Ahrens, M., Hendricks, A., Röcken, C., Szymczak, S., Hampe, J., and Schafmayer, C. (2018). Histologic improvement of NAFLD in patients with obesity after bariatric surgery based on standardized NAS (NAFLD activity score). Surg. Obes. Relat. Dis. *14*, 1607–1616.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Schuppan, D., and Afdhal, N.H. (2008). Liver cirrhosis. The Lancet *371*, 838–851.

Schwen, L.O., Homeyer, A., Schwier, M., Dahmen, U., Dirsch, O., Schenk, A., Kuepfer, L., Preusser, T., and Schenk, A. (2016). Zonated quantification of steatosis in an entire mouse liver. Comput. Biol. Med. *73*, 108–118.

Seitz, H.K., Bataller, R., Cortez-Pinto, H., Gao, B., Gual, A., Lackner, C., Mathurin, P., Mueller, S., Szabo, G., and Tsukamoto, H. (2018). Alcoholic liver disease. Nat. Rev. Dis. Primer *4*.

Self, M., Lagutin, O.V., Bowling, B., Hendrix, J., Cai, Y., Dressler, G.R., and Oliver, G. (2006). Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. EMBO J. *25*, 5214–5228.

Self, M., Geng, X., and Oliver, G. (2009). Six2 activity is required for the formation of the mammalian pyloric sphincter. Dev. Biol. *334*, 409–417.

Sell, S., Osborn, K., and Leffert, H.L. (1981). Autoradiography of "oval cells" appearing rapidly in the livers of rats fed N-2-fluorenylacetamide in a choline devoid diet. Carcinogenesis *2*, 7–14.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shayevitch, R., Askayo, D., Keydar, I., and Ast, G. (2018). The importance of DNA methylation of exons on alternative splicing. RNA *24*, 1351–1362.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

Skvortsova, K., Zotenko, E., Luu, P.-L., Gould, C.M., Nair, S.S., Clark, S.J., and Stirzaker, C. (2017). Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. Epigenetics Chromatin *10*.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820.

Snykers, S., Henkens, T., De Rop, E., Vinken, M., Fraczek, J., De Kock, J., De Prins, E., Geerts, A., Rogiers, V., and Vanhaecke, T. (2009). Role of epigenetics in liver-specific gene transcription, hepatocyte differentiation and stem cell reprogrammation. J. Hepatol. *51*, 187–211.

Sokal, E.M., Trivedi, P., Cheeseman, P., Portmann, B., and Mowat, A.P. (1989). The application of quantitative cytochemistry to study the acinar distribution of enzymatic activities in human liver biopsy sections. J. Hepatol. *9*, 42–48.

Sokal, E.M., Collette, E., and Buts, J.P. (1991). Persistence of a Liver Metabolic Zonation in Extra-Hepatic Biliary Atresia Cirrhotic Livers. Pediatr. Res. *30*, 286–289.

Sokal, E.M., Collette, E., and Buts, J.P. (1993). Continuous increase of alcohol dehydrogenase activity along the liver plate in normal and cirrhotic human livers. Hepatology *17*, 202–205.

Sooknanan, R., Pease, J., and Doyle, K. (2010). Novel methods for rRNA removal and directional, ligation-free RNA-seq library preparation. Nat. Methods *7*, i–ii.

Spruijt, C.G., and Vermeulen, M. (2014). DNA methylation: old dog, new tricks? Nat. Struct. Mol. Biol. *21*, 949–954.

Stacchiotti, A., Favero, G., Lavazza, A., Golic, I., Aleksic, M., Korac, A., Rodella, L.F., and Rezzani, R. (2016). Hepatic Macrosteatosis Is Partially Converted to Microsteatosis by Melatonin Supplementation in ob/ob Mice Non-Alcoholic Fatty Liver Disease. PLOS ONE *11*, e0148115.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. MMBR *64*, 435–459.

Stöckel, D., Kehl, T., Trampert, P., Schneider, L., Backes, C., Ludwig, N., Gerasch, A., Kaufmann, M., Gessler, M., Graf, N., et al. (2016). Multi-omics enrichment analysis using the GeneTrail2 web service. Bioinformatics *32*, 1502–1508.

Strålfors, A., and Ekwall, K. (2011). Heterochromatin and Euchromatin-Organization, Boundaries, and Gene Regulation. In Encyclopedia of Molecular Cell Biology and Molecular Medicine, R.A. Meyers, ed. (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), p.

Stringer, E.J., Duluc, I., Saandi, T., Davidson, I., Bialecka, M., Sato, T., Barker, N., Clevers, H., Pritchard, C.A., Winton, D.J., et al. (2012). Cdx2 determines the fate of postnatal intestinal endoderm. Dev. Camb. Engl. *139*, 465–474.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226.

Stunnenberg, H.G., Hirst, M., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., Amit, I., Antonarakis, S.E., Aparicio, S., et al. (2016). The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell *167*, 1145–1149.

Suzuki, T., and Ono, T. (1987). Immunohistochemical studies on the distribution and frequency of fatty-acid-binding protein positive cells in human fetal, newborn and adult liver tissues. J. Pathol. *153*, 385–394.

Suzuki, A., Sekiya, S., Buscher, D., Izpisua Belmonte, J.C., and Taniguchi, H. (2008). Tbx3 controls the fate of hepatic progenitor cells in liver development by suppressing p19ARF expression. Development *135*, 1589–1595.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Taberlay, P.C., Achinger-Kawecka, J., Lun, A.T.L., Buske, F.A., Sabir, K., Gould, C.M., Zotenko, E., Bert, S.A., Giles, K.A., Bauer, D.C., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. *26*, 719–731.

Tabibian, J.H., Masyuk, A.I., Masyuk, T.V., O'Hara, S.P., and LaRusso, N.F. (2013). Physiology of Cholangiocytes. In Comprehensive Physiology, R. Terjung, ed. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Takiguchi, M. (1998). The C/EBP family of transcription factors in the liver and other organs. Int. J. Exp. Pathol. *79*, 369–391.

Talal, A.H., Venuto, C.S., and Younis, I. (2017). Assessment of Hepatic Impairment and Implications for Pharmacokinetics of Substance Use Treatment: Clinical Pharmacology in Drug Development. Clin. Pharmacol. Drug Dev. *6*, 206–212.

Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. Nat. Rev. Mol. Cell Biol. *18*, 115–126.

Tan, X., Yuan, Y., Zeng, G., Apte, U., Thompson, M.D., Cieply, B., Stolz, D.B., Michalopoulos, G.K., Kaestner, K.H., and Monga, S.P.S. (2008). β-Catenin deletion in hepatoblasts disrupts hepatic morphogenesis and survival during mouse development. Hepatology *47*, 1667–1679.

Tarlow, B.D., Finegold, M.J., and Grompe, M. (2014). Clonal tracing of Sox9 $^+$ liver progenitors in mouse oval cell injury. Hepatology *60*, 278–289.

Tarrats, N., Moles, A., Morales, A., García-Ruiz, C., Fernández-Checa, J.C., and Marí, M. (2011). Critical role of tumor necrosis factor receptor 1, but not 2, in hepatic stellate cell proliferation, extracellular matrix remodeling, and liver fibrogenesis. Hepatology *54*, 319–327.

Tavian, M., and Péault, B. (2005). Embryonic development of the human hematopoietic system. Int. J. Dev. Biol. *49*, 243–250.

Teissandier, A., and Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. EMBO J. *36*, 1471–1473.

Teufel, A., Itzel, T., Erhart, W., Brosch, M., Wang, X.Y., Kim, Y.O., von Schönfels, W., Herrmann, A., Brückner, S., Stickel, F., et al. (2016). Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. Gastroenterology *151*, 513-525.e0.

Thoma, F., and Koller, T. (1977). Influence of histone H1 on chromatin structure. Cell *12*, 101–107.

Tie, F., Banerjee, R., Saiakhova, A.R., Howard, B., Monteith, K.E., Scacheri, P.C., Cosgrove, M.S., and Harte, P.J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Dev. Camb. Engl. *141*, 1129–1139.

Tien, E.S., and Negishi, M. (2006). Nuclear receptors CAR and PXR in the regulation of hepatic metabolism. Xenobiotica *36*, 1152–1163.

Tierling, S., Schmitt, B., and Walter, J. (2018). Comprehensive Evaluation of Commercial Bisulfite-Based DNA Methylation Kits and Development of an Alternative Protocol With Improved Conversion Performance. Genet. Epigenetics *10*, 1179237X1876609.

Tijsterman, M., Ketting, R.F., and Plasterk, R.H.A. (2002). The Genetics of RNA Silencing. Annu. Rev. Genet. *36*, 489–519.

Topol, L., Chen, W., Song, H., Day, T.F., and Yang, Y. (2009). Sox9 Inhibits Wnt Signaling by Promoting β-Catenin Phosphorylation in the Nucleus. J. Biol. Chem. *284*, 3323–3333.

Trapnell, C. (2017). monocle.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. Nat. Rev. Genet. *19*, 535–548.

Valderrama-Trevino, A.I., Barrera-Mera, B., Ceballos-Villalva, J.C., and Montalvo-Jave, E.E. (2017). Hepatic Metastasis from Colorectal Cancer. Euroasian J. Hepato-Gastroenterol. *7*, 166–175.

VanHook, A.M. (2012). Dishevelled Mediates Wnt-Notch Crosstalk. Sci. Signal. *5*, ec298–ec298.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. *16*, 178–189.

Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc. Natl. Acad. Sci. *106*, 21453–21458.

Waddington, C.H. (1968). Towards a theoretical biology. Nature *218*, 525–527.

Waddington, C.H. (2012). The epigenotype. 1942. Int. J. Epidemiol. *41*, 10–13.

Wang, B., Zhao, L., Fish, M., Logan, C.Y., and Nusse, R. (2015). Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver. Nature *524*, 180–185.

Wang, L., Ozark, P.A., Smith, E.R., Zhao, Z., Marshall, S.A., Rendleman, E.J., Piunti, A., Ryan, C., Whelan, A.L., Helmin, K.A., et al. (2018). TET2 coactivates gene expression through demethylation of enhancers. Sci. Adv. *4*, eaau6986.

Wang, Y., Alhaque, S., Cameron, K., Meseguer-Ripolles, J., Lucendo-Villarin, B., Rashidi, H., and Hay, D.C. (2017). Defined and Scalable Generation of Hepatocyte-like Cells from Human Pluripotent Stem Cells. J. Vis. Exp.

Weijun Luo (2017). pathview.

Weiskirchen, R., and Tacke, F. (2014). Cellular and molecular functions of hepatic stellate cells in inflammatory responses and liver immunology. Hepatobiliary Surg. Nutr. *3*, 344–363.

Weitzman, S.A., Turk, P.W., Milkowski, D.H., and Kozlowski, K. (1994). Free radical adducts induce alterations in DNA cytosine methylation. Proc. Natl. Acad. Sci. *91*, 1261–1264.

WHO (2018). Controlling the global obesity epidemic.

Wickham, H. (2009). Ggplot2: elegant graphics for data analysis (New York: Springer).

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*.

Wossidlo, M. (2011). 5-hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun *2*, 241.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature *532*, 329–333.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. *36*, 3573–3599.

Xie, N., Zhou, Y., Sun, Q., and Tang, B. (2018). Novel Epigenetic Techniques Provided by the CRISPR/Cas9 System. Stem Cells Int. *2018*, 1–12.

Ye, D., Li, T., Heraud, P., and Parnpai, R. (2016). Effect of Chromatin-Remodeling Agents in Hepatic Differentiation of Rat Bone Marrow-Derived Mesenchymal Stem Cells *In Vitro* and *In Vivo*. Stem Cells Int. *2016*, 1–11.

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., Muthurajan, U.M., Nie, X., Kawashima, T., Groth, M., et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell *158*, 98–109.

Yeo, S., Jeong, S., Kim, J., Han, J.-S., Han, Y.-M., and Kang, Y.-K. (2007). Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells. Biochem. Biophys. Res. Commun. *359*, 536–542.

Yi, F., Brubaker, P.L., and Jin, T. (2005). TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. J. Biol. Chem. *280*, 1457–1464.

Yin, C. (2017). Molecular mechanisms of Sox transcription factors during the development of liver, bile duct, and pancreas. Semin. Cell Dev. Biol. *63*, 68–78.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. TIG *13*, 335–340.

Yoon, S.-M., Gerasimidou, D., Kuwahara, R., Hytiroglou, P., Yoo, J.E., Park, Y.N., and Theise, N.D. (2011). Epithelial cell adhesion molecule (EpCAM) marks hepatocytes newly derived from stem/progenitor cells in humans. Hepatology *53*, 964–973.

Young, M.D., Willson, T.A., Wakefield, M.J., Trounson, E., Hilton, D.J., Blewitt, M.E., Oshlack, A., and Majewski, I.J. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic Acids Res. *39*, 7415–7427.

Younossi, Z.M., Loomba, R., Anstee, Q.M., Rinella, M.E., Bugianesi, E., Marchesini, G., Neuschwander-Tetri, B.A., Serfaty, L., Negro, F., Caldwell, S.H., et al. (2018). Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. Hepatology *68*, 349–360.

Zhang, Y., Qiao, H.-X., Zhou, Y.-T., Hong, L., and Chen, J.-H. (2018). Fibrinogen-like-protein 1 promotes the invasion and metastasis of gastric cancer and is associated with poor prognosis. Mol. Med. Rep. *18*, 1465–1472.

Zhao, Y.-J., Ju, Q., and Li, G.-C. (2013). Tumor markers for hepatocellular carcinoma. Mol. Clin. Oncol. *1*, 593–598.

Zhulidov, P.A. (2004). Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. *32*, 37e–337.

Zong, Y., Panikkar, A., Xu, J., Antoniou, A., Raynaud, P., Lemaigre, F., and Stanger, B.Z. (2009). Notch signaling controls liver development by regulating biliary differentiation. Development *136*, 1727–1739.

6. Supplement

6.1 Supplementary Figures

A

Celltype	log(CPM+1)	
	V0.1	V0.2
Cholangiocytes (KRT19)	1.09	1.98
Cholangiocytes (KRT7)	2.67	4.11
Fibroblasts (THY1)	1.85	2.93
Fibroblasts (S100A4 (FSP1))	3.77	3.88
Hepatic stellate cells (CD38)	1.98	1.77
Hepatocytes (ALB)	16.81	16.56
Hepatocytes (APOA1)	14.11	14.37
Kupffer cells (CLEC4F)	0.45	0.40
Kupffer cells / macrophages (CD68)	0.02	5.50
Mast cells (FCER2)	0.20	0.17
Mast cells (IL2RA)	0.25	0.34
mature dendritic cells (CD83)	2.40	2.19
NK cells (NCR1)	0.44	0.31
Sinusoidal endothel cells (AOC3)	3.15	2.91
Sinusoidal endothel cells (STAB1)	4.21	4.48
Sinusoidal endothel cells (CLEC4M)	5.19	4.98
T cells (CD3G)	1.24	1.05

Supplementary Figure 1: Marker transcript expression levels for hepatocytes ad non-parenchymal cell-types in experiment V0.1 and V0.2. A – RNA abundance is displayed as log (CPM + 1). B – Kupffer cell and macrophage marker CD69 expression levels as CPM by phenotypic group. C – Heatmap of average marker gene expression normalized for sequencing depth as log (CPM + 1) by individual phenotypes.

A

expression		Variance [%]	Zone	Phenotype	Donor	Sex	BMI	Age	Diabetes
PC1	correlation	90.27	0.473	0.706	0.110	0.178	0.082	0.097	0.004
	p value		0.000	0.051	0.413	0.184	0.547	0.475	0.976
PC2	correlation	1.51	0.861	0.272	0.149	0.279	0.145	0.114	0.277
	p value		0.000	0.041	0.268	0.035	0.283	0.400	0.037
PC3	correlation	0.94	0.017	0.213	0.089	0.025	0.464	0.034	0.227
	p value		0.901	0.112	0.509	0.854	0.000	0.801	0.090

methylation		Variance [%]	Zone	Phenotype	Donor	Sex	BMI	Age	Diabetes
PC1	correlation	88.23	0.437	0.025	0.275	0.287	0.120	0.081	0.070
	p value		0.001	0.852	0.038	0.030	0.373	0.550	0.604
PC2	correlation	0.88	0.271	0.360	0.159	0.580	0.011	0.332	0.559
	p value		0.042	0.006	0.236	0.000	0.934	0.012	0.000
PC3	correlation	0.50	0.263	0.204	0.377	0.384	0.272	0.111	0.445
	p value		0.048	0.129	0.004	0.003	0.041	0.410	0.001

B

expression		Variance [%]	Zone	Phenotype	Donor	Sex	BMI	Age	HOMA	Batch
PC1	correlation	89.43	0.112	0.439	0.249	0.061	0.400	0.214	0.349	0.161
	p value		0.189	0.000	0.003	0.482	0.000	0.012	0.000	0.058
PC2	correlation	2.58	0.148	0.682	0.327	0.194	0.238	0.393	0.221	0.306
	p value		0.082	0.000	0.000	0.024	0.006	0.000	0.018	0.000
PC3	correlation	0.79	0.048	0.118	0.266	0.236	0.100	0.007	0.153	0.264
	p value		0.577	0.168	0.002	0.006	0.257	0.936	0.105	0.002

methylation		Variance [%]	Zone	Phenotype	Donor	Sex	BMI	Age	HOMA	Batch	CpG Coverage
PC1	correlation	98.20	0.113	0.319	0.277	0.060	0.024	0.159	0.263	0.277	0.053
	p value		0.247	0.001	0.004	0.540	0.810	0.105	0.010	0.004	0.588
PC2	correlation	0.08	0.167	0.651	0.305	0.136	0.187	0.403	0.065	0.349	0.241
	p value		0.087	0.000	0.001	0.166	0.064	0.000	0.533	0.000	0.013
PC3	correlation	0.07	0.172	0.126	0.010	0.060	0.142	0.003	0.405	0.112	0.115
	p value		0.078	0.197	0.918	0.541	0.161	0.976	0.000	0.253	0.242

Supplementary Figure 2: Pearson correlation with corresponding p-values of principle components obtained from PCA of zonal expression and DNA methylation data with zonation, phenotype, donor, sex, BMI, Age, and diabetes diagnosis. A – V0.1. B – V0.2, including correlations with batch and CpG Coverage to identify possible batch effects between V0.1 and V0.2. Increasing correlation values are indicated in a white to red to dark red color ramp. Significant p values (< 0.05) are highlighted in yellow.
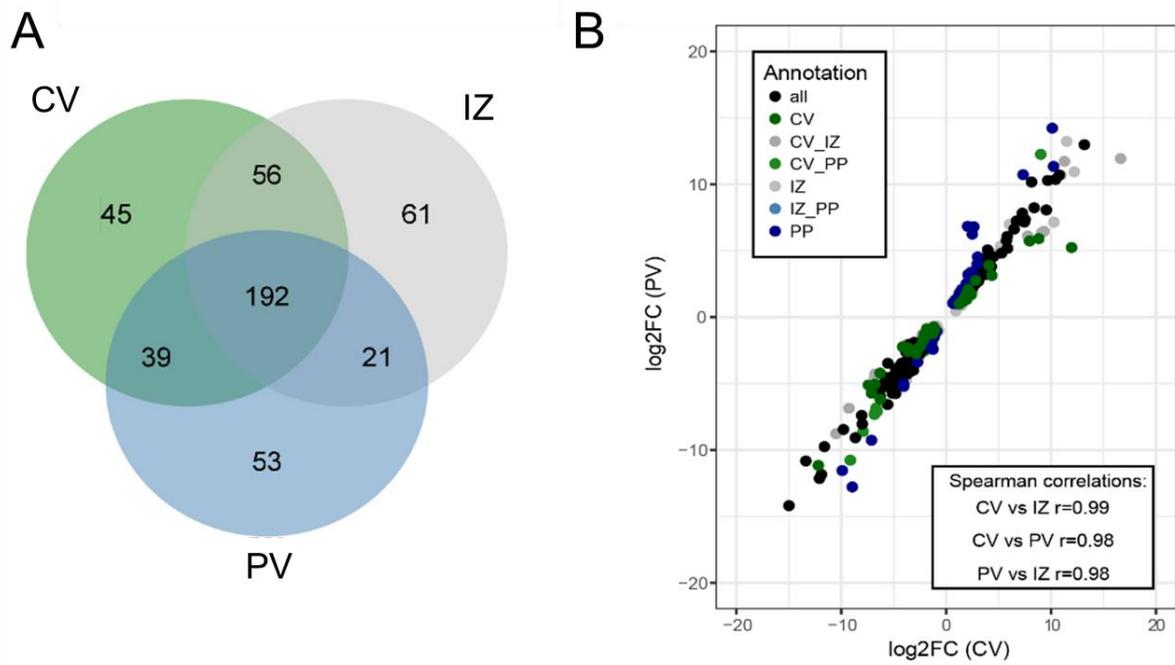
Supplementary Figure 3: Differential gene expression in steatosis (Brosch et al., 2018). A – Venn diagram depicting the overlap of DEGs between steatotic samples (STEA and EARLY, n = 10) and samples with normal liver histology (NC and HO, n = 9) for individual zones (CV, IZ and PV) determined using edgeR with |log2FC| > 1 and FDR < 0.01. B – Scatterplot of CV versus PV log2 fold changes of genes significantly deregulated in steatosis in any of the three zones with corresponding Spearman correlations between zones.

NC		Variance [%]	Zone	Donor
PC1	correlation	89.9	0.325	0.145
	p value		0.175	0.554
PC2	correlation	2.05	0.101	0.782
	p value		0.680	0.000
PC3	correlation	1.86	0.592	0.172
	p value		0.008	0.483

HO		Variance [%]	Zone	Donor
PC1	correlation	89.45	0.252	0.060
	p value		0.246	0.787
PC2	correlation	2.19	0.884	0.251
	p value		0.000	0.247
PC3	correlation	1.37	0.415	0.130
	p value		0.179	0.555

STEA		Variance [%]	Zone	Donor
PC1	correlation	88.53	0.202	0.338
	p value		0.379	0.134
PC2	correlation	2.17	0.799	0.456
	p value		0.000	0.038
PC3	correlation	1.52	0.164	0.440
	p value		0.478	0.046

EARLY		Variance [%]	Zone	Donor
PC1	correlation	88.26	0.077	0.252
	p value		0.740	0.271
PC2	correlation	2.35	0.828	0.444
	p value		0.000	0.044
PC3	correlation	1.5	0.164	0.507
	p value		0.478	0.019

NASH		Variance [%]	Zone	Donor
PC1	correlation	89.89	0.026	0.542
	p value		0.918	0.020
PC2	correlation	1.99	0.131	0.630
	p value		0.604	0.005
PC3	correlation	1.59	0.721	0.135
	p value		0.001	0.594

FIB3		Variance [%]	Zone	Donor
PC1	correlation	88.52	0.000	0.022
	p value		1.000	0.947
PC2	correlation	2.62	0.030	0.777
	p value		0.927	0.003
PC3	correlation	2.08	0.118	0.583
	p value		0.714	0.047

ACI		Variance [%]	Zone	Donor
PC1	correlation	89.67	0.079	0.461
	p value		0.722	0.027
PC2	correlation	5.2	0.031	0.147
	p value		0.887	0.504
PC3	correlation	1.26	0.086	0.674
	p value		0.695	0.000

Supplementary Figure 4: Pearson correlation with corresponding p-values of group-wise principal components derived from PCA of the 1000 most variable transcripts from LCM-RNA-seq data (V0.2) with hepatic zonation (Zone) and donor. High correlation values are colored in red, significant p values (< 0.05) are highlighted in yellow.
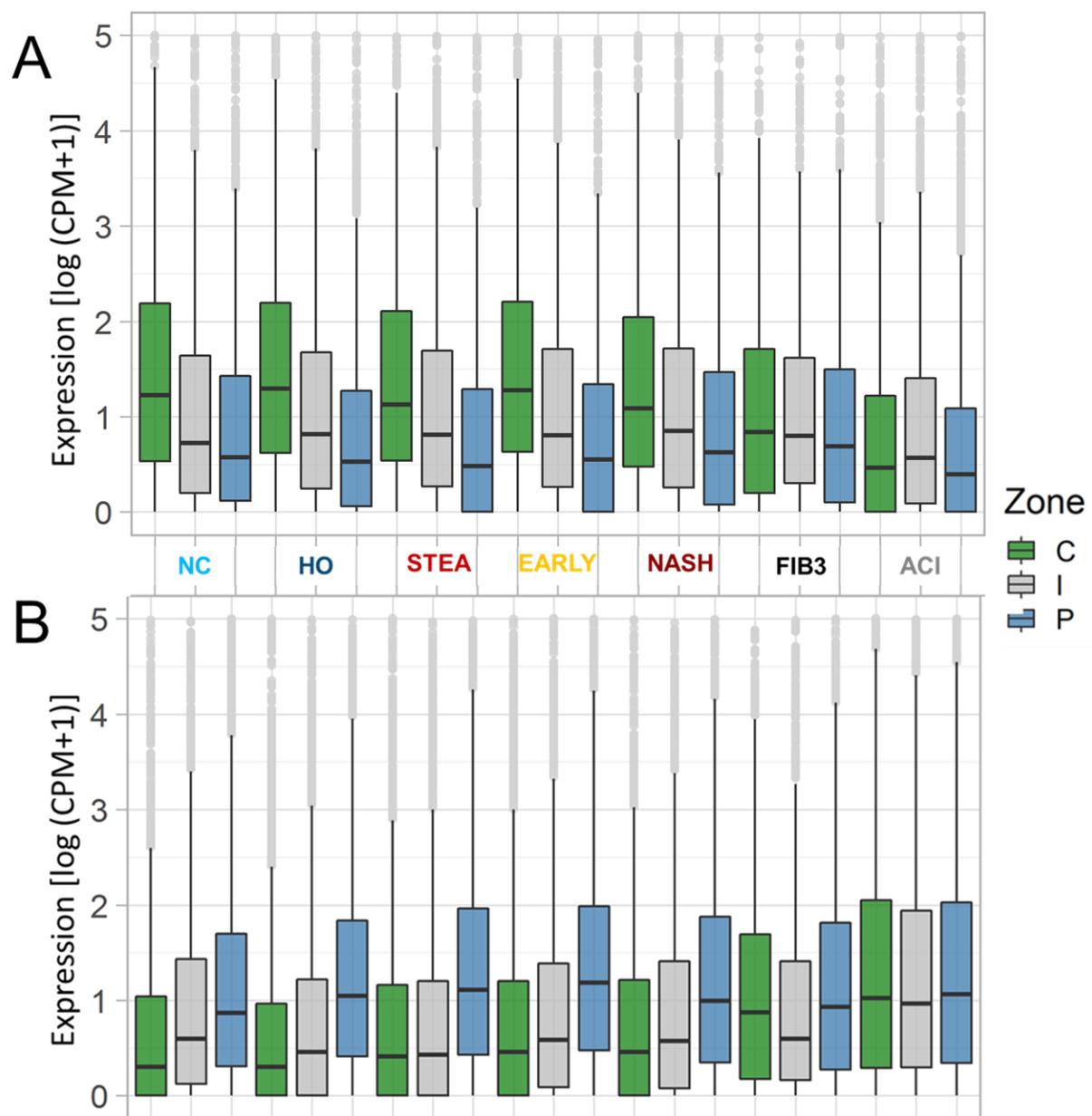
Supplementary Figure 5: Zone-specific expression levels of all pericentrally (A) and periportally (B) zonated genes during fatty liver disease progression from normal controls (NC) to cirrhosis (ACI). Expression levels are displayed as log (CPM + 1). The analysis is based on the zonated gene list described in chapter 3.1.1 (805 genes).
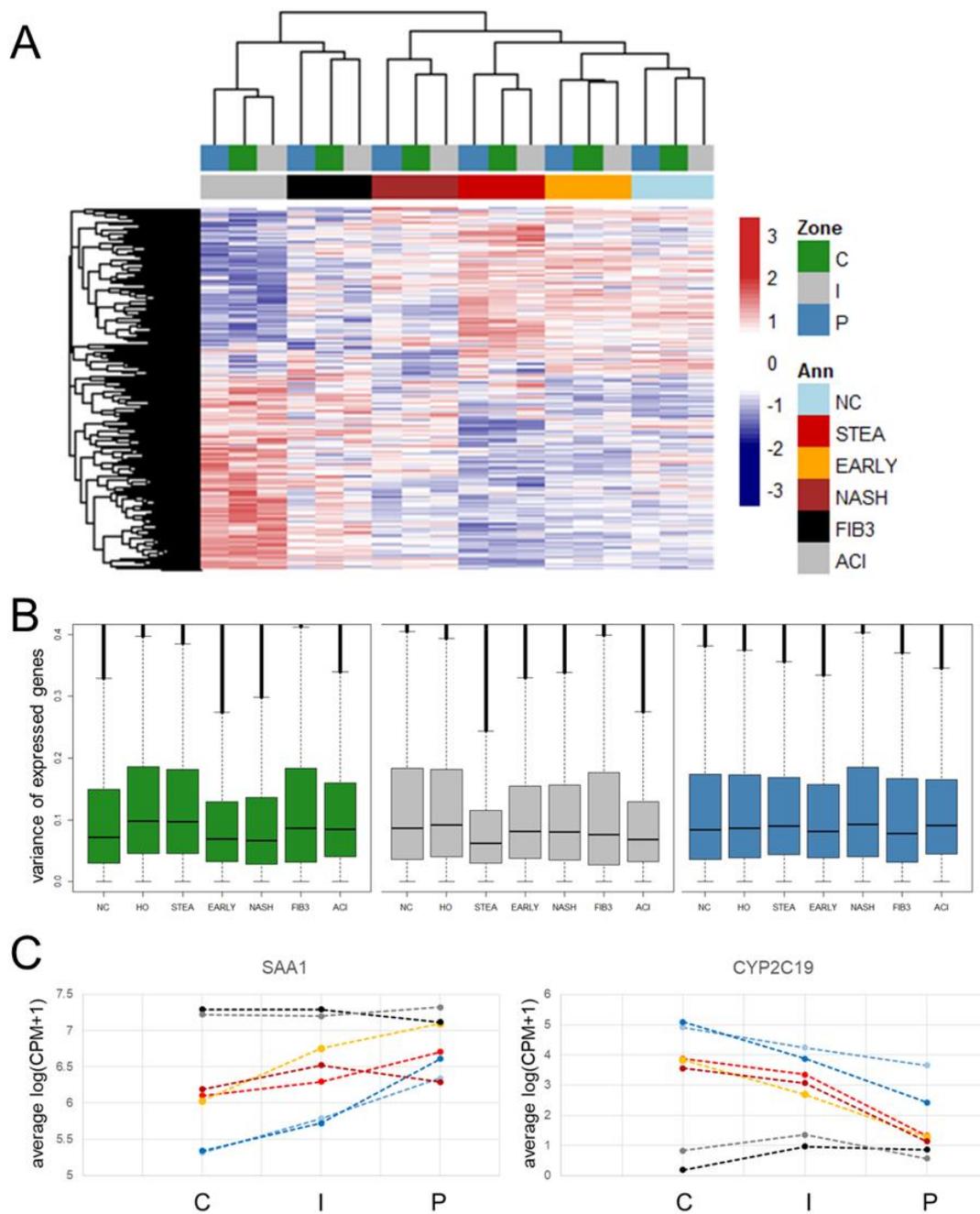
Supplementary Figure 6: Differential gene expression during fatty liver disease progression. The annotation legend displays controls (NC) in light blue, healthy obese (HO) in dark blue, non-alcoholic steatosis (STEA) in red, early NASH (EARLY) in orange, progressed NASH (NASH) in dark red, fibrosis (FIB3) in black and alcoholic cirrhosis (ACI) in grey. Pericentral samples are annotated in green, intermediate in grey and periportal in blue. A – Canberra clustering of scaled $\log_2$ fold changes (logFC) of all DEGs detected in any comparison described in Figure 30 (n = 2941, grey). Red indicates downregulation in disease, white corresponds to no expression change compared to HO (|logFC| < 1), and blue to upregulation in disease. B – Boxplots of inter-individual variance of expressed genes within groups after filtering of unexpressed genes. C – Average zonal expression levels of exemplary DEGs for individual phenotypic groups, which are annotated as described in A.
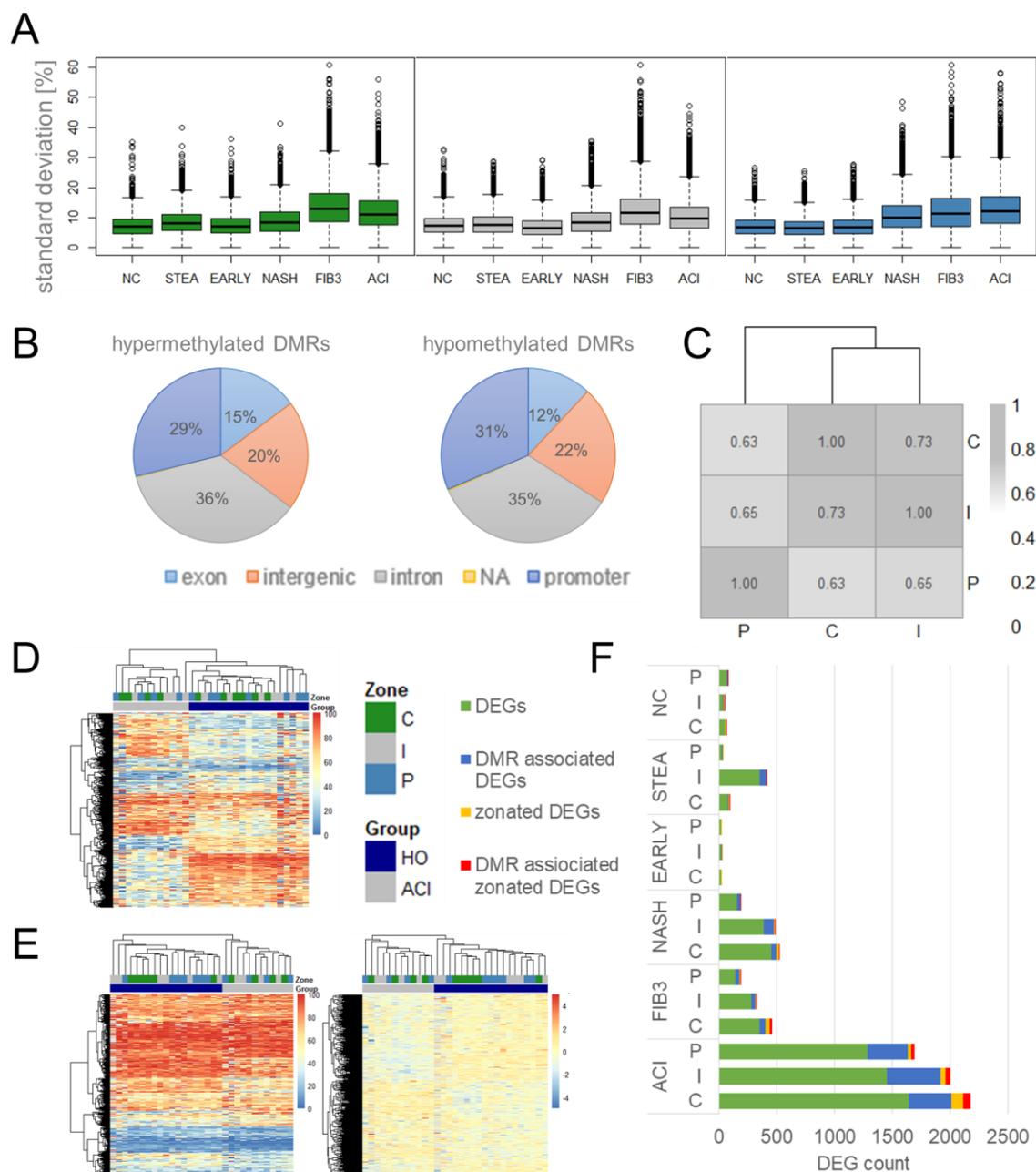
Supplementary Figure 7: DNA methylation in progressed fatty liver disease stages. A – Boxplot of absolute standard deviation of zone-specific DMRs detected between phenotypes and healthy obese controls (HO) as described in Figure 31. B – Exemplary genomic annotation of DMRs detected in the pericentral zone between alcoholic cirrhosis and HO for hypomethylated and hypermethylated regions. Proportions of genomic annotations of DMRs detected in other comparisons (Figure 31) were distributed similarly. C – Pearson correlations between zone-specific methylation differences between ACI and HO. D – Methylation level heatmap of the 1000 most variable DMRs between ACI and HO in the pericentral zone. E – Methylation level heatmap (left panel) and corresponding z-score illustration (right panel) of the 1239 pericentral DMRs between ACI and HO associated with genes that are zonally expressed in healthy liver tissue. F – Association between zonation in healthy liver, DEGs and DMRs of the comparisons described in Figures 30 and 31. DEGs not associated with zonation or DMRs are colored in green, DMR associated DEGs in blue, DEGs zonated in normal liver in yellow and DMR associated with DEGs zonated in normal liver in red.

WNT signaling pathway
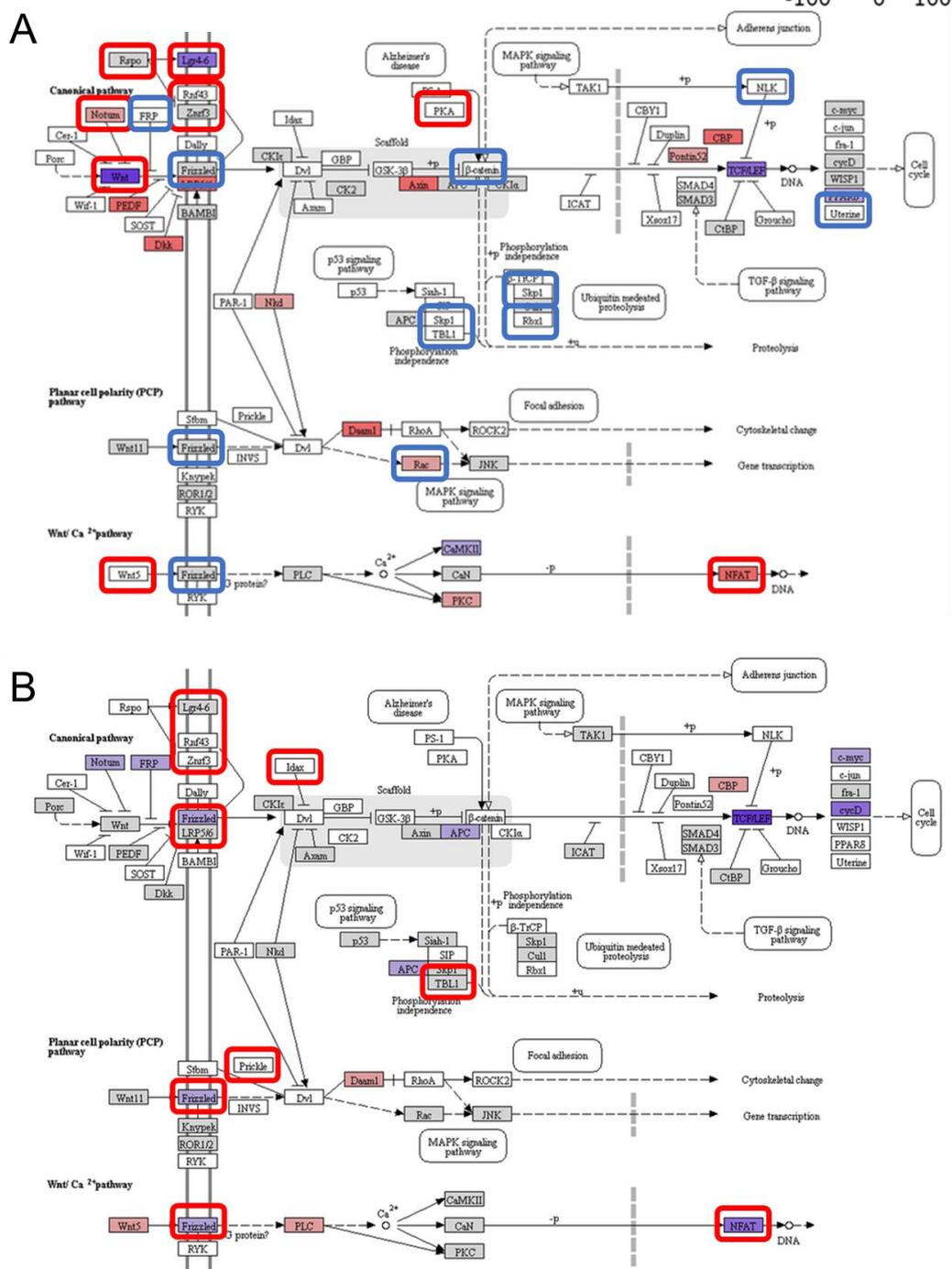
Supplementary Figure 8: Transcriptional and epigenetic deregulation of the Wnt signaling pathway in pericentral (A) and periportal (B) hepatocytes in alcoholic cirrhosis. Fill color illustrates differential DNA methylation between alcoholic cirrhosis and healthy obese controls, with red corresponding to hypermethylation and blue to hypomethylation, respectively. Accordingly, genes framed in red are transcriptionally downregulated in alcoholic cirrhosis, while blue edges illustrate upregulated gene expression.
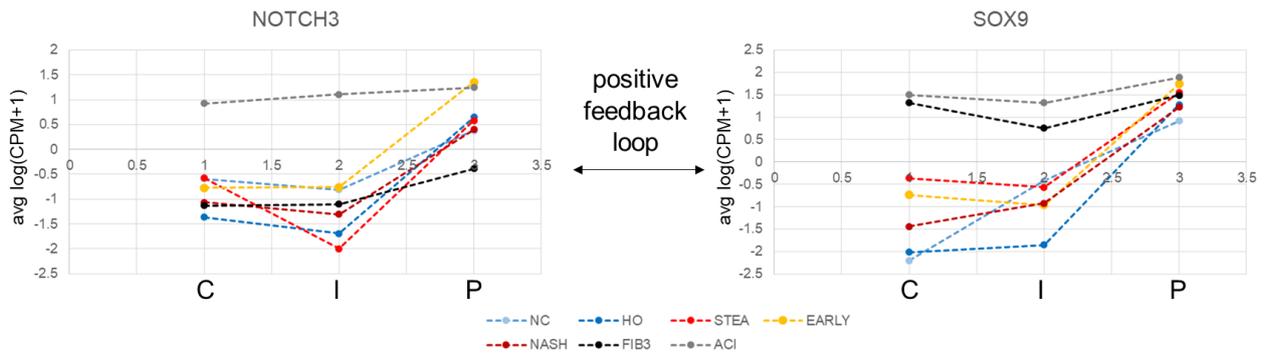
Supplementary Figure 9: Zone-specific NOTCH3 and SOX9 expression levels as average log (CPM + 1) of individual phenotypes during progression of fatty liver disease. The annotation legend displays controls (NC) in light blue, healthy obese (HO) in dark blue, non-alcoholic steatosis (STEA) in red, early NASH (EARLY) in orange, progressed NASH (NASH) in dark red, fibrosis (FIB3) in black and alcoholic cirrhosis (ACI) in grey.
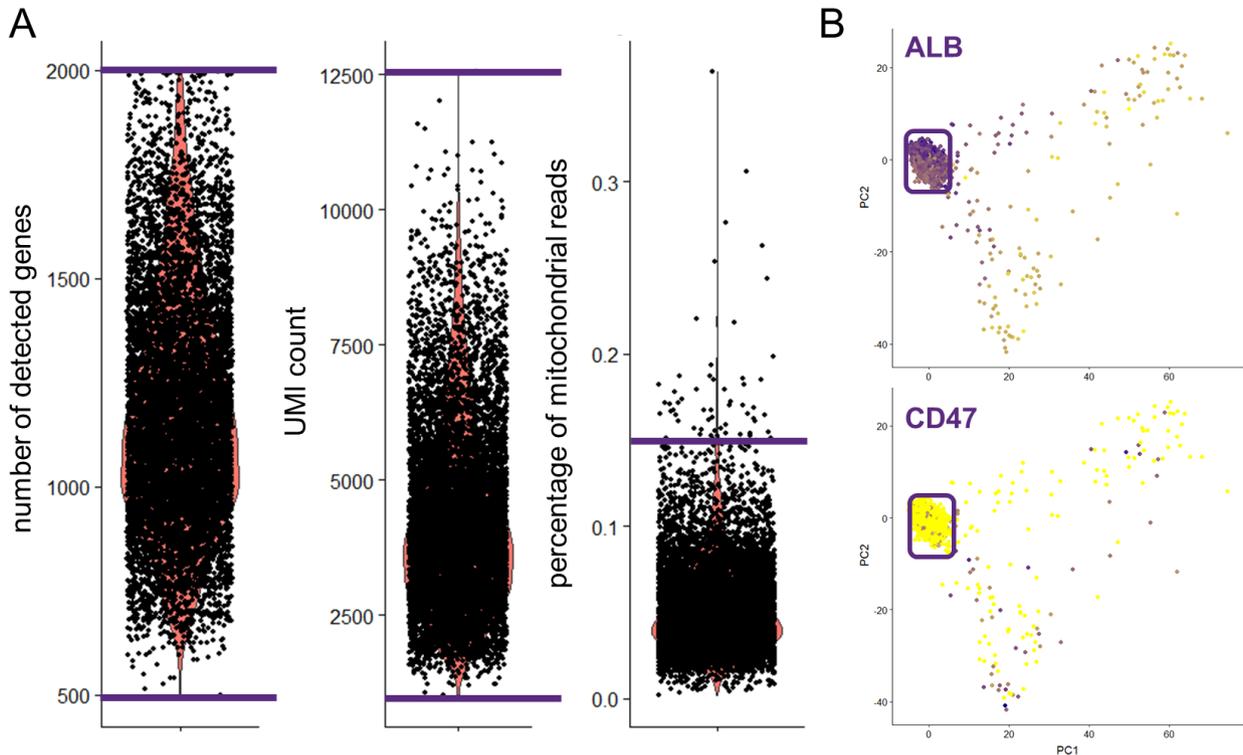
Supplementary Figure 10: Quality metrics of a single cell dataset generated from human hepatocytes using the 10X Genomics Chromium platform. A – Number of detected genes, UMI count and percentage of mitochondrial reads for 8411 cells already pre-filtered for low respective high gene and UMI counts. Purple lines indicate chosen thresholds for filtering of low quality cells. B – PCA on HVGs detected in 8181 single cells after initial filtering of low quality cells. The dataset was further filtered for PCA outliers, retaining only cells encircled purple. Additionally cells with log2 normalized UMI count < 4 for Albumin (yellow in upper panel) and > 4 for CD47 (purple in lower panel) were excluded to obtain only high confidence hepatocytes without any non-parenchymal cell contamination.
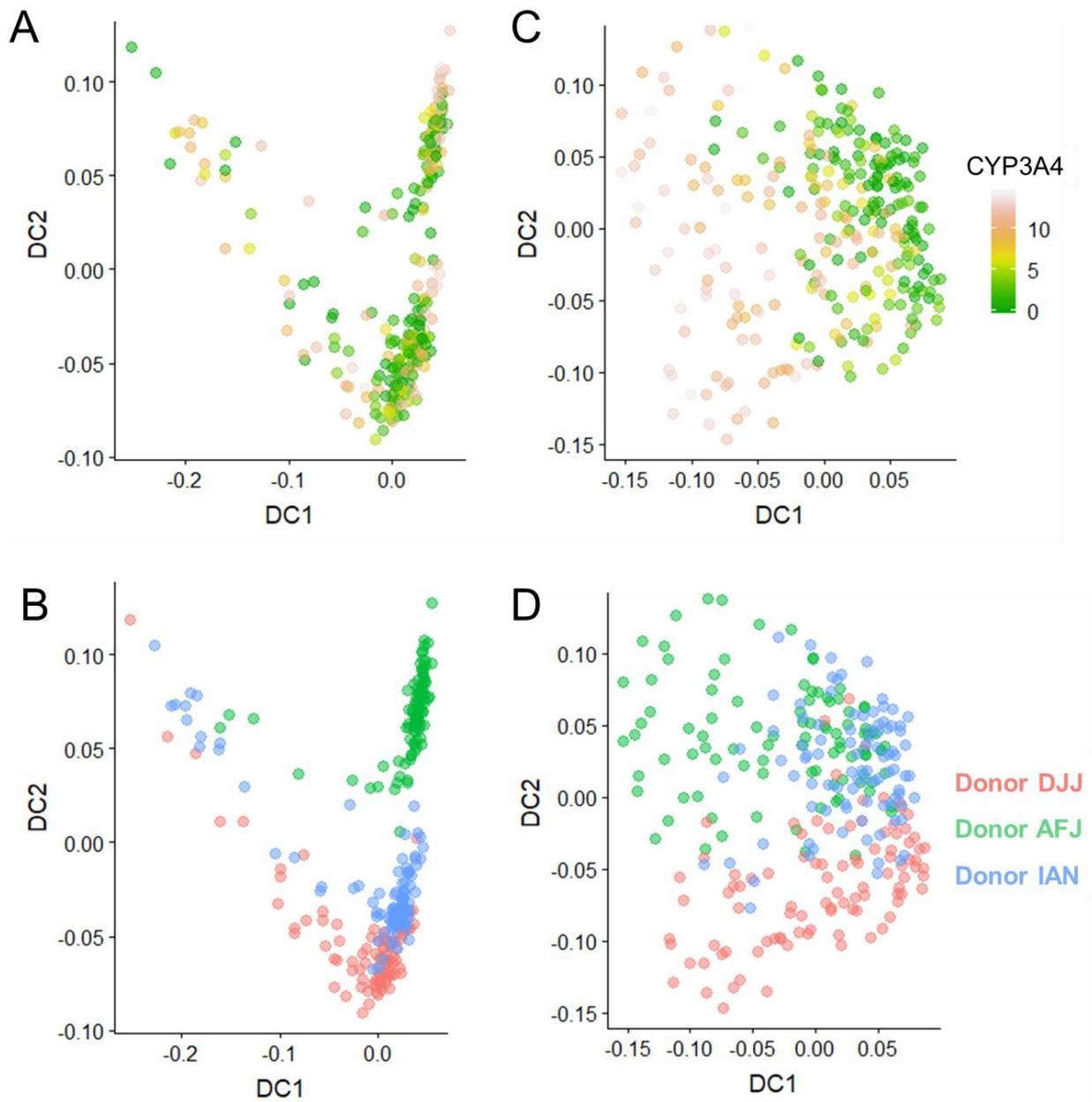
Supplementary Figure 11: Diffusion map of a small, deeply sequenced scRNA-seq dataset comprising 3 donors of PHH (286 cells, in average 1.8 million reads per cell) based on all HVGs (A and B) or 38 HVGs with zonated expression profiles in healthy human liver (C and D). Plots are colored by normalized CYP3A4 expression (pericentral marker, A and C) or by Donor ID (B and D).
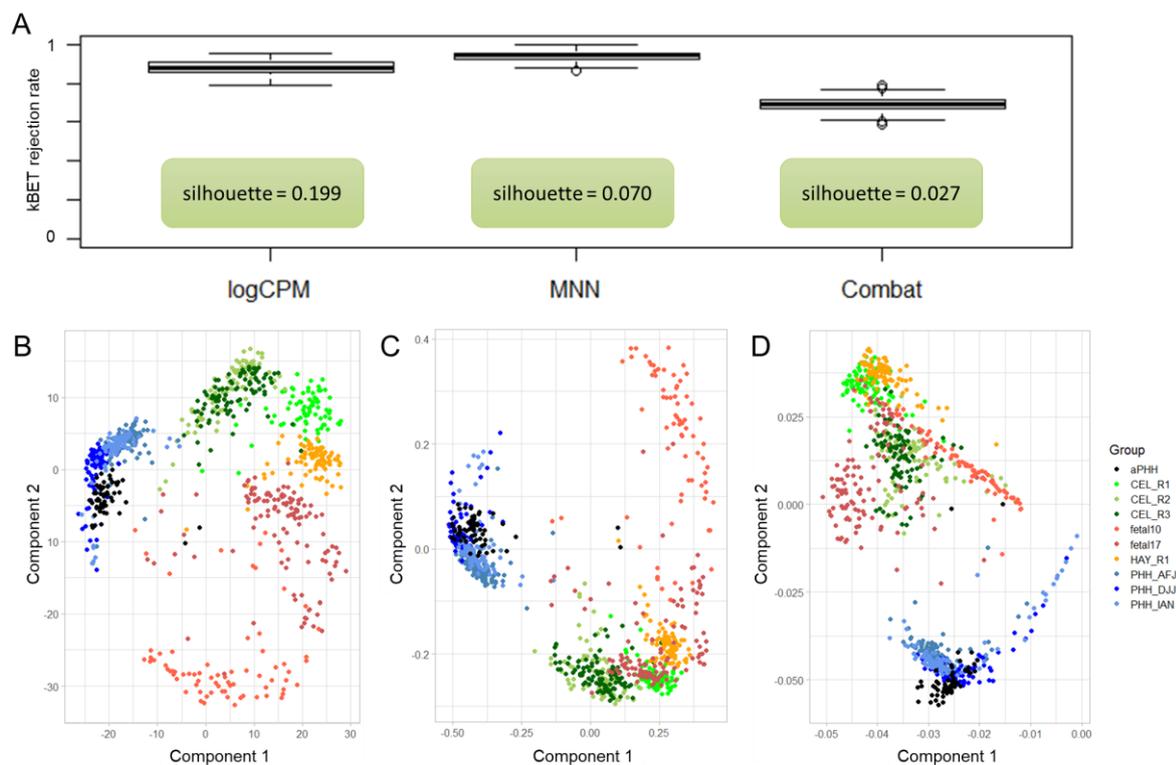
Supplementary Figure 12: Batch effect assessment and correction of the StemNet scRNA-seq dataset including external data from Camp et al. (2017). A – kBET rejection rate of the logCPM normalized dataset, and after MNN and Combat correction. Corresponding PCA silhouette scores are highlighted in green. Details are described in chapter 2.2.2.2. B – PCA of the 1000 most variable transcripts of the dataset normalized as log counts per million (logCPM, same panel as Figure 47 D more convenient comparison with batch corrected datasets). Cells are colored by sample annotation (black – external adult PHH (Camp et al., 2017), external fetal PHH – shades of red, CEL HLCs – shades of green, HAY HLCs – orange, PHH – shades of blue). C –Dimensionality reduction plot after MNN correction. D – Dimensionality reduction plot after Combat correction.
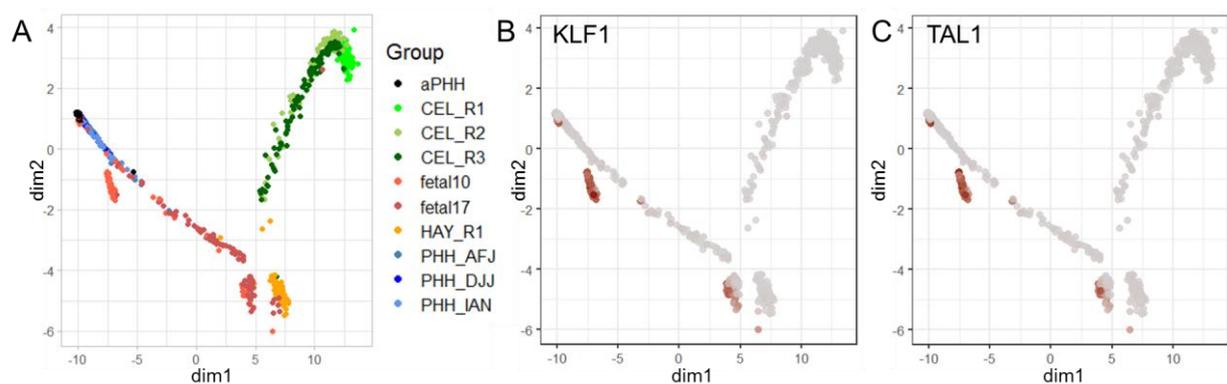
Supplementary Figure 13: Fetal hepatocyte subgroup (week 10.5) detected by pseudotime analysis (Figure 49) expresses markers of fetal liver erythropoiesis. A – Figure 49 A. B – Reduced representation map colored by KLF1 and TAL1 expression. Red corresponds to high expression levels.
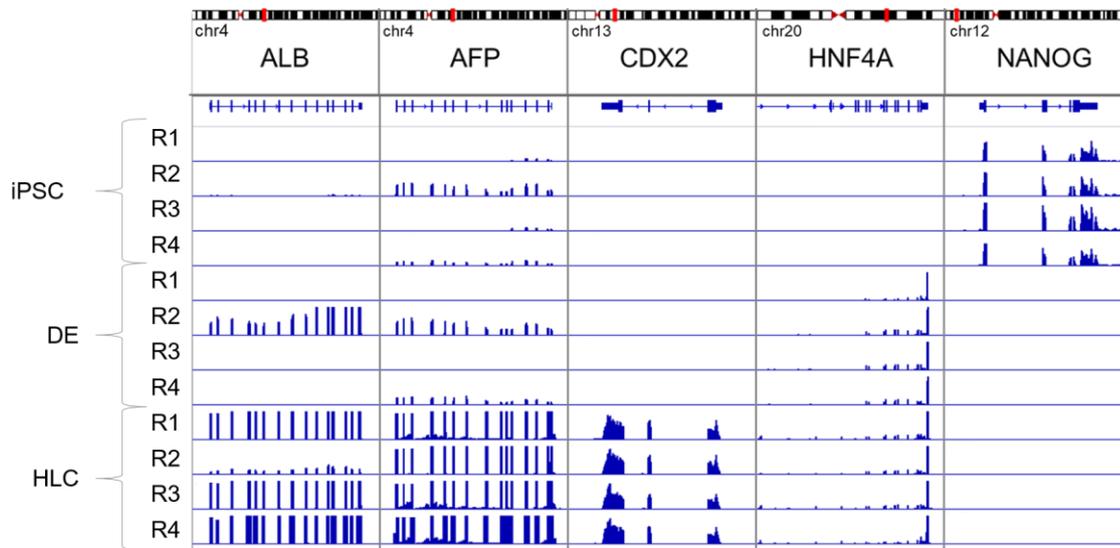
Supplementary Figure 14: Exemplary genome browser view of expression levels of the differentiation marker genes ALB, AFP, CDX2, HNF4A and NANOG. The figure displays the CEL differentiation process from iPSCs to DE to HLC in 4 replicates. Coverage tracks were normalized for sequencing depth.
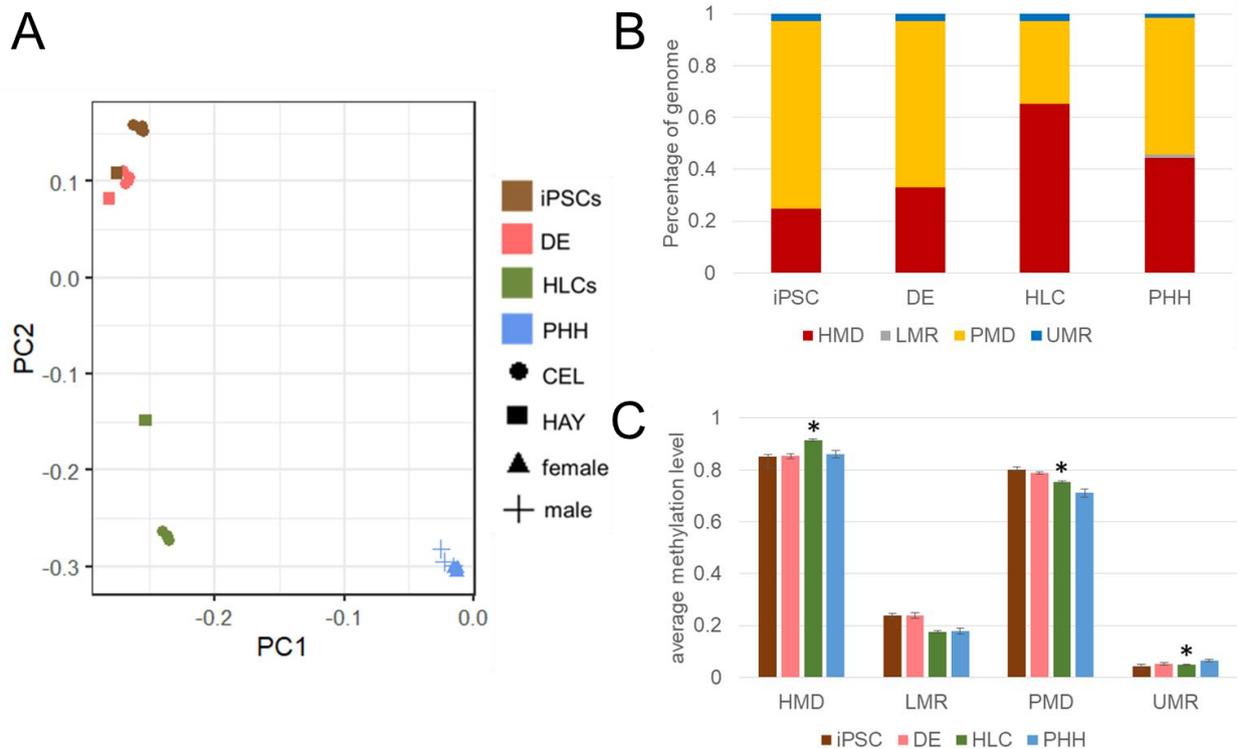
Supplementary Figure 15: Segmentation of DNA methylation patterns during HLC differentiation. A – PCA of DNA methylation levels of the 5,000 most variable CpGs. iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue. The annotation legend displays corresponding symbols for CEL (n = 4), HAY (n = 1) as well as female (n = 3) and male PHH (n = 3). B – Percentage of the genome segmented into highly methylated domains (HMDs, red), lowly methylated domains (LMDs, grey), partially methylated domains (PMDs, yellow) and unmethylated regions (UMRs, blue) in iPSCs, DE, HLCs and PHH. C – Average methylation levels in HMDs, LMRs, PMDs and UMRs. Standard deviations are depicted as error bars. Asterisks mark significant differences between HLCs and the other cell-types (ttest, p < 0.05).
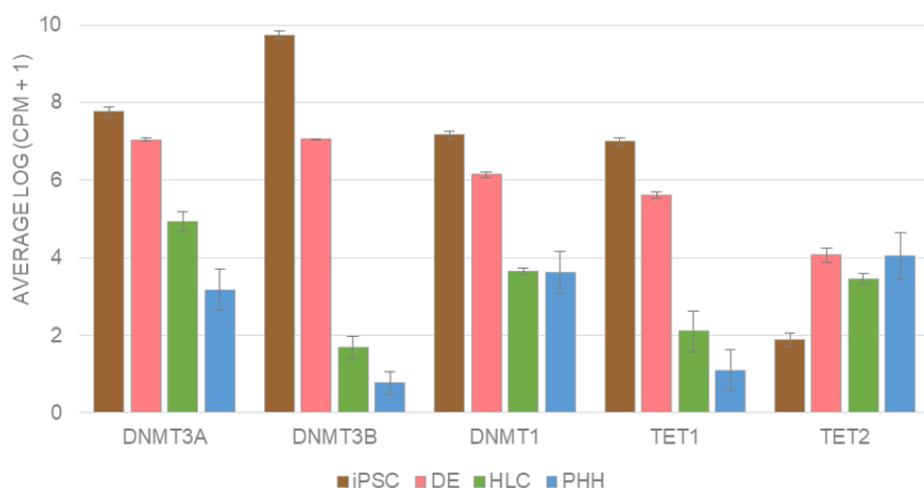
Supplementary Figure 16: Expression levels of DNA methyltransferases and Tet Methyl-cytosine Dioxygenases. Expression is displayed as average log (CPM + 1). Error bars illustrate standard deviations. iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue.
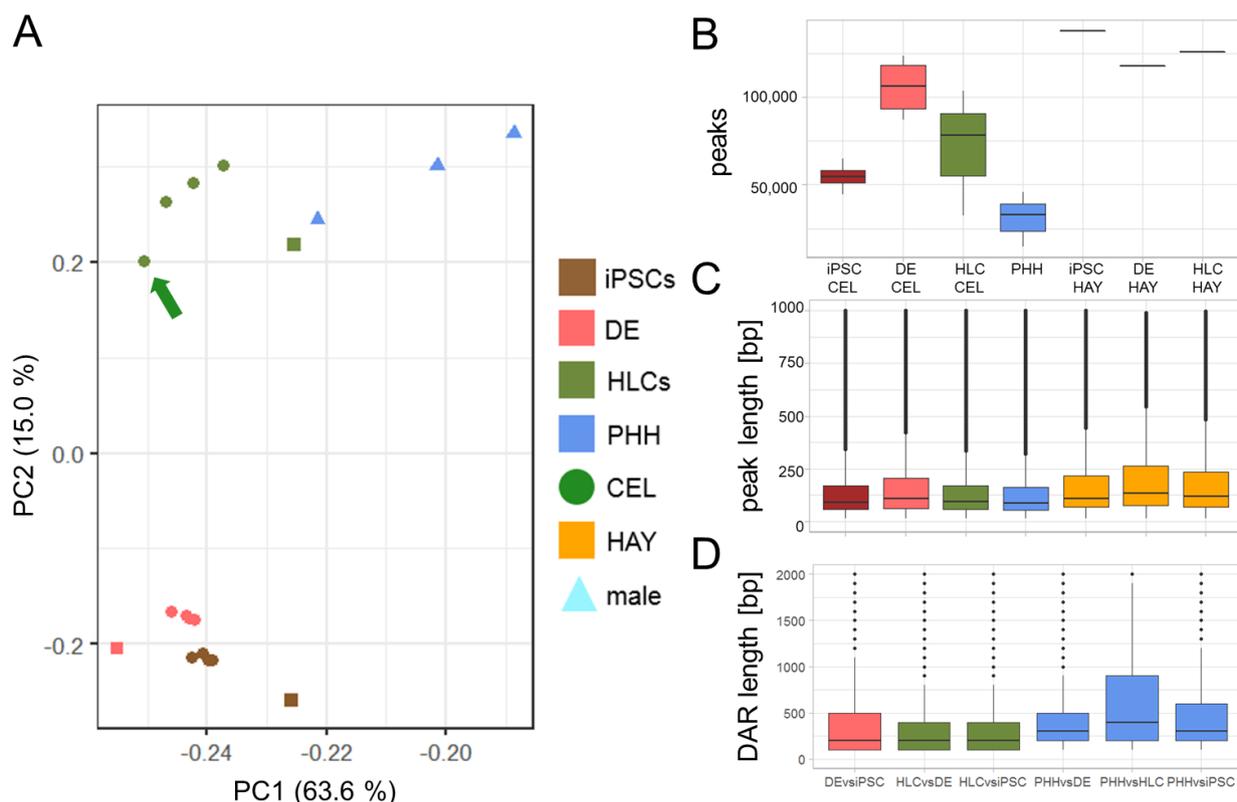
Supplementary Figure 17: Quality assessment of chromatin accessibility analysis by ATAC-seq. A – Principal component analysis of ATAC-seq coverage in the 50,000 most variable 100 bp windows, based on log (CPM + 1) counts calculated after removal of blacklist regions. iPSCs are colored in brown, DE in red, HLCs in olive and PHH in blue. The annotation legend displays corresponding symbols for CEL (n = 4), HAY (n = 1) and PHH (n = 3). B – Boxplot of MACS2 peak counts per sample group. C – Boxplot of peak length per sample group. D – Length of differentially accessible regions (DARs) detected in the comparisons further described in Figure 44.
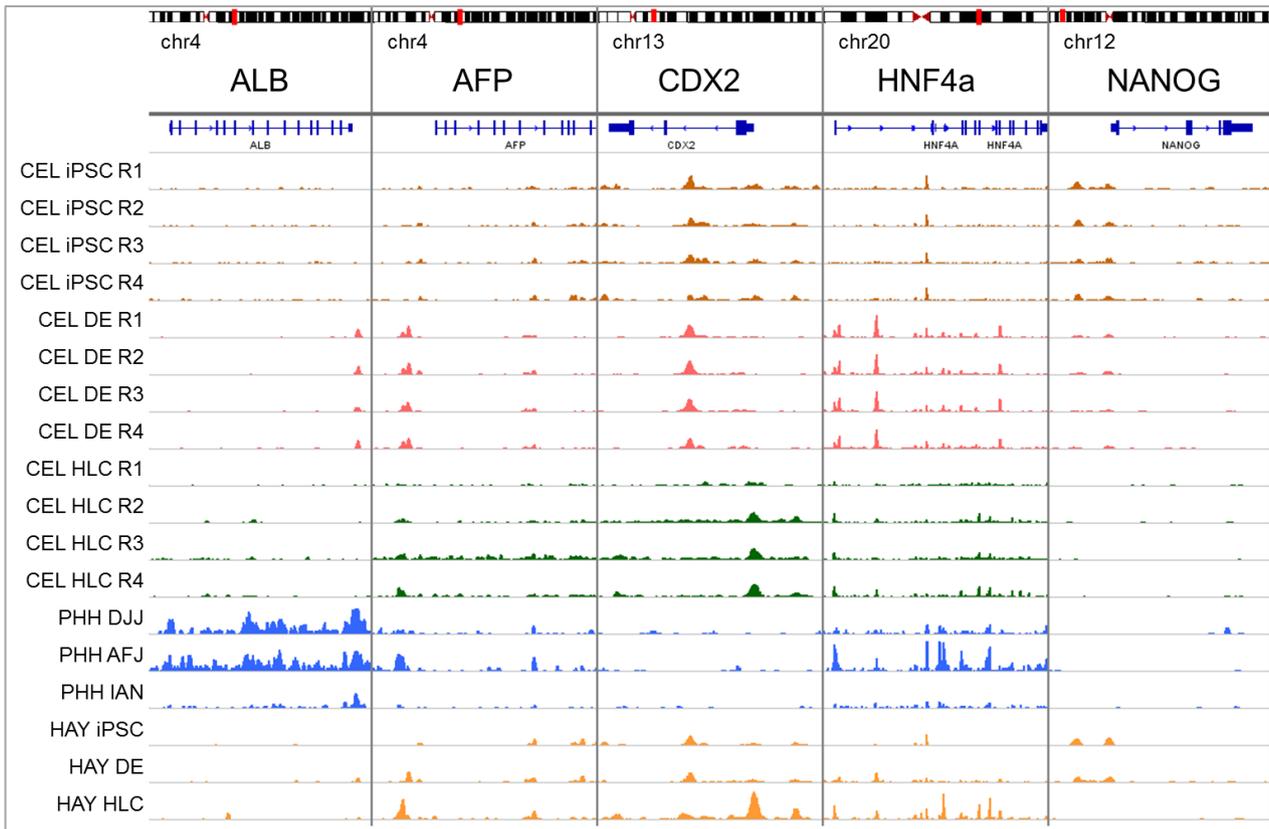
Supplementary Figure 18: Exemplary genome browser representation of accessible chromatin detected by ATAC-seq coverage at the marker genes ALB, AFP, CDX2, HNF4a and NANOG that exhibit strong expression differences during *in vitro* HLC differentiation and in comparison of HLCs with PHH. CEL iPSC coverage tracks are colored in brown, CEL DE in red, CEL HLCs in olive, PHH in blue and HAY samples in orange. Tracks were normalized by sequencing depth.
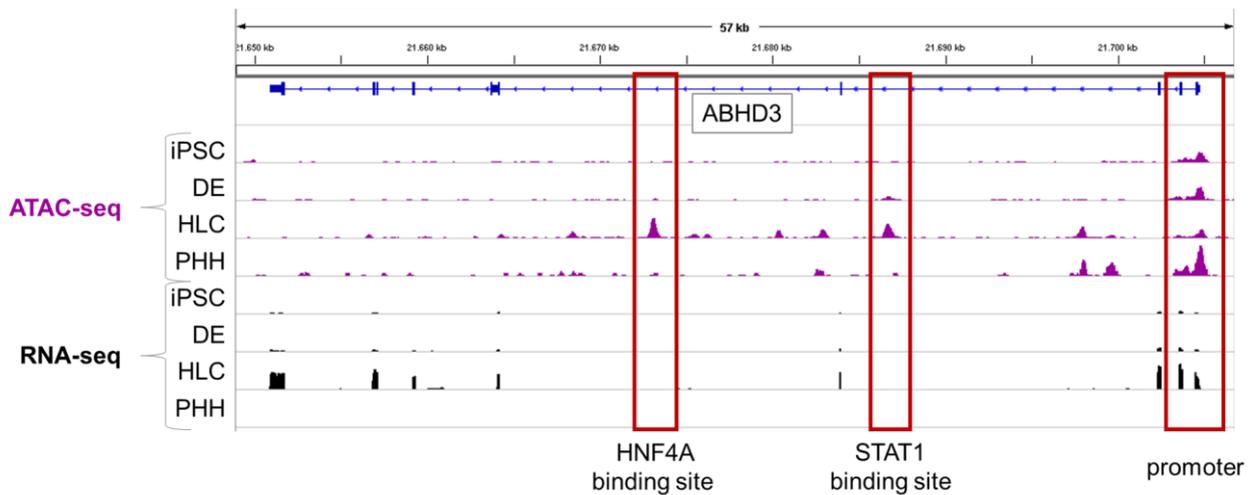
Supplementary Figure 19: Exemplary genome browser representation of multiple putative regulatory regions with differential accessibility within a single gene. ATAC-seq coverage is illustrated in purple, RNA-seq in black. Tracks were normalized by sequencing depth. The red boxes mark DARs with putative regulatory potential.

6.2 Supplementary Tables

Due to page limitations tables with differentially expressed genes (DEGs), differentially methylated regions (DMRs) and differentially accessible regions (DARs) are not included in this thesis. Already published zonal DEGs and DMRs (chapter 3.1.1) can be accessed in the supplement of the online version of the publication (DOI:10.1038/s41467-018-06611-5).

Supplementary Table 1: LCM RRBS data quality parameters including total read count, mapping rate, average (avg) genome coverage, average CpG coverage and number of called CpGs. All libraries featured a bisulfite conversion rate of 0.99.

sample	total read count	mapping rate	avg genome coverage	avg CpG coverage	called CpGs
6963_I	31,237,654	0.90	3.71	4.32	6,750,373
6976_C	52,370,899	0.89	4.78	6.35	10,484,681
6976_I	46,628,637	0.86	4.81	6.40	7,725,009
6976_P	45,382,784	0.88	5.28	6.98	8,200,269
7185_C	53,035,309	0.85	4.86	6.23	8,183,361
7185_I	41,534,700	0.89	4.46	5.55	7,895,321
7196_C	41,878,187	0.87	4.76	5.84	6,114,403
7196_I	62,237,205	0.92	5.38	6.95	11,479,617
7196_P	42,534,844	0.91	4.02	4.97	9,702,319
7203_C	59,458,140	0.90	5.42	7.40	10,633,989
7203_I	50,701,560	0.90	5.08	6.96	10,210,581
7203_P	55,597,887	0.88	5.35	7.32	10,512,605
7259_C	48,430,881	0.86	4.87	6.64	9,004,706
7259_I	60,341,166	0.88	5.41	7.29	9,484,792
7263_C	40,343,552	0.85	4.71	6.01	6,807,134
7263_I	46,800,889	0.85	4.07	5.09	10,018,604
7263_P	52,841,913	0.89	4.84	6.30	10,714,432
7286_C	39,269,942	0.85	3.62	4.57	9,596,088
7299_C	42,691,345	0.86	4.31	5.57	8,133,411
7299_I	36,432,089	0.86	4.04	4.94	6,863,182
7299_P	41,053,207	0.84	5.50	7.48	6,241,629
7319_C	14,411,270	0.85	2.20	2.48	5,300,883
7319_I	36,065,390	0.86	3.79	4.73	8,889,714
7319_P	43,218,212	0.87	4.11	5.10	9,916,216
7322_C	36,871,314	0.91	4.03	5.26	9,580,698
7322_I	52,590,839	0.85	4.63	5.83	10,545,381
7322_P	55,128,731	0.88	6.15	7.91	7,813,691
7373_C	32,165,152	0.86	4.09	4.98	5,800,733
7373_I	48,867,245	0.88	4.34	5.40	10,426,844
7373_P	35,426,905	0.86	3.91	5.19	8,311,389
7415_C	59,192,095	0.88	5.51	7.69	11,104,197
7415_P	59,587,046	0.91	5.02	6.44	11,085,143

7425_C	46,543,806	0.83	4.30	5.68	9,532,779
7425_I	45,527,226	0.87	4.72	6.09	9,623,470
7425_P	46,701,866	0.93	4.92	6.42	9,996,168
7468_P	50,169,169	0.91	4.48	5.56	10,406,403
7479_C	40,702,090	0.82	4.03	5.01	8,646,800
7479_I	71,540,290	0.91	6.33	8.74	11,135,028
7479_P	45,411,948	0.90	4.15	5.10	10,453,561
7484_C	51,679,137	0.87	5.23	7.08	10,446,553
7484_P	41,204,288	0.82	4.87	7.06	9,033,837
7490_C	66,161,203	0.73	5.17	6.90	10,137,990
7490_I	62,015,639	0.83	6.55	10.28	8,337,233
7490_P	53,625,135	0.91	4.84	6.40	11,085,013
7519_C	64,626,975	0.87	5.61	7.58	10,879,717
7519_I	53,084,167	0.93	4.93	6.47	10,723,037
7519_P	53,162,202	0.87	4.75	6.05	9,047,933
7580_P	46,340,321	0.81	4.25	5.45	9,333,561
7522_C	46,315,010	0.89	4.64	6.38	10,063,482
7522_I	56,646,933	0.90	5.44	7.23	11,216,226
7522_P	54,107,863	0.83	5.19	7.18	9,816,943
6610_C	78,873,787	0.95	6.95	11.50	11,072,674
6758_C	65,118,522	0.97	6.26	8.02	12,988,344
6922_C	53,535,977	0.96	5.09	6.63	12,407,130
6967_C	60,589,957	0.97	5.55	7.09	12,672,545
7012_C	86,281,707	0.95	7.79	11.97	12,132,444
7041_C	50,432,583	0.97	4.95	6.00	12,115,510
7137_C	54,099,099	0.96	5.36	7.22	11,545,614
7157_C	54,893,801	0.97	5.36	6.73	12,378,537
7172_C	54,508,741	0.97	4.85	5.63	12,414,924
7173_C	42,424,640	0.96	4.46	5.81	10,988,989
7181_C	50,658,273	0.96	4.74	6.12	11,775,374
7188_C	53,048,049	0.96	4.76	5.83	12,343,298
7194_C	49,942,423	0.96	5.27	6.83	11,719,014
7213_C	53,473,699	0.97	4.97	5.83	12,506,037
7230_C	79,439,403	0.96	6.75	8.62	13,540,026
7252_C	50,772,112	0.97	4.80	6.05	11,646,830
7279_C	47,178,956	0.97	4.44	5.03	11,888,831
7344_C	45,618,003	0.97	4.49	5.07	11,327,475
6610_I	80,974,977	0.97	7.27	10.49	12,776,895
6758_I	51,114,249	0.95	5.65	8.55	11,292,503
6922_I	63,000,603	0.96	6.34	9.07	12,131,656
6967_I	53,426,925	0.96	5.32	7.71	10,831,087
7012_I	114,390,363	0.94	9.37	15.51	11,592,568
7041_I	47,927,194	0.97	4.69	5.90	11,531,708
7137_I	31,550,093	0.97	4.14	5.33	9,712,824
7157_I	49,649,860	0.97	5.02	6.63	12,112,574

7172_I	45,150,619	0.97	4.53	5.28	11,646,988
7173_I	47,267,857	0.97	4.53	5.65	12,191,182
7181_I	57,127,165	0.97	5.44	6.82	12,235,195
7188_I	57,094,561	0.95	5.25	7.79	11,277,766
7213_I	57,242,623	0.97	5.98	8.41	11,440,290
7230_I	71,647,920	0.96	6.18	7.63	13,588,307
7251_I	52,293,589	0.97	5.08	6.09	12,070,208
7252_I	54,044,401	0.97	4.78	5.88	12,330,420
7279_I	48,115,574	0.97	4.74	5.81	11,457,334
7344_I	46,550,121	0.97	4.80	5.75	11,559,383
6610_P	105,023,932	0.97	8.68	12.35	13,915,388
6758_P	59,318,066	0.96	5.94	8.91	11,217,679
6922_P	49,744,185	0.97	5.14	6.75	11,969,638
6967_P	50,403,324	0.96	4.83	6.09	11,995,433
7012_P	82,860,127	0.95	7.35	10.25	12,713,596
7041_P	59,789,547	0.97	5.73	7.20	12,513,637
7137_P	47,437,176	0.96	4.84	6.37	10,950,516
7157_P	52,059,151	0.97	5.29	6.86	12,243,916
7173_P	53,230,505	0.96	5.00	6.54	11,542,335
7181_P	46,074,659	0.97	4.32	5.12	11,773,561
7188_P	34,069,091	0.95	4.12	5.83	9,378,995
7194_P	55,465,687	0.93	5.66	9.19	9,439,465
7213_P	48,550,564	0.96	5.32	7.92	9,570,016
7230_P	164,377,644	0.94	12.58	19.87	13,078,626
7251_P	51,879,923	0.97	4.73	5.68	12,265,899
7252_P	45,474,395	0.97	4.17	4.70	11,839,828
7279_P	60,866,656	0.97	5.00	5.66	12,876,738
7344_P	53,437,003	0.97	4.99	6.12	12,818,384
6963_P	49,323,351	0.94	4.33	5.35	11,000,350

Supplementary Table 2: LCM RNA-seq data quality control parameters including total read count, number of uniquely mapped reads, duplication rate, ribosomal (rRNA) rate, expression profiling efficiency (ratio of exonic reads of total reads), and the numbers of detected transcripts and genes.

Sample	total read count	Uniquely mapped reads	Duplication rate	rRNA rate	expression profiling efficiency	detected transcripts	detected genes
6610_C	20,372,974	19,405,165	0.048	0.053	0.837	106,543	18,925
6610_I	8,997,657	8,446,670	0.061	0.081	0.782	85,446	15,651
6610_P	15,875,136	15,055,647	0.052	0.094	0.794	102,268	18,323
6758_C	21,209,402	20,118,254	0.051	0.055	0.822	105,342	18,599
6758_I	13,499,965	12,797,554	0.052	0.048	0.808	90,306	16,589
6758_P	20,721,728	19,249,572	0.071	0.084	0.741	95,909	18,561
6922_C	24,849,367	23,641,405	0.049	0.057	0.838	107,414	18,909

6922_I	29,287,936	28,248,682	0.035	0.037	0.875	109,388	18,921
6922_P	17,851,441	16,918,434	0.052	0.074	0.815	96,768	17,422
6963_C	20,533,767	19,327,156	0.059	0.033	0.855	101,172	18,723
6963_I	27,688,134	26,335,880	0.049	0.028	0.876	108,232	19,458
6967_C	15,587,489	14,679,054	0.058	0.042	0.804	95,682	17,647
6967_I	23,564,707	22,286,815	0.054	0.081	0.788	104,342	18,568
6967_P	17,585,780	16,564,799	0.058	0.076	0.756	99,397	18,625
6975_C	24,425,410	23,693,616	0.03	0.021	0.871	117,409	19,777
6975_I	14,835,718	14,001,266	0.056	0.041	0.782	95,475	17,377
6975_P	19,797,264	18,843,494	0.048	0.063	0.808	106,872	18,741
6976_C	22,631,156	21,608,765	0.045	0.024	0.873	112,094	19,687
6976_I	24,587,418	23,265,796	0.054	0.031	0.854	110,551	19,676
6976_P	29,224,382	27,562,639	0.057	0.055	0.823	114,823	20,559
7012_C	28,841,327	27,806,751	0.036	0.036	0.865	112,439	19,475
7012_I	9,524,841	8,933,248	0.062	0.037	0.779	84,566	15,669
7012_P	27,484,742	26,059,686	0.052	0.048	0.839	112,551	19,521
7012_P	24,626,743	23,446,757	0.048	0.027	0.886	106,947	19,185
7041_C	18,995,240	18,221,008	0.041	0.035	0.855	106,075	18,174
7041_I	14,834,771	14,094,137	0.05	0.04	0.837	100,221	17,605
7041_P	24,861,409	23,525,946	0.054	0.058	0.807	104,064	18,732
7137_C	27,308,516	26,097,660	0.044	0.046	0.869	110,941	18,957
7137_I	24,932,574	23,976,057	0.038	0.069	0.854	105,911	18,214
7137_P	12,803,812	12,061,260	0.058	0.035	0.813	93,847	16,735
7157_C	23,196,447	22,287,414	0.039	0.041	0.876	112,205	19,037
7157_I	25,372,454	24,405,073	0.038	0.042	0.868	109,482	18,820
7157_P	12,436,822	11,901,599	0.043	0.068	0.838	96,189	17,043
7172_C	11,631,332	11,036,908	0.051	0.041	0.831	95,566	17,083
7172_I	21,421,605	20,461,138	0.045	0.039	0.861	100,969	17,921
7172_P	26,011,043	24,721,176	0.05	0.038	0.838	106,344	18,608
7173_C	17,106,733	16,265,526	0.049	0.047	0.823	102,283	17,906
7173_I	26,712,874	25,285,076	0.053	0.049	0.833	111,054	19,330
7173_P	22,758,133	21,470,373	0.057	0.068	0.812	106,046	18,856
7181_C	22,693,806	21,831,524	0.038	0.042	0.856	109,889	18,852
7181_I	26,132,767	24,975,446	0.044	0.053	0.841	106,978	18,617
7181_P	17,098,492	16,286,521	0.047	0.049	0.818	102,159	18,044
7185_C	28,614,025	26,519,642	0.073	0.086	0.777	107,464	19,785
7185_I	31,973,693	29,850,373	0.066	0.093	0.774	109,408	19,962
7185_P	35,228,513	32,963,779	0.064	0.105	0.775	111,388	20,252
7188_C	18,677,971	17,401,819	0.068	0.056	0.766	101,071	18,865
7188_I	20,401,058	19,016,033	0.068	0.045	0.807	105,444	18,733
7188_P	22,758,239	21,321,940	0.063	0.061	0.787	108,174	19,319
7194_C	15,271,797	14,831,959	0.029	0.024	0.901	98,711	16,558
7194_I	12,400,366	11,863,069	0.043	0.011	0.86	91,276	16,108
7194_P	17,128,790	16,475,601	0.038	0.033	0.887	99,764	17,099
7196_C	32,905,874	30,955,883	0.059	0.033	0.86	113,067	20,133

7196_I	30,223,755	28,642,811	0.052	0.03	0.874	111,735	19,514
7196_P	33,338,563	31,362,948	0.059	0.074	0.819	109,004	19,689
7197_C	23,966,527	23,146,573	0.034	0.036	0.897	112,395	19,137
7197_I	17,279,503	16,747,163	0.031	0.023	0.903	112,433	18,805
7197_P	23,849,106	22,873,247	0.041	0.03	0.885	111,843	19,012
7203_C	25,254,071	24,383,578	0.034	0.029	0.903	112,065	18,979
7203_I	18,590,241	18,045,090	0.029	0.025	0.921	107,588	18,175
7203_P	22,958,604	22,141,837	0.036	0.034	0.903	106,952	18,228
7213_C	19,309,703	18,602,695	0.037	0.027	0.868	104,062	18,276
7213_I	21,410,104	20,457,136	0.045	0.037	0.832	106,438	19,125
7213_P	24,525,289	23,615,864	0.037	0.037	0.869	109,233	19,065
7230_C	17,941,689	17,253,606	0.038	0.022	0.86	103,509	17,843
7230_I	20,543,200	19,716,310	0.04	0.028	0.87	103,501	18,025
7230_P	20,598,592	19,785,637	0.039	0.034	0.874	101,850	18,211
7251_C	23,909,560	22,839,134	0.045	0.041	0.846	112,255	19,116
7251_I	11,673,679	11,001,378	0.058	0.04	0.815	95,699	16,670
7251_P	24,678,676	23,424,149	0.051	0.043	0.843	109,620	18,835
7252_C	24,215,508	23,444,087	0.032	0.025	0.875	114,709	19,237
7252_I	17,506,347	16,963,448	0.031	0.027	0.877	108,192	18,104
7252_P	23,468,671	22,807,197	0.028	0.029	0.898	110,737	18,431
7259_C	26,515,318	25,710,690	0.03	0.027	0.918	112,388	18,703
7259_I	20,387,796	19,729,099	0.032	0.025	0.918	105,426	17,762
7259_P	23,976,406	23,261,309	0.03	0.032	0.917	109,973	18,358
7263_C	24,182,550	23,482,575	0.029	0.025	0.921	104,925	17,627
7263_I	19,630,794	19,149,193	0.025	0.019	0.932	110,904	18,372
7263_P	29,395,685	28,508,284	0.03	0.027	0.918	106,439	17,705
7265_C	28,585,322	27,738,533	0.03	0.033	0.913	118,878	19,851
7265_I	23,719,174	22,974,159	0.031	0.03	0.919	118,533	19,874
7265_P	24,590,322	23,799,093	0.032	0.034	0.909	115,785	19,178
7279_C	23,718,587	22,966,163	0.032	0.022	0.879	120,027	19,997
7279_I	18,167,341	17,714,739	0.025	0.019	0.882	112,481	18,574
7279_P	19,405,294	18,762,239	0.033	0.032	0.878	109,473	18,576
7286_C	39,366,753	37,224,458	0.054	0.038	0.868	113,893	20,364
7286_I	21,834,569	21,009,220	0.038	0.03	0.884	107,599	18,880
7286_P	31,605,763	30,299,156	0.041	0.044	0.877	111,701	19,891
7299_C	26,190,796	24,203,843	0.076	0.034	0.852	109,762	19,455
7299_I	27,967,935	26,608,751	0.049	0.029	0.864	112,989	20,022
7299_P	29,517,612	28,178,451	0.045	0.026	0.891	114,323	20,080
7319_C	30,765,509	28,399,363	0.077	0.054	0.821	108,826	19,980
7319_I	29,868,677	28,623,584	0.042	0.031	0.886	115,861	20,152
7319_P	20,784,206	19,588,923	0.058	0.056	0.842	105,166	19,327
7322_C	34,612,804	32,929,900	0.049	0.028	0.879	114,422	19,869
7322_I	36,861,112	35,117,865	0.047	0.031	0.879	111,805	19,797
7322_P	18,659,566	17,549,091	0.06	0.049	0.837	100,050	18,241
7326_C	25,566,708	24,686,220	0.034	0.036	0.894	107,516	18,419

7326_I	19,884,561	19,348,350	0.027	0.026	0.92	110,806	18,675
7326_P	25,014,359	24,198,610	0.033	0.03	0.905	108,251	18,609
7344_C	20,031,835	19,208,074	0.041	0.025	0.879	112,469	18,850
7344_I	21,041,237	20,338,404	0.033	0.03	0.874	115,262	19,208
7344_P	9,905,518	9,391,313	0.052	0.029	0.844	98,126	16,751
7373_C	22,062,382	21,124,461	0.043	0.02	0.892	110,052	19,289
7373_I	23,970,163	23,089,257	0.037	0.021	0.907	108,725	18,813
7373_P	24,604,473	23,669,820	0.038	0.025	0.906	108,204	18,877
7383_C	24,943,235	24,149,834	0.032	0.03	0.912	111,010	18,544
7383_I	19,187,601	18,630,416	0.029	0.028	0.918	111,965	18,747
7383_P	22,859,272	22,198,371	0.029	0.028	0.92	110,486	18,453
7415_C	43,520,134	41,396,777	0.049	0.028	0.876	118,655	21,123
7415_I	37,249,272	35,642,501	0.043	0.021	0.888	116,181	20,393
7415_P	38,347,162	34,329,132	0.105	0.038	0.81	111,334	20,204
7425_C	31,577,255	30,138,217	0.046	0.028	0.879	119,588	20,963
7425_I	27,099,881	25,421,483	0.062	0.037	0.847	111,862	19,972
7425_P	40,708,236	38,288,682	0.059	0.054	0.835	113,391	20,911
7468_C	24,305,220	23,101,839	0.05	0.035	0.866	110,613	19,722
7468_I	20,638,278	19,788,220	0.041	0.028	0.888	109,869	19,058
7468_P	21,405,864	20,428,346	0.046	0.037	0.886	106,766	18,867
7479_C	22,262,283	21,637,572	0.028	0.031	0.921	112,100	18,777
7479_C	25,600,534	24,817,200	0.031	0.037	0.922	115,225	19,519
7479_I	19,305,954	18,735,219	0.03	0.041	0.918	113,366	18,956
7479_P	23,767,148	23,087,716	0.029	0.033	0.921	109,505	18,526
7479_P	19,376,288	18,707,456	0.035	0.034	0.917	107,497	18,314
7484_C	24,307,239	23,032,112	0.052	0.038	0.874	110,682	19,673
7484_I	24,067,021	23,020,500	0.043	0.034	0.889	109,163	19,120
7484_P	21,755,073	20,441,234	0.06	0.055	0.838	104,979	19,069
7490_C	27,408,005	26,603,200	0.029	0.033	0.917	113,965	19,106
7490_I	19,181,172	18,674,884	0.026	0.023	0.933	113,733	18,798
7490_P	26,361,599	25,502,535	0.033	0.032	0.912	113,554	19,107
7505_C	26,659,140	25,823,080	0.031	0.03	0.909	107,070	18,008
7505_C	22,003,484	21,230,286	0.035	0.032	0.912	105,152	17,922
7505_I	17,080,722	16,635,370	0.026	0.023	0.927	104,550	17,474
7505_P	27,252,275	26,511,917	0.027	0.026	0.921	108,117	18,134
7505_P	21,645,094	20,951,310	0.032	0.03	0.916	104,387	17,875
7519_C	30,296,908	28,431,965	0.062	0.03	0.833	110,779	20,191
7519_I	27,325,559	25,933,976	0.051	0.028	0.863	109,488	19,587
7519_P	24,931,257	22,781,072	0.086	0.063	0.785	101,215	19,293
7522_C	11,092,691	10,749,825	0.031	0.037	0.9	96,836	16,855
7522_I	20,163,430	19,564,184	0.03	0.022	0.917	108,983	18,484
7522_P	24,354,070	23,587,355	0.031	0.028	0.906	110,251	18,840

Supplementary Table 3: *In silico* analysis of TF binding motives in pericentrally (CV) hypomethylated DMRs using HOMER (Brosch et al., 2018). The table includes TFs with enriched binding motives, p-values and the percentage of target sequences (= DMRs) comprising the respective motive. TFs printed in bold are detected exclusively in pericentral DMRs.

CV hypomethylated DMRs								
enriched TF	P-value	target [%]	enriched TF	P-value	target [%]	enriched TF	P-value	target [%]
NF1	1.00E-65	48.82%	**TCFL2**	1.00E-16	1.83%	Gata4	1.00E-06	16.77%
HNF4a	1.00E-62	15.04%	Olig2	1.00E-15	43.78%	**HNF6**	1.00E-05	8.30%
c-Myc	1.00E-56	14.10%	HIF2a	1.00E-15	9.76%	**Pbx3**	1.00E-05	5.29%
SCL	1.00E-54	82.37%	**RUNX-AML**	1.00E-14	17.13%	**TR4**	1.00E-05	3.70%
RXR	1.00E-52	34.09%	BMYB	1.00E-14	27.91%	ARE	1.00E-05	5.92%
ZFX	1.00E-48	44.20%	**Tcf4 (TCF7L2)**	1.00E-13	9.22%	SPDEF	1.00E-05	22.15%
Erra	1.00E-46	61.69%	Smad4	1.00E-12	39.66%	AP-1	1.00E-05	14.00%
Tcf12	1.00E-46	31.50%	**RUNX**	1.00E-12	16.45%	STAT4	1.00E-05	18.21%
MyoG	1.00E-40	32.02%	**CEBP:AP1**	1.00E-11	14.61%	Stat3	1.00E-05	11.32%
n-Myc	1.00E-39	20.97%	**Nur77**	1.00E-11	3.62%	ETV1	1.00E-05	30.87%
USF1	1.00E-39	13.99%	FOXP1	1.00E-11	7.86%	CHR	1.00E-05	10.78%
Ap4	1.00E-39	35.20%	AP2gamma	1.00E-11	30.72%	**HRE**	1.00E-04	3.61%
Myf5	1.00E-39	20.99%	Stat3+il23	1.00E-10	15.56%	Smad3	1.00E-04	59.64%
FOXA1	1.00E-36	23.07%	MafA	1.00E-10	18.54%	Sox3	1.00E-04	29.59%
Foxa2	1.00E-35	14.89%	**Usf2**	1.00E-10	10.48%	Elk4	1.00E-04	11.68%
Max	1.00E-34	18.60%	NF1:FOXA1	1.00E-10	1.19%	Fli1	1.00E-03	24.22%
Esrrb	1.00E-31	18.22%	Gata1	1.00E-10	10.15%	**VDR**	1.00E-03	5.95%
MyoD	1.00E-30	23.57%	Smad2	1.00E-10	38.66%	**Atf4**	1.00E-03	4.44%
NF1	1.00E-30	15.79%	RBPJ:Ebox	1.00E-09	10.26%	GRE	1.00E-03	2.82%
Tcf3	1.00E-30	5.48%	**Bach1**	1.00E-09	1.31%	**E-box**	1.00E-03	2.00%
ZNF711	1.00E-29	58.77%	**Tbx20**	1.00E-08	4.81%	**p53**	1.00E-03	0.21%
AR	1.00E-29	72.75%	Nanog	1.00E-08	68.70%	**PAX5**	1.00E-03	3.14%
FOXA1	1.00E-28	19.46%	**GATA-IR4**	1.00E-08	1.21%	GRE	1.00E-03	4.50%
PPARE	1.00E-27	27.85%	**E2A**	1.00E-08	52.05%	**ERE**	1.00E-02	7.73%
bHLHE40	1.00E-26	4.93%	Bach2	1.00E-08	4.41%	PU.1	1.00E-02	10.60%
Nr5a2	1.00E-26	20.12%	**RUNX2**	1.00E-07	17.99%	**Chop**	1.00E-02	3.23%
Fox:Ebox	1.00E-25	21.06%	**CTCF**	1.00E-07	3.51%	**RARg**	1.00E-02	1.15%
Nr5a2	1.00E-25	14.94%	Bcl6	1.00E-07	27.88%	Hoxc9	1.00E-02	7.73%
HIF-1a	1.00E-25	7.69%	AP-2alpha	1.00E-07	29.24%	GABPA	1.00E-02	19.96%
EBF1	1.00E-23	35.23%	PR	1.00E-07	42.42%	EBF	1.00E-02	6.99%
NeuroD1	1.00E-23	24.62%	Tcfcp2l1	1.00E-07	3.73%	Ets1	1.00E-02	6.15%
Tlx	1.00E-22	15.35%	Gata2	1.00E-07	11.11%	ETS1	1.00E-02	22.55%
ATF3	1.00E-22	5.58%	**RUNX1**	1.00E-07	22.11%	NFAT	1.00E-02	15.29%
FXR	1.00E-21	12.92%	**HEB**	1.00E-07	20.42%	**Tbx5**	1.00E-02	73.18%
Atoh1	1.00E-21	31.31%	ERG	1.00E-06	35.37%	NF-E2	1.00E-02	1.30%
c-Myc	1.00E-20	15.84%	Hnf1	1.00E-06	2.14%	MafK	1.00E-02	5.86%
Ptf1a	1.00E-18	70.78%	Foxo1	1.00E-06	37.94%	TEAD4	1.00E-02	18.92%
CEBP	1.00E-17	17.16%	Arnt:Ahr	1.00E-06	15.47%	**p63**	1.00E-02	9.41%
MYB	1.00E-17	33.13%	Jun-AP1	1.00E-06	5.08%	**FOXA1:AR**	1.00E-02	1.67%

Supplementary Table 4: *In silico* analysis of TF binding motives in periportally (PV) hypomethylated DMRs using HOMER (Brosch et al., 2018). The table includes TFs with enriched binding motives, p-values and the percentage of target sequences (= DMRs) comprising the respective motive. TFs printed in bold are detected exclusively in periportal DMRs.

PV hypomethylated DMRs								
enriched TF	P-value	target [%]	enriched TF	P-value	target [%]	enriched TF	P-value	target [%]
Fli1	1.00E-35	31.82%	**Sox6**	1.00E-08	30.46%	Hoxc9	1.00E-04	8.86%
NF1	1.00E-35	17.54%	HIF2a	1.00E-08	10.01%	RXR	1.00E-04	31.54%
Ap4	1.00E-34	38.87%	Nanog	1.00E-08	71.75%	USF1	1.00E-04	12.28%
ERG	1.00E-28	42.40%	FOXP1	1.00E-08	8.74%	c-Myc	1.00E-03	15.03%
MyoD	1.00E-27	26.97%	**Pdx1**	1.00E-08	17.68%	EBF	1.00E-03	7.57%
GABPA	1.00E-25	25.67%	**Sox2**	1.00E-08	17.90%	NF-E2	1.00E-03	1.51%
Tcf12	1.00E-25	33.48%	Nr5a2	1.00E-07	14.96%	**Cdx2**	1.00E-03	12.03%
ETV1	1.00E-25	37.32%	**HOXA9**	1.00E-07	12.43%	Tcfcp2l1	1.00E-03	3.83%
SCL	1.00E-24	83.85%	**ELF1**	1.00E-07	14.09%	**Maz**	1.00E-03	37.07%
EWS:FLI1	1.00E-23	17.11%	Sox3	1.00E-07	33.33%	TEAD4	1.00E-03	20.49%
Myf5	1.00E-23	23.83%	AP-1	1.00E-07	15.88%	NFAT	1.00E-03	18.00%
EWS:ERG	1.00E-22	19.30%	Smad3	1.00E-07	61.28%	ATF3	1.00E-03	4.80%
ZFX	1.00E-21	44.46%	**GATA3**	1.00E-07	28.29%	**Rfx1**	1.00E-03	4.87%
MyoG	1.00E-20	34.11%	BMYB	1.00E-07	29.73%	PR	1.00E-03	44.48%
NF1	1.00E-20	47.35%	Gata4	1.00E-07	19.06%	**HOXD13**	1.00E-03	19.13%
ETS1	1.00E-20	28.31%	Bach2	1.00E-06	5.00%	FXR	1.00E-02	12.26%
Atoh1	1.00E-19	33.96%	**STAT1**	1.00E-06	6.67%	GRE	1.00E-02	3.36%
Ets1	1.00E-16	8.88%	Max	1.00E-06	17.92%	AP-2alpha	1.00E-02	28.23%
SPDEF	1.00E-16	27.06%	c-Myc	1.00E-06	12.58%	GRE	1.00E-02	5.21%
Elk4	1.00E-16	15.81%	Stat3	1.00E-06	12.82%	STAT6	1.00E-02	10.95%
Tlx	1.00E-16	15.98%	**Rbpj1**	1.00E-06	37.05%	**NFkB-p65**	1.00E-02	14.56%
Elk1	1.00E-14	15.79%	**ELF5**	1.00E-06	17.87%	**ETS:E-box**	1.00E-02	2.78%
Foxa2	1.00E-14	16.15%	Arnt:Ahr	1.00E-06	15.96%	**CArG**	1.00E-02	6.06%
Olig2	1.00E-14	46.29%	n-Myc	1.00E-06	19.49%	CHR	1.00E-02	11.65%
Fox:Ebox	1.00E-13	21.49%	Jun-AP1	1.00E-06	5.80%	Erra	1.00E-02	56.69%
FOXA1	1.00E-12	23.85%	EBF1	1.00E-06	35.07%	STAT6	1.00E-02	10.65%
AR	1.00E-12	74.22%	Gata2	1.00E-05	12.58%	PU.1	1.00E-02	11.92%
Foxo1	1.00E-12	42.40%	**Egr2**	1.00E-05	5.36%	**Unknown**	1.00E-02	11.63%
FOXA1	1.00E-11	20.15%	Hoxb4	1.00E-05	3.89%	**TATA-Box**	1.00E-02	23.59%
ZNF711	1.00E-11	56.60%	**STAT5**	1.00E-05	7.40%	**RFX**	1.00E-02	1.83%
NeuroD1	1.00E-11	25.50%	Esrrb	1.00E-05	17.09%	**X-box**	1.00E-02	2.32%
EGR	1.00E-10	6.40%	**NFkB-p65-Rel**	1.00E-04	1.87%	Hnf1	1.00E-02	2.15%
STAT4	1.00E-10	21.23%	**TEAD**	1.00E-04	15.52%	**IRF4**	1.00E-02	8.69%
Smad4	1.00E-10	41.49%	**ETS**	1.00E-04	8.50%	MafK	1.00E-02	6.18%
Bcl6	1.00E-10	31.78%	RBPJ:Ebox	1.00E-04	11.07%	**PU.1-IRF**	1.00E-02	30.10%
MafA	1.00E-09	20.00%	ARE	1.00E-04	6.74%	**HOXA2**	1.00E-02	2.02%
Lhx3-like	1.00E-09	18.79%	**HEB?**	1.00E-04	20.09%	bHLHE40	1.00E-02	4.46%
Nr5a2	1.00E-09	20.23%	Gata1	1.00E-04	11.22%	AP2gamma	1.00E-02	28.82%
Smad2	1.00E-09	40.34%	NF1:FOXA1	1.00E-04	1.30%	HIF-1a	1.00E-02	6.80%
Stat3+il23	1.00E-08	17.66%	MYB	1.00E-04	33.26%			

Supplementary Table 5: The table summarized published experimental and *in silico* data for TF binding preferences to methylated or unmethylated DNA with respective references provided as Pubmed IDs and the full citations below the table. In addition, an analysis of methylation in HepG2 cells at respective TFBS was performed using DEEP WGBS data (01_HepG2_LiHG_Ct1_WGBS_S_1.fullDEEP) tiled into 500 bp windows (coverage > 10) overlapping TF ChIP-seq peaks (ENCODE V3). TFs preferentially binding methylated CpGs are printed in red.

TF	Methylation status preference	Assay	Methylation binding sites in HepG2 [%]	Reference
ARID3A	NA	NA	23,01	NA
ATF3	unmethylated DNA	*in vitro*	10,91	28473536
BHLHE40	unmethylated DNA	*in vitro*	14,84	28473536
BRCA1	unmethylated DNA	predicted	5,55	29145608
CEBPB	methylated CpGs	*in vitro*	33,38	28473536
CEBPD	moderate influence of CpG methylation / only at specific CpG positions	*in vitro*	11,81	28473536
CHD2	unmethylated DNA	predicted	6,92	29145608
CTCF	moderate influence of CpG methylation / only at specific CpG positions	*in vivo*	31,74	26257180
ELF1	unmethylated DNA	*in vitro*	13,18	28473536
EP300	NA	NA	23,22	NA
ESRRA	NA	NA	18,04	NA
EZH2	NA	NA	21,23	NA
FOSL2	unmethylated DNA	*in vitro*	25,74	23434322
FOXA1	moderate influence of CpG methylation / only at specific CpG positions	*in vitro*	29,43	21029866
FOXA2	NA	NA	27,69	NA
GABPA	unmethylated DNA	*in vitro*	9,00	28473536
GRp20	NA	NA	3,71	NA
HDAC2	NA	NA	16,77	NA
HNF4A	NA	NA	22,62	NA
HNF4G	unmethylated DNA	predicted	22,12	29145608
HSF1	inconclusive	*in vitro*	13,73	28473536
IRF3	inconclusive	*in vitro*	7,01	28473536
JUN	unmethylated DNA	*in vitro*	27,53	28473536
JUND	unmethylated DNA	*in vitro*	27,26	28473536
MAFF	unmethylated DNA	*in vitro*	34,58	28473536
MAFK	moderate influence of CpG methylation / only at specific CpG positions	predicted	35,16	29145608
MAX	unmethylated DNA	*in vitro*	11,17	28473536
MAZ	NA	NA	9,68	NA
MBD4	moderate influence of CpG methylation / only at specific CpG positions	*in vitro*	19,91	21029866
MXI1	unmethylated DNA	predicted	12,86	29145608
MYBL2	unmethylated DNA	*in vitro*	19,26	28473536
MYC	unmethylated DNA	*in vitro*	8,99	28473536
NFIC	methylated CpGs	*in vitro*	19,80	24015356
NR2C2	NA	*NA*	9,66	NA
NRF1	unmethylated DNA	*in vitro*	5,49	23434322
POLR2A	NA	NA	29,74	NA
PPARGC1A	NA	NA	16,97	NA
RAD21	moderate influence of CpG methylation / only at specific CpG positions	predicted	28,38	29145608
RCOR1	NA	NA	12,07	NA
REST	unmethylated DNA	predicted	14,05	29145608
RFX5	methylated CpGs	*in vitro*	15,43	23434322
RXRA	methylated CpGs	*in vitro*	19,49	24015356
SIN3AK20	NA	NA	8,75	NA
SMC3	NA	NA	26,08	NA
SP1	methylated CpGs	*in vitro*	18,52	28473536
	unmethylated DNA	*in vivo*		8090226
SP2	methylated CpGs	*in vitro*	17,65	28473536
SREBP1	NA	NA	7,35	NA
SRF	unmethylated DNA	predicted	18,69	29145608
TAF1	unmethylated DNA	predicted	6,64	29145608
TBP	unmethylated DNA	predicted	7,74	29145608
TCF12	unmethylated DNA	*in vitro*	13,73	28473536
TCF7L2	unmethylated DNA	predicted	20,04	29145608
TEAD4	NA	NA	23,16	NA
USF1	unmethylated DNA	*in vitro*	24,00	28473536
USF2	unmethylated DNA	*in vitro*	18,55	28473536
YY1	unmethylated DNA	*in vitro*	11,77	28473536
ZBTB33	methylated CpGs	*in vitro*	10,57	28473536
	unmethylated DNA	*in vivo*		23693142
ZBTB7A	unmethylated DNA	*in vitro*	11,38	28473536
ZNF274	methylated CpGs	*in vitro*	58,22	28473536

References (Sup. Table 5)

Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143,** 470–484 (2010). [PMID: 21029866]

Blattler, A. *et al.* ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. *Epigenetics Chromatin* **6,** 13 (2013). [PMID: 23693142]

Brandeis, M. *et al.* Sp1 elements protect a CpG island from de novo methylation. *Nature* **371,** 435–438 (1994). [PMID: 8090226]

Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2,** e00726 (2013). [PMID: 24015356]

Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep* **12,** 1184–1195 (2015). [PMID: 26257180]

Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013). [PMID: 23434322]

Wang, G. *et al.* MeDReaders: a database for transcription factors that bind to methylated DNA. *Nucleic Acids Res.* **46,** D146–D151 (2018). [PMID: 29145608]

Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356,** (2017). [PMID: 28473536]

Supplementary Table 6: Gene ontology enrichment analysis of genes with upregulated expression in pericentral hepatocytes. GO terms detected as significantly enriched (p ≤ 0.05) in the gene set associated to pericentrally hypomethylated DMRs are printed in red.

Term	Description	Gene count	PValue	Fold Enrichment
GO:0006805	xenobiotic metabolic process	9	5.6E-06	9.23
GO:0055114	oxidation-reduction process	22	1.6E-05	2.97
GO:0046483	heterocycle metabolic process	4	3.7E-05	53.31
GO:0016098	monoterpenoid metabolic process	4	3.7E-05	53.31
GO:0017144	drug metabolic process	5	3.3E-04	14.81
GO:0016055	Wnt signaling pathway	10	5.7E-04	4.28
GO:0019373	epoxygenase P450 pathway	4	1.4E-03	17.77
GO:0006706	steroid catabolic process	3	1.5E-03	47.98
GO:0097267	omega-hydroxylase P450 pathway	3	5.2E-03	26.65
GO:0006954	inflammatory response	12	8.0E-03	2.53
GO:0030282	bone mineralization	4	8.6E-03	9.41
GO:0042738	exogenous drug catabolic process	3	9.4E-03	19.99
GO:0089711	L-glutamate transmembrane transport	3	1.1E-02	18.45
GO:0071300	cellular response to retinoic acid	5	1.1E-02	5.71
GO:0042573	retinoic acid metabolic process	3	1.3E-02	17.13
GO:0008202	steroid metabolic process	4	1.6E-02	7.44
GO:0090263	positive regulation of canonical Wnt signaling pathway	6	1.7E-02	4.00
GO:0032332	positive regulation of chondrocyte differentiation	3	2.3E-02	12.63
GO:0051968	positive regulation of synaptic transmission. glutamatergic	3	2.3E-02	12.63
GO:0050776	regulation of immune response	7	2.4E-02	3.14
GO:0021761	limbic system development	2	2.5E-02	79.96
GO:0038018	Wnt receptor catabolic process	2	2.5E-02	79.96
GO:0019226	transmission of nerve impulse	3	2.5E-02	11.99
GO:0030178	negative regulation of Wnt signaling pathway	4	2.7E-02	6.15
GO:0007155	cell adhesion	12	2.9E-02	2.09
GO:0045163	clustering of voltage-gated potassium channels	2	3.7E-02	53.31
GO:0060349	bone morphogenesis	3	4.4E-02	8.88
GO:0014047	glutamate secretion	3	4.7E-02	8.57
GO:0043276	anoikis	2	4.9E-02	39.98
GO:0034653	retinoic acid catabolic process	2	4.9E-02	39.98
GO:0006778	porphyrin-containing compound metabolic process	2	4.9E-02	39.98
GO:0038026	reelin-mediated signaling pathway	2	4.9E-02	39.98
GO:0021800	cerebral cortex tangential migration	2	4.9E-02	39.98
GO:0042904	9-cis-retinoic acid biosynthetic process	2	4.9E-02	39.98

Supplementary Table 7: Gene ontology enrichment analysis of genes with upregulated expression in periportal hepatocytes. GO terms detected as significantly enriched (p ≤ 0.05) in the gene set associated to periportally hypomethylated DMRs are printed in red.

Term	Description	Gene count	PValue	Fold Enrichment
GO:0001525	angiogenesis	19	1.4E-06	4.01
GO:0007155	cell adhesion	27	5.7E-06	2.77
GO:0030198	extracellular matrix organization	16	2.0E-05	3.84
GO:0019370	leukotriene biosynthetic process	6	5.0E-05	14.11
GO:0007219	Notch signaling pathway	11	1.7E-04	4.50
GO:0070374	positive regulation of ERK1 and ERK2 cascade	13	3.6E-04	3.49
GO:0006939	smooth muscle contraction	5	4.8E-04	13.07
GO:0006749	glutathione metabolic process	7	1.2E-03	5.88
GO:0051216	cartilage development	7	1.5E-03	5.58
GO:0009612	response to mechanical stimulus	7	1.5E-03	5.58
GO:0042127	regulation of cell proliferation	12	2.0E-03	3.05
GO:0030818	negative regulation of cAMP biosynthetic process	4	2.3E-03	14.47
GO:0030217	T cell differentiation	5	3.5E-03	7.84
GO:0070098	chemokine-mediated signaling pathway	7	3.9E-03	4.64
GO:0010628	positive regulation of gene expression	14	4.1E-03	2.51
GO:0003203	endocardial cushion morphogenesis	4	4.3E-03	11.76
GO:0006750	glutathione biosynthetic process	4	4.3E-03	11.76
GO:0007220	Notch receptor processing	4	4.3E-03	11.76
GO:0010634	positive regulation of epithelial cell migration	5	5.0E-03	7.13
GO:0048146	positive regulation of fibroblast proliferation	6	5.6E-03	5.23
GO:0001503	ossification	7	7.0E-03	4.12
GO:0023019	signal transduction involved in regulation of gene expression	4	7.1E-03	9.90
GO:0019229	regulation of vasoconstriction	4	8.2E-03	9.41
GO:0001942	hair follicle development	5	8.3E-03	6.19
GO:0001657	ureteric bud development	5	8.3E-03	6.19
GO:0007399	nervous system development	14	8.5E-03	2.29
GO:0007626	locomotory behavior	7	8.8E-03	3.92
GO:0030199	collagen fibril organization	5	9.1E-03	6.03
GO:0006520	cellular amino acid metabolic process	5	1.0E-02	5.88
GO:0090023	positive regulation of neutrophil chemotaxis	4	1.1E-02	8.55
GO:0030574	collagen catabolic process	6	1.1E-02	4.41
GO:0001504	neurotransmitter uptake	3	1.2E-02	17.64
GO:0010976	positive regulation of neuron projection development	7	1.2E-02	3.70
GO:0030155	regulation of cell adhesion	5	1.3E-02	5.47
GO:0007165	signal transduction	37	1.4E-02	1.50
GO:0061314	Notch signaling involved in heart development	3	1.5E-02	15.68
GO:0031638	zymogen activation	3	1.5E-02	15.68
GO:0006954	inflammatory response	16	1.5E-02	1.99
GO:0030900	forebrain development	5	1.6E-02	5.11
GO:0007507	heart development	10	1.6E-02	2.57
GO:0071300	cellular response to retinoic acid	6	1.6E-02	4.03
GO:0030335	positive regulation of cell migration	10	1.7E-02	2.56
GO:0006260	DNA replication	9	1.8E-02	2.73
GO:0060174	limb bud formation	3	1.8E-02	14.11
GO:0003184	pulmonary valve morphogenesis	3	1.8E-02	14.11
GO:0001771	immunological synapse formation	3	1.8E-02	14.11
GO:0060411	cardiac septum morphogenesis	3	1.8E-02	14.11
GO:0006268	DNA unwinding involved in DNA replication	3	1.8E-02	14.11
GO:0008347	glial cell migration	3	1.8E-02	14.11
GO:0048863	stem cell differentiation	4	1.9E-02	6.97

ID	GO term		count	p-value	z-score
GO:0001569	patterning of blood vessels		4	2.1E-02	6.72
GO:0048247	lymphocyte chemotaxis		4	2.1E-02	6.72
GO:0007204	positive regulation of cytosolic calcium ion concentration		8	2.4E-02	2.81
GO:0006936	muscle contraction		7	2.6E-02	3.08
GO:0043552	positive regulation of phosphatidylinositol 3-kinase activity		4	2.7E-02	6.07
GO:0001558	regulation of cell growth		6	2.7E-02	3.53
GO:0006814	sodium ion transport		6	2.9E-02	3.48
GO:0071356	cellular response to tumor necrosis factor		7	3.0E-02	2.99
GO:0001974	blood vessel remodeling		4	3.0E-02	5.88
GO:0003215	cardiac right ventricle morphogenesis		3	3.0E-02	10.85
GO:0033630	positive regulation of cell adhesion mediated by integrin		3	3.0E-02	10.85
GO:1900025	negative regulation of cell spreading		3	3.0E-02	10.85
GO:0031274	positive regulation of pseudopodium assembly		3	3.0E-02	10.85
GO:0051926	negative regulation of calcium ion transport		3	3.0E-02	10.85
GO:0001666	response to hypoxia		9	3.0E-02	2.46
GO:0045747	positive regulation of Notch signaling pathway		4	3.2E-02	5.70
GO:2000147	positive regulation of cell motility		3	3.4E-02	10.08
GO:0001837	epithelial to mesenchymal transition		4	3.5E-02	5.53
GO:0030097	hemopoiesis		5	3.6E-02	3.99
GO:0017158	regulation of calcium ion-dependent exocytosis		4	3.8E-02	5.38
GO:0048661	positive regulation of smooth muscle cell proliferation		5	3.8E-02	3.92
GO:0050679	positive regulation of epithelial cell proliferation		5	3.8E-02	3.92
GO:0035050	embryonic heart tube development		3	3.9E-02	9.41
GO:0048469	cell maturation		4	4.0E-02	5.23
GO:0008284	positive regulation of cell proliferation		17	4.1E-02	1.72
GO:0070378	positive regulation of ERK5 cascade		2	4.2E-02	47.04
GO:0048866	stem cell fate specification		2	4.2E-02	47.04
GO:0035759	mesangial cell-matrix adhesion		2	4.2E-02	47.04
GO:0051965	positive regulation of synapse assembly		5	4.2E-02	3.79
GO:0006940	regulation of smooth muscle contraction		3	4.4E-02	8.82
GO:0001568	blood vessel development		4	4.6E-02	4.95
GO:1902476	chloride transmembrane transport		6	4.8E-02	3.03
GO:0014068	positive regulation of phosphatidylinositol 3-kinase signaling		5	4.9E-02	3.62
GO:0006953	acute-phase response		4	4.9E-02	4.82
GO:0048791	calcium ion-regulated exocytosis of neurotransmitter		4	4.9E-02	4.82
GO:0048265	response to pain		3	4.9E-02	8.30
GO:0008584	male gonad development		6	5.0E-02	3.00

Supplementary Table 8: Zone-specific gene ontology enrichment of DEGs in alcoholic cirrhosis (ACI) compared to healthy obese donors (HO). Positive z-scores (grey) for significantly enriched (p ≤ 0.05) GO terms represent upregulation in cirrhosis, while negative z-scores (blue) relate to downregulation. The table displays z-scores with corresponding log10 p-values for the top 100 GO terms enriched in the pericentral (C), intermediate (I) or periportal (P) zone. Z-scores for GO terms also enriched in the corresponding zonal comparisons of NASH or FIB3 with HO are printed in bold.

ID	GO term	log10(p-value) C	zscore C	log10(p-value) I	zscore I	log10(p-value) P	zscore P
GO:0043401	steroid hormone mediated signaling pathway	4.24	-3.36	3.45	-2.50	3.41	-2.89
GO:0030522	intracellular receptor signaling pathway	3.45	**-3.32**	1.39	-1.89	1.67	-2.65
GO:0045893	positive regulation of transcription, DNA-templated	4.16	-3.00	1.49	**-2.48**	3.57	-3.22
GO:0060070	canonical Wnt signaling pathway	1.36	-2.89	1.62	-2.31	0.00	0.00
GO:0008152	metabolic process	5.64	-2.83	4.64	**-2.27**	3.94	-1.22
GO:0000122	negative regulation of transcription from RNA polymerase II pro	1.32	-2.81	0.00	0.00	1.48	-4.18

GO ID	Description						
GO:0006631	fatty acid metabolic process	2.34	2.71	5.22	2.84	2.55	2.53
GO:0006805	xenobiotic metabolic process	5.49	2.68	5.47	1.61	4.43	0.50
GO:0046487	glyoxylate metabolic process	1.39	2.45	0.00	0.00	0.00	0.00
GO:0006699	bile acid biosynthetic process	1.77	2.45	1.31	2.24	0.00	0.00
GO:0006069	ethanol oxidation	2.93	2.45	4.21	2.65	2.41	1.34
GO:0042632	cholesterol homeostasis	2.14	2.31	2.97	3.05	0.00	0.00
GO:0009083	branched-chain amino acid catabolic process	1.31	2.24	0.00	0.00	0.00	0.00
GO:0043691	reverse cholesterol transport	1.39	2.24	0.00	0.00	0.00	0.00
GO:0030193	regulation of blood coagulation	1.47	2.24	1.61	1.34	2.64	2.45
GO:0006810	transport	2.15	2.21	1.51	2.74	1.48	2.92
GO:0006629	lipid metabolic process	2.18	2.13	1.43	0.94	1.68	0.24
GO:0017144	drug metabolic process	2.52	2.12	0.00	0.00	0.00	0.00
GO:0006633	fatty acid biosynthetic process	2.34	2.11	2.65	1.51	1.52	1.41
GO:0055114	oxidation-reduction process	11.40	2.06	10.99	2.47	9.07	0.00
GO:0006706	steroid catabolic process	2.43	2.00	0.00	0.00	0.00	0.00
GO:0042135	neurotransmitter catabolic process	1.76	2.00	1.87	2.00	0.00	0.00
GO:0070327	thyroid hormone transport	1.61	2.00	1.72	2.00	1.90	1.00
GO:0055085	transmembrane transport	2.32	1.98	0.00	0.00	0.00	0.00
GO:0005978	glycogen biosynthetic process	4.16	1.90	4.50	2.53	1.80	2.45
GO:0008209	androgen metabolic process	3.15	1.89	2.49	1.63	1.93	1.34
GO:0006210	thymine catabolic process	1.80	1.73	1.88	1.73	0.00	0.00
GO:0032922	circadian regulation of gene expression	2.53	1.73	0.00	0.00	2.27	1.90
GO:0010906	regulation of glucose metabolic process	1.68	1.63	0.00	0.00	1.44	1.34
GO:0015695	organic cation transport	2.44	1.63	0.00	0.00	2.91	1.63
GO:0009636	response to toxic substance	2.00	1.60	0.00	0.00	1.59	0.90
GO:0008285	negative regulation of cell proliferation	1.34	1.26	0.00	0.00	1.82	2.54
GO:0006071	glycerol metabolic process	1.36	1.00	2.33	1.34	0.00	0.00
GO:0007157	heterophilic cell-cell adhesion via plasma membrane cell adhes	1.51	1.00	1.74	1.67	0.00	0.00
GO:0051443	positive regulation of ubiquitin-protein transferase activity	1.68	0.82	2.56	0.38	1.44	1.34
GO:0070328	triglyceride homeostasis	1.39	0.82	1.55	1.63	1.80	0.82
GO:1901215	negative regulation of neuron death	1.57	0.71	1.78	1.41	0.00	0.00
GO:0001678	cellular glucose homeostasis	1.77	0.45	0.00	0.00	1.35	2.00
GO:0036498	IRE1-mediated unfolded protein response	0.00	0.00	1.35	3.00	1.69	3.00
GO:0098609	cell-cell adhesion	0.00	0.00	1.55	3.78	1.35	2.04
GO:0000302	response to reactive oxygen species	0.00	0.00	1.34	2.65	1.62	2.65
GO:0051603	proteolysis involved in cellular protein catabolic process	0.00	0.00	1.38	1.41	2.80	3.16
GO:0090263	positive regulation of canonical Wnt signaling pathway	0.00	0.00	1.81	1.50	2.80	2.67
GO:0002223	stimulatory C-type lectin receptor signaling pathway	0.00	0.00	0.00	0.00	2.51	3.87
GO:0002479	antigen processing and presentation of exogenous peptide anti	0.00	0.00	0.00	0.00	4.25	3.74
GO:0031145	anaphase-promoting complex-dependent catabolic process	0.00	0.00	0.00	0.00	2.71	3.61
GO:0038061	NIK/NF-kappaB signaling	0.00	0.00	0.00	0.00	3.42	3.61
GO:0006521	regulation of cellular amino acid metabolic process	0.00	0.00	0.00	0.00	3.86	3.46
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in	0.00	0.00	0.00	0.00	2.36	3.46
GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in	0.00	0.00	0.00	0.00	2.59	3.46
GO:0000209	protein polyubiquitination	0.00	0.00	0.00	0.00	2.27	3.27
GO:0050852	T cell receptor signaling pathway	0.00	0.00	0.00	0.00	2.62	2.98
GO:0038095	Fc-epsilon receptor signaling pathway	0.00	0.00	0.00	0.00	2.44	2.84
GO:0043488	regulation of mRNA stability	0.00	0.00	0.00	0.00	2.17	2.67
GO:0060071	Wnt signaling pathway, planar cell polarity pathway	0.00	0.00	0.00	0.00	4.13	2.67
GO:0045454	cell redox homeostasis	0.00	0.00	0.00	0.00	1.46	2.53

GO ID	Description						
GO:0051092	positive regulation of NF-kappaB transcription factor activity	0.00	0.00	1.45	2.50	0.00	0.00
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic pr	0.00	0.00	0.00	0.00	1.82	2.40
GO:0000165	MAPK cascade	0.00	0.00	0.00	0.00	1.98	2.35
GO:0032930	positive regulation of superoxide anion generation	0.00	0.00	1.72	2.00	0.00	0.00
GO:0030970	retrograde protein transport, ER to cytosol	0.00	0.00	0.00	0.00	1.35	2.00
GO:0060337	type I interferon signaling pathway	0.00	0.00	0.00	0.00	1.94	1.90
GO:0007017	microtubule-based process	0.00	0.00	1.49	1.89	0.00	0.00
GO:0006099	tricarboxylic acid cycle	0.00	0.00	0.00	0.00	2.24	1.89
GO:0006487	protein N-linked glycosylation	0.00	0.00	0.00	0.00	1.62	1.89
GO:0060736	prostate gland growth	0.00	0.00	1.87	2.00	0.00	0.00
GO:0090009	primitive streak formation	0.00	0.00	1.59	2.00	0.00	0.00
GO:0046889	positive regulation of lipid biosynthetic process	0.00	0.00	1.37	2.00	0.00	0.00
GO:0046628	positive regulation of insulin receptor signaling pathway	0.00	0.00	0.00	0.00	1.44	-2.00
GO:0006367	transcription initiation from RNA polymerase II promoter	0.00	0.00	0.00	0.00	1.50	-2.00
GO:0033539	fatty acid beta-oxidation using acyl-CoA dehydrogenase	0.00	0.00	1.53	2.24	0.00	0.00
GO:0005978	glycogen biosynthetic process	0.00	0.00	0.00	0.00	1.80	-2.45
GO:0001889	liver development	0.00	0.00	1.97	1.73	2.96	-0.83
GO:0005977	glycogen metabolic process	0.00	0.00	1.93	2.65	0.00	0.00
GO:0031016	pancreas development	0.00	0.00	1.31	1.34	1.51	-1.34
GO:0007623	circadian rhythm	0.00	0.00	1.54	1.51	1.53	-1.26
GO:0040018	positive regulation of multicellular organism growth	0.00	0.00	1.55	1.13	1.84	-1.89
GO:0050821	protein stabilization	1.90	0.23	4.40	1.22	3.04	0.69
GO:0034614	cellular response to reactive oxygen species	2.57	0.38	1.37	1.34	1.58	0.45
GO:0008286	insulin receptor signaling pathway	1.53	0.58	1.81	1.15	1.43	0.63
GO:0050918	positive chemotaxis	1.87	0.71	1.55	0.38	1.84	1.89
GO:0006103	2-oxoglutarate metabolic process	2.06	0.82	1.53	0.45	1.74	1.34
GO:0098869	cellular oxidant detoxification	2.27	0.83	0.00	0.00	2.64	1.73
GO:2000379	positive regulation of reactive oxygen species metabolic proces	1.67	1.13	1.31	1.63	0.00	0.00
GO:0071222	cellular response to lipopolysaccharide	1.38	1.29	1.38	1.60	0.00	0.00
GO:0043066	negative regulation of apoptotic process	3.57	1.34	1.60	1.68	0.00	0.00
GO:2001238	positive regulation of extrinsic apoptotic signaling pathway	2.63	1.41	2.17	1.13	0.00	0.00
GO:0002576	platelet degranulation	1.68	1.50	1.67	0.53	1.43	1.94
GO:0071260	cellular response to mechanical stimulus	1.43	1.51	1.31	1.26	0.00	0.00
GO:0006102	isocitrate metabolic process	1.32	1.73	1.40	1.73	1.52	1.73
GO:0019673	GDP-mannose metabolic process	1.52	1.73	1.60	1.73	0.00	0.00
GO:0033591	response to L-ascorbic acid	1.32	1.73	1.40	1.73	0.00	0.00
GO:0006935	chemotaxis	1.42	2.00	0.00	0.00	0.00	0.00
GO:0006163	purine nucleotide metabolic process	1.47	2.00	0.00	0.00	0.00	0.00
GO:0006954	inflammatory response	1.78	2.16	0.00	0.00	0.00	0.00
GO:0030198	extracellular matrix organization	1.44	2.29	0.00	0.00	0.00	0.00
GO:0006953	acute-phase response	3.35	2.31	3.02	1.26	1.62	2.12
GO:0070527	platelet aggregation	2.01	2.33	0.00	0.00	0.00	0.00
GO:0010951	negative regulation of endopeptidase activity	1.44	2.67	2.13	1.21	0.00	0.00
GO:0033209	tumor necrosis factor-mediated signaling pathway	1.53	3.50	0.00	0.00	2.87	4.12

Supplementary Table 9: Spectrophotometric quality control of DNA and RNA extracted from iPSC, DE and HLC samples. Concentrations are given in ng/µl. The ratio of absorbance at 260 nm and 280 nm (A260/280) indicates high purity.

Sample	DNA		RNA	
	Conc. [ng/µl]	A260/280	Conc. [ng/µl]	A260/280
CEL R1 DE	153	1.89	199	1.94
CEL R1 HLC	138	1.90	219	1.98
CEL R1 iPSC	62	1.94	233	1.92
CEL R2 DE	127	1.98	311	1.99
CEL R2 HLC	108	1.89	198	1.99
CEL R2 iPSC	120	1.90	348	1.99
CEL R3 DE	69	1.93	260	1.93
CEL R3 HLC	122	1.89	237	1.96
CEL R3 iPSC	88	1.94	388	2.01
CEL R4 DE	111	1.91	344	1.97
CEL R4 HLC	72	1.96	259	1.98
CEL R4 iPSC	131	1.90	398	1.99
HAY R1 DE	75	1.96	173	1.95
HAY R1 HLC	100	1.94	237	1.96
HAY R1 iPSC	116	1.90	349	1.99

Supplementary Table 10: StemNet RNA-seq data quality control parameters including total read count, number of uniquely mapped reads, duplication rate, ribosomal (rRNA) rate, expression profiling efficiency (ratio of exonic reads of total reads), and the numbers of detected transcripts and genes.

Sample	total read count	Uniquely mapped reads	Duplication rate	rRNA rate	expression profiling efficiency	detected transcripts	detected genes
CEL_DE_R1	66,635,306	65,104,822	0.023	0.111	0.605	141,251	29,845
CEL_HLC_R1	68,603,284	67,363,592	0.018	0.094	0.742	138,926	27,655
CEL_iPSC_R1	28,015,364	26,942,622	0.038	0.080	0.502	134,295	28,102
CEL_DE_R2	37,601,502	36,568,469	0.027	0.087	0.563	133,645	26,886
CEL_HLC_R2	53,916,362	52,925,958	0.018	0.072	0.740	135,544	26,568
CEL_iPSC_R2	23,930,588	23,239,334	0.029	0.114	0.620	126,785	23,147
CEL_DE_R3	40,466,027	39,524,554	0.023	0.092	0.640	129,342	25,064
CEL_HLC_R3	47,798,580	46,803,429	0.021	0.090	0.697	135,755	26,876
CEL_iPSC_R3	70,461,192	68,777,602	0.024	0.105	0.658	144,244	29,552
CEL_DE_R4	29,206,262	28,315,835	0.030	0.062	0.522	129,230	26,360
CEL_HLC_R4	64,299,419	63,180,075	0.017	0.114	0.764	132,874	24,519
CEL_iPSC_R4	44,840,143	43,691,370	0.026	0.122	0.626	137,336	26,541
HAY_DE_R1	21,946,193	21,162,853	0.036	0.054	0.459	128,546	26,276
HAY_HLC_R1	34,390,444	32,921,281	0.043	0.029	0.351	146,440	32,942
HAY_iPSC_R1	59,987,661	58,544,857	0.024	0.108	0.610	148,696	30,600

Supplementary Table 11: StemNet RRBS data quality parameters including total read count, mapping rate, bisulfite conversion rate, average (avg) genome coverage, average CpG coverage and number of called CpGs.

Sample	total read count	mapping rate	conversion rate	avg genome coverage	avg CpG coverage	called CpGs
CEL_DE_R1	43,557,490	0.94	0.99	7.58	10.22	6,566,662
CEL_HLC_R1	53,836,678	0.97	0.99	9.52	14.68	7,699,159
CEL_iPSC_R1	48,966,686	0.97	0.98	8.99	13.68	7,315,052
CEL_DE_R2	59,,447,079	0.96	0.99	9.63	14.3	7,457,099
CEL_HLC_R2	47,986,185	0.94	0.98	7.44	9.7	7,501,635
CEL_iPSC_R2	72,185,776	0.94	0.97	9.25	12.93	7,948,112
CEL_DE_R3	59,588,906	0.94	0.99	9.45	14.39	7,331,907
CEL_HLC_R3	52,633,477	0.97	0.98	9.44	13.77	7,600,784
CEL_iPSC_R3	50,878,758	0.97	0.97	8.92	13.2	7,476,408
CEL_DE_R4	44,957,261	0.92	0.98	7.21	10.06	6,642,690
CEL_HLC_R4	36,710,571	0.97	0.99	6.05	8.13	7,288,060
CEL_iPSC_R4	38,108,376	0.9	0.97	6.25	7.89	5,936,358
HAY_DE_R1	28,173,493	0.9	0.99	4.8	5.7	5,460,617
HAY_HLC_R1	41,799,606	0.95	0.99	8.01	11.58	6,607,657
HAY_iPSC_R1	52,655,392	0.94	0.98	8.71	12.81	7,150,398

Supplementary Table 12: StemNet ATAC-seq data quality parameters including total read count, uniquely mapped reads, duplication rate, the ratio of mitochondrial (mt) reads, MACS2 peak count and fraction of reads located in peaks (FriP). Critical values are printed in bold.

Sample	total read count	uniquely mapped reads	duplication rate	ratio of mt reads	peak count	FriP score
CEL_DE_R1	62,211,956	28,314,190	0.545	0.376	116,564	0.251
CEL_DE_R2	49,604,414	21,706,009	0.562	0.403	95,624	0.224
CEL_DE_R3	71,404,549	31,432,607	0.560	0.382	123,399	0.248
CEL_DE_R4	47,113,792	19,818,272	0.579	0.420	86,897	0.197
CEL_HLC_R1	43,803,980	15,157,323	0.654	0.349	**7,334**	**0.016**
CEL_HLC_R2	42,952,657	28,202,218	0.343	0.181	78,007	0.108
CEL_HLC_R3	23,724,295	13,888,911	0.415	0.159	32,561	0.064
CEL_HLC_R4	41,383,867	27,992,852	0.324	0.180	103,601	0.165
CEL_iPSC_R1	39,558,795	12,284,214	0.689	0.513	55,762	0.135
CEL_iPSC_R2	52,749,164	17,305,708	0.672	0.511	64,965	0.132
CEL_iPSC_R3	32,870,552	11,787,419	0.641	0.491	44,215	0.099
CEL_iPSC_R4	41,895,393	11,782,586	0.719	0.552	53,292	0.128
HAY_DE_R1	48,286,327	20,974,314	0.566	0.386	117,677	0.202
HAY_HLC_R1	39,342,022	16,864,979	0.571	0.359	125,903	0.386
HAY_iPSC_R1	57,626,560	24,863,112	0.569	0.397	137,610	0.253
PHH_AFJ	51,382,001	**4,060,977**	**0.921**	**0.737**	45,610	0.277
PHH_DJJ	82,714,835	**4,224,927**	**0.949**	**0.728**	32,675	0.177
PHH_IAN	53,530,355	**7,868,751**	**0.853**	**0.662**	14,639	**0.049**

Supplementary Table 13: Spearman correlations of CDX2, ALB and AFP expression levels and TF regulon activities in single cells. Positive correlations above 0.5 are highlighted in blue, negative correlations below 0.5 in grey.

Regulon	CDX2	ALB	AFP	Regulon	CDX2	ALB	AFP
HLF	-0.45	0.82	-0.43	CREB5	0.36	-0.32	0.60
NR1I2	-0.41	0.81	-0.40	EBF4	0.56	-0.33	0.57
NFIC	-0.45	0.81	-0.44	ATF3	0.68	-0.36	0.74
NR1I3	-0.40	0.81	-0.40	KLF6	0.64	-0.36	0.77
AR	-0.42	0.80	-0.43	RELB	0.41	-0.37	0.52
NFIL3	-0.38	0.78	-0.31	CREB3L2	0.54	-0.38	0.71
CEBPD	-0.47	0.76	-0.31	ETV5	0.52	-0.39	0.48
CEBPB	-0.33	0.75	-0.37	CDX2	0.73	-0.40	0.65
HSF4	-0.50	0.75	-0.48	CREB3L1	0.54	-0.40	0.56
RXRB	-0.44	0.74	-0.48	GATA5	0.66	-0.41	0.62
PPARG	-0.48	0.74	-0.27	CEBPG	0.50	-0.43	0.68
PPARGC1A	-0.42	0.68	-0.50	SP1	0.54	-0.43	0.57
CUX2	-0.49	0.68	-0.59	ARID5B	0.42	-0.44	0.55
NR1H4	-0.17	0.67	-0.21	ARID3A	0.51	-0.45	0.63
KLF15	-0.59	0.65	-0.67	BHLHE40	0.68	-0.47	0.73
ETV2	-0.55	0.65	-0.62	KLF7	0.39	-0.48	0.51
CREBL2	-0.56	0.64	-0.64	FOSL2	0.71	-0.50	0.68
PPARA	-0.11	0.61	-0.11	GTF3C2	0.20	-0.50	0.09
SRY	-0.39	0.57	-0.30	POLR2A	0.27	-0.51	0.40
ESR1	-0.33	0.57	-0.52	ELK3	0.17	-0.52	0.30
RXRA	-0.21	0.57	-0.39	E2F8	0.10	-0.52	0.08
FOXP3	-0.42	0.56	-0.59	HDAC2	0.05	-0.52	0.12
RORA	-0.47	0.55	-0.58	TAF7	0.13	-0.53	0.20
NR2F2	-0.45	0.55	-0.66	E2F4	0.10	-0.53	0.22
FOXA3	-0.41	0.54	-0.43	CLOCK	0.50	-0.53	0.55
CFL2	-0.44	0.51	-0.37	FOSL1	0.28	-0.54	0.47
ZNF841	-0.41	0.50	-0.44	ZNF516	0.34	-0.54	0.18
RARA	-0.14	0.50	-0.18	IRF1	0.58	-0.54	0.49
POU2F3	-0.34	0.49	-0.52	MAZ	0.11	-0.54	0.14
KLF11	-0.62	0.49	-0.69	ETS1	0.26	-0.54	0.40
HSF1	-0.42	0.43	-0.62	BCLAF1	0.23	-0.55	0.41
POLE4	0.32	0.14	0.50	SRF	0.41	-0.56	0.58
SOX9	0.38	-0.08	0.64	JDP2	0.34	-0.56	0.39
ZNF189	0.32	-0.11	0.52	RUNX1	0.29	-0.56	0.43
HNF4G	0.64	-0.15	0.62	CREB3	0.52	-0.56	0.65
MAF	0.52	-0.17	0.48	BCL11A	0.39	-0.57	0.31
TFF3	0.62	-0.22	0.66	CREB1	0.16	-0.57	0.23
TCF7L2	0.63	-0.24	0.75	KLF13	0.30	-0.57	0.37
ETS2	0.36	-0.28	0.52	E2F7	0.12	-0.59	0.14
KLF5	0.69	-0.28	0.63	SOX4	0.65	-0.60	0.67
ELF3	0.64	-0.28	0.66	CTCF	0.25	-0.60	0.39
IRF6	0.54	-0.30	0.53	ELF1	0.39	-0.63	0.51
EGR1	0.65	-0.32	0.62	MAX	0.67	-0.63	0.65
ESRRG	0.59	-0.32	0.76	SOX11	0.37	-0.64	0.40

Supplementary Table 14: Top 50 most significant (adjusted p-value < 0.05) key regulators of HLC differentiation identified by REGGAE as described in Figure 46. Transcription factors identified uniquely in one cluster are highlighted in yellow.

Cluster 1 TF	adj. p	Cluster 2 TF	adj. p	Cluster 3 TF	adj. p	Cluster 4 TF	adj. p	Cluster 5 TF	adj. p	Cluster 6 TF	adj. p	Cluster 7 TF	adj. p	Cluster 8 TF	adj. p
JARID2	1.24E-16	KDM2B	2.87E-22	KDM2B	3.79E-25	JARID2	7.65E-25	SIX2	0.026	CXXC1	1.81E-07	JARID2	2.18E-13	FXR	0.003
LMNB1	8.30E-16	MBD4	1.22E-21	CNOT3	8.00E-25	KMT2A	3.22E-16			ASXL1	1.81E-07	SUMO2	3.55E-09	SMC1A	0.022
MBD4	8.99E-15	KMT2A	5.71E-19	KMT2A	3.80E-24	CNOT3	3.38E-16			RAG2	9.67E-07	LMNB1	5.82E-09		
KMT2A	1.71E-14	CNOT3	7.21E-19	TFAP2A	6.24E-23	HID1	2.62E-14			CDK9	1.17E-06	MBD4	1.26E-08		
EOMES	1.43E-13	SUMO2	8.24E-19	SUMO2	9.45E-22	KDM2B	3.63E-14			ARNTL	1.17E-06	TFAP2A	3.71E-08		
SMC1A	2.56E-13	SMC1A	9.54E-19	SMC1A	1.33E-20	TFAP2A	1.44E-13			GRHL2	1.70E-06	SMC1A	7.48E-08		
CTNNB1	1.18E-12	TET2	4.01E-18	DPEP3	6.52E-19	PMEPA1	1.66E-12			KDM4C	1.70E-06	CTNNB1	1.52E-07		
KDM2B	1.40E-12	TFAP2A	5.71E-17	CHD8	6.52E-19	SUMO2	3.09E-12			KMT2A	1.75E-06	KDM2B	1.76E-07		
PMEPA1	2.00E-12	JARID2	7.26E-17	RUNX1T1	6.80E-19	DMC1	4.76E-12			CNOT3	1.75E-06	KMT2A	7.26E-07		
CNOT3	3.07E-12	MYH11	3.60E-16	NIPBL	9.03E-19	SMC1A	2.83E-11			BRPF3	1.75E-06	HID1	7.26E-07		
TFAP2A	3.07E-12	ARNTL	3.60E-16	ASXL1	1.64E-18	ERG	7.27E-11			ZNF750	1.84E-06	CNOT3	8.12E-07		
CHD8	3.61E-12	SMAD4	2.57E-15	MBD4	1.64E-18	ARNTL	8.66E-11			ING3	1.87E-06	ERG	1.16E-06		
GRHL2	9.60E-12	DPEP3	2.84E-15	RAG2	2.73E-18	HIF1A	5.90E-10			KMT2D	2.57E-06	LMNA	1.34E-06		
SMAD4	1.30E-11	ERG	1.03E-14	PMEPA1	6.64E-18	ZFP42	1.37E-09			ING5	4.27E-06	RUNX2	1.34E-06		
SNAI2	2.25E-11	SNAI2	1.38E-14	ARNTL	2.09E-17	NIPBL	2.15E-09			BRD1	4.27E-06	PMEPA1	2.03E-06		
MEF2B	2.76E-11	ZMYND8	1.38E-14	E2F3	2.77E-17	KDM4A	2.57E-09			ZMYND8	4.71E-06	CHD8	2.12E-06		
TCF4	3.27E-11	GARS	1.78E-14	TET2	4.74E-17	CHD8	2.75E-09			PHF2	4.78E-06	TOP2B	7.44E-06		
ZNF217	4.47E-11	SMAD3	2.05E-14	ING3	4.74E-17	TET2	2.94E-09			KDM2B	6.32E-06	YAP1	8.67E-06		
RUNX2	4.47E-11	HID1	2.32E-14	PBX1	4.74E-17	CDK9	7.09E-09			SUMO2	7.68E-06	EOMES	9.80E-06		
LMNA	5.17E-11	CHD8	2.54E-14	HID1	7.81E-17	SNAI2	3.83E-08			KLF5	9.65E-06	TET2	2.71E-05		
HIF1A	1.06E-10	YAP1	2.64E-14	CDK9	7.81E-17	TFAP2A	3.83E-08			TFAP2A	1.29E-05	TCF4	3.36E-05		
TP63	1.43E-10	KDM4A	3.82E-14	FLI1	8.24E-17	TFAP2C	3.83E-08			MYH11	1.43E-05	SMAD4	1.12E-04		
ESR2	1.54E-10	KDM4C	4.41E-14	KDM4C	8.24E-17	SMAD4	5.69E-08			FLI1	1.47E-05	DMC1	1.37E-04		
DPEP3	2.50E-10	TFAP2C	1.11E-13	CXXC1	8.44E-17	GARS	7.23E-08			MBD4	1.47E-05	ZMYND8	4.42E-04		
ING3	5.19E-10	NR1H3	1.64E-13	DMC1	9.66E-17	GRHL2	9.34E-08			SMC1A	1.47E-05	GTF2I	4.42E-04		
PPARD	5.88E-10	FLI1	2.34E-13	TFAP2C	1.80E-16	ESR2	2.33E-07			ZNF217	1.47E-05	MYH11	5.57E-04		
PDX1	5.96E-10	MEF2B	2.79E-13	HOXA4	2.37E-16	PPARD	4.79E-07			RUNX1T1	1.47E-05	FLI1	0.002		
YAP1	9.97E-10	CTNNB1	3.06E-13	MYH11	3.06E-16	CTBP2	7.30E-07			BCL6	1.47E-05	PPARG	0.002		
ARNTL	1.39E-09	DMC1	4.00E-13	KMT2D	6.26E-16	RUNX1T1	9.49E-07			MYOD1	1.47E-05	ESR2	0.002		
CDK9	2.12E-09	ING3	1.68E-12	HIF1A	7.13E-16	SPDEF	1.06E-06			BRD3	1.47E-05	TP63	0.002		
NIPBL	2.86E-09	CXXC1	2.00E-12	KDM4A	7.16E-16	DPEP3	1.15E-06			NELFE	1.79E-05	CDK9	0.002		
SIX2	3.26E-09	TCF4	2.61E-12	GTF2I	9.10E-16	CTNNB1	1.40E-06			GTF2B	1.79E-05	SNAI2	0.002		
GARS	7.95E-09	BRPF3	2.92E-12	GRHL2	1.45E-15	FLI1	1.52E-06			HMGN3	1.79E-05	NIPBL	0.002		
FOXO1	8.72E-09	CDK9	5.51E-12	BRD3	2.34E-15	YAP1	2.19E-06			CHD8	1.99E-05	PSIP1	0.002		
SMAD3	9.35E-09	GRHL2	5.66E-12	SNAI2	2.77E-15	JMJD1C	2.25E-06			TCF3	2.29E-05	CREBBP	0.003		
HID1	1.03E-08	NIPBL	7.32E-12	PPARG	3.28E-15	TCF3	2.39E-06			NIPBL	3.04E-05	HIF1A	0.003		
KDM4A	1.30E-08	RAG2	7.37E-12	JARID2	6.42E-15	RAG2	2.66E-06			TET2	3.09E-05	CTBP2	0.003		
SUMO2	1.83E-08	HIF1A	9.48E-12	NR1H3	1.52E-14	CREBBP	2.77E-06			LEO1	3.16E-05	DAXX	0.004		
MYOD1	2.28E-08	SIX2	1.66E-11	BRPF3	2.35E-14	NR1H4	2.91E-06			AFF4	3.20E-05	PIAS1	0.004		
KLF5	3.29E-08	PPARG	2.28E-11	MEF2B	3.63E-14	TCF4	2.91E-06			GTF2I	4.26E-05	KDM4A	0.004		
FLI1	5.22E-08	PMEPA1	2.33E-11	YAP1	4.45E-14	PDX1	6.40E-06			NFKB1	4.64E-05	RREB1	0.005		
TET2	5.84E-08	PBX1	6.12E-11	CTNNB1	4.45E-14	ASXL1	6.69E-06			VDR	4.84E-05	ARNTL	0.006		
ERG	6.32E-08	ASXL1	8.10E-11	ERG	4.81E-14	PPARG	7.96E-06			SNAI2	5.39E-05	SIX2	0.006		
AHR	7.05E-08	NKX2-1	8.93E-11	HMGN3	5.21E-14	PBX1	1.07E-05			LYL1	5.77E-05	GATA6	0.009		
TFAP2C	7.12E-08	PGR	8.93E-11	BCL6	1.10E-13	MYH11	1.29E-05			PMEPA1	6.91E-05	PPARD	0.011		
PSIP1	1.01E-07	CDK8	1.15E-10	KDM6B	1.68E-13	OTX2	1.68E-05			ESR2	7.16E-05	KDM3A	0.012		
GTF2I	1.43E-07	E2F3	1.71E-10	LMO2	1.94E-13	LMNA	1.80E-05			NR1H3	8.82E-05	GRHL2	0.012		
DAXX	1.43E-07	KLF5	2.07E-10	JMJD1C	2.07E-13	ZNF92	2.20E-05			MECOM	1.45E-04	PDX1	0.014		
PPARG	1.48E-07	BCL6	2.08E-10	KLF5	2.94E-13	GTF2I	2.33E-05			DPEP3	1.48E-04	ZFP42	0.014		
BRPF3	1.48E-07	TP63	2.14E-10	SMAD4	4.88E-13	MYOD1	2.33E-05			HOXA4	1.57E-04	RAG2	0.015		

6.3 List of Abbreviations

%	percent
° C	degree Celsius
µg	microgram
µl	microliter
µM	micromolar
5mC	methylcytosine
5hmC	hydroxy-methylcytosine
5fC	formy-cytosine
5cC	carboxy-cytosine
A	adenine
ATAC-seq	Assay for transposase-accessible chromatin followed by sequencing
ATP	adenosintriphosphate
bp	base pair
C	cytosine
Cat.Number	catalogue number
CEL	Cellartis protocol
DAR	differentially accessible region
ddH2O	double distilled water
DE	definitive endoderm
DEG	differentially expressed gene
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
dNTPs	deoxyribonucleotides
G	guanine

H2O	water
HAY	HAY protocol
HLC	hepatocyte-like cell
HVG	highly variable genes
iPSC	induced pluripotent stem cell
l	liter
LCM	laser capture microdissection
M	molar
mg	milligram
ml	milliliter
mM	millimolar
mRNA	messenger RNA
NGS	Next generation sequencing
PHH	primary human hepatocyte
RNA	ribonucleic acid
RNase	ribonuclease
RNA-seq	RNA sequencing
RP-HPLC	reverse phase high-performance liquid chromatography
RRBS	Reduced representation bisulfite sequencing
scRNA-seq	single cell RNA sequencing
Sup.	supplementary
T	thymine
TF	transcription factor
TFBS	TF binding sites
U	unit

6.4 List of Figures

6.5 List of Tables

7. Curriculum Vitae

First Name: Kathrin
Surname: Kattler
Date of Birth: 17.02.1991
Country: Germany
Email address: Kathrin.Kattler@googlemail.com

EDUCATION

10/2015 – 2019	Ph.D. student
	Supervisor: Prof. Dr. Jörn Walter
	Department of Epigenetics, Saarland University, Germany
	Topic: <u>Epigenetic characterization of (zonal) hepatocyte subpopulations in context of human health and complex metabolic diseases</u>
10/2013 - 10/2015	Master of Science (1,1) in human and molecular biology
	Supervisor: Prof. Dr. Jörn Walter
	Department of Epigenetics, Saarland University, Germany
	Topic: <u>Genome-wide chromatin accessibility studies in mammalian cells</u>
10/2010- 07/2013	Bachelor of Science (1,2) in human and molecular biology
	Supervisor: Prof. Dr. Jörn Walter
	Department of Epigenetics, Saarland University, Germany
	Topic: <u>Creation of histone modification reference maps in the HepaRG cell line using ChIP-sequencing</u>
10/2009-09/2010	Pharmaceutical studies, Saarland University, Germany
06/2009	Abitur (1,6), Albert-Einstein Gymnasium, Völklingen, Germany

FURTHER ACTIVITIES

2013 - 2015	Student research assistant, Saarland University, Germany
2014	ZHMB student's mini research grant
2012 - 2014	Students' council of the biology department
2012 – 2013	Shop assistant, Weltbild Plus Medienvertrieb

FURTHER QUALIFICATIONS

Languages	German (native language)
	English (fluent spoken and written)
	French (basic)

	Spanish (basic)
Programming skills	R (advanced)
	Python (basic)

TEACHING

Grundpraktikum Genetik, Biologie Bachelor (Leitung)

Seminar zum Grundpraktikum Genetik, Biologie Bachelor (Leitung)

Aufbaupraktikum I Genetik, Biologie Bachelor (Praktikumsbetreuung)

Advanced Modul I Genetik, Biologie Master (Praktikumsbetreuung)

Seminar Bio- und Gentechnologie, Lehramt (Organisation und Leitung)

Aufbaupraktikum Genetik für Bioinformatiker (Praktikumsbetreuung)

7.1 List of publications with contributions

Research articles

First author and second author [* shared first author]

Brosch*, M., **Kattler*, K.**, Herrmann, A., von Schönfels, W., Nordström, K., Seehofer, D., Damm, G., Becker, T., Zeissig, S., Nehring, S., et al. (2018). Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. **Nature Communications**.

> Contributions: Preparation of RRBS libraries on microdissected hepatocytes, data analysis, figure panels (except Figure 3 b, Sup. Figures 1, 2 a, 9 and 10), Sup. Tables, manuscript of methods and results, contributions to introduction and discussion.

Behjati Ardakani F, **Kattler K**, Nordström K, Gasparoni N, Gasparoni G, Fuchs S, Sinha A, Barann M, Ebert P, Fischer J, et al. (2018). Integrative analysis of single-cell expression data reveals distinct regulatory states in bidirectional promoters. **Epigenetics Chromatin**.

> Contributions: Establishment of scRNA-seq data and preparation of HepG2 scRNA-seq libraries, methods section on scRNA-seq library preparation and processing, critical review of the manuscript.

Contributing author

Fischer, AU., Müller, NIC., Deller, T., Del Turco, D., Fisch, JO., Griesemer, D., **Kattler, K.**, Maraslioglu, A., Roemer, V., Xu-Friedman, MA., et al. (2019). GABA is a Modulator, Rather than a Classical Transmitter, in the MNTB-LSO Sound Localization Circuit. **J Physiol.**

> Contributions: RNA-seq library preparation of microdissected samples, consultation of NGS data analysis.

Kehl, T., Schneider, L., **Kattler, K.**, Stöckel, D., Wegert, J., Gerstner, N., Ludwig, N., Distler, U., Schick, M., Keller, U., et al. (2018a). REGGAE: a novel approach for the identification of key transcriptional regulators. **Bioinformatics**.

Contributions: ChIP-seq library preparation, critical review of the manuscript.

Kehl, T., Schneider, L., **Kattler, K.**, Stöckel, D., Wegert, J., Gerstner, N., Ludwig, N., Distler, U., Tenzer, S., Gessler, M., et al. (2018b). The role of TCF3 as potential master regulator in blastemal Wilms tumors. **Int. J. Cancer**.

Contributions: ChIP-seq library preparation, method section on ChIP-seq data preparation and processing.

Salhab, A., Nordström, K., Gasparoni, G., **Kattler, K.**, Ebert, P., Ramirez, F., Arrigoni, L., Müller, F., Polansky, J.K., Cadenas, C., et al. (2018). A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. **Genome Biol**.

Contributions: ChIP-seq library preparation, method section on ChIP-seq data preparation and processing, critical review of the manuscript.

Nordström KJ, Schmidt F, Gasparoni N, Salhab A, Gasparoni G, **Kattler K**, Müller F, Ebert P, Costa IG, Pfeifer N, et al. (2019). Unique and assay specific features of NOMe-, ATAC- and DNase I-seq data. Nucleic Acids Res. [accepted].

Contributions: ATAC-seq library preparation, method section on ATAC-seq data preparation, critical review of the manuscript.

Review articles

Godoy, P., Schmidt-Heck, W., Hellwig, B., Nell, P., Feuerborn, D., Rahnenführer, J., **Kattler, K.**, Walter, J., Blüthgen, N., and Hengstler, J.G. (2018). Assessment of stem cell differentiation based on genome-wide expression profiles. **Philos. Trans. R. Soc. Lond., B, Biol. Sci**. *373*.

Contributions: Critical review of the manuscript.

7.2 List of conference talks, poster presentations and hackathons

2019 – LifeTime Opening Conference, Berlin, Germany: "Spatial transcriptomic and epigenomic maps of human liver: blueprints for projection of single cell data into hepatic pseudospace" [selected talk and poster].

2019 – SCOG Hackathon, Munich, Germany: "Computational single cell biology" [hackathon].

2019 – 3rd annual EpiTriO Meeting, Munich, Germany: "Signatures of liver zonation in health and metabolic disease" [talk].

2018 – Eurolife Winter School on Genome Stability: Structure and Function of Genomes in Homeostasis and Disease, Obergurgl, Austria: "Characterization of *in vitro* differentiated hepatocyte-like cells on single cell level" [short talk].

2018 – SCOG workshop: Advances in single cell epigenomics, Überherrn, Germany: "Integrated epigenomic map of the human liver" [talk].

2018 – SCOG Steering Committee Meeting, Munich, Germany: "There and back again – single cell data in spatial tissue context of human liver" [talk].

2018 – 2nd annual EpiTriO Meeting, Paris, France: "Signatures of liver zonation in health and metabolic disease" [talk].

2017 – annual IHEC Science Days, Berlin, Germany: "NexT-WGBS – A new tagmentation based approach for whole genome bisulfite sequencing" [poster].

2017 – IMB Conference Gene Regulation by the numbers, Mainz, Germany: "Zonal hepatocyte subpopulations in context of human metabolic diseases" [poster].