




2019

Roles Of Euchromatin And Heterochromatin In Hepatocyte Maturation And Liver Fibrosis

Jessica Mae Grindheim

University of Pennsylvania, jmgrindheim@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Cell Biology Commons](#), and the [Developmental Biology Commons](#)

Recommended Citation

Grindheim, Jessica Mae, "Roles Of Euchromatin And Heterochromatin In Hepatocyte Maturation And Liver Fibrosis" (2019).

Publicly Accessible Penn Dissertations. 3258.

<https://repository.upenn.edu/edissertations/3258>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/3258>

For more information, please contact repository@pobox.upenn.edu.

Roles Of Euchromatin And Heterochromatin In Hepatocyte Maturation And Liver Fibrosis

Abstract

Liver transplantation is the main treatment for acute liver failure patients; however, there is an insufficient supply of donor livers. Since transplanting hepatocytes, the main liver cell type, provides therapeutic effect and can be a bridge to transplant or recovery, scientists are working on generating replacement hepatocytes from stem cells and other cell types through reprogramming protocols. Currently, replacement hepatocytes recapitulate a subset of natural hepatocyte features, yet are still in an immature state, as they have not silenced all immature hepatocyte genes and activated all mature hepatocyte genes. Consequently, replacement hepatocytes do not perform as well as natural hepatocytes in transplant experiments. Despite these shortcomings, relatively little is known about how natural hepatic maturation is regulated, particularly at the chromatin level.

We discovered extensive chromatin dynamics during hepatic postnatal maturation, including changes in H3K9me₃-marked and H3K27me₃-marked heterochromatin, and transcription. Heterochromatin is of particular interest, as we found that it guards cell identity by repressing lineage-inappropriate or temporally-inappropriate genes. We further classified H3K9me₃- and H3K27me₃-marked chromatin by compaction state with a novel assay, termed srHC-seq. In postnatal hepatocyte maturation H3K27me₃-marked heterochromatin represses early maturation genes, late maturation genes, and alternative lineage genes to both regulate timing of hepatic maturation and repress alternate fates. Significantly, we identify a euchromatic H3K27me₃⁺ promoter signature that predicts which H3K27me₃-marked genes will derepress in response ablation of the enzymes that deposit H3K27me₃. Disruption of either H3K9me₃- or H3K27me₃-marked chromatin leads to liver damage, and in the case of H3K27me₃ this is likely due to the aberrant derepression of genes associated with fibrosis that normally have a euchromatic H3K27me₃⁺ promoter signature. Our results emphasize the role of heterochromatin in regulating liver development, maturation, and fibrosis, and highlight the need to identify factors controlling heterochromatin formation and breakdown, both for the purposes of enhancing in vitro hepatic maturation and for understanding factors which predispose humans to disease.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Kenneth S. Zaret

Keywords

chromatin, hepatocyte, liver, maturation

Subject Categories

Cell Biology | Developmental Biology

**ROLES OF EUCHROMATIN AND HETEROCHROMATIN IN
HEPATOCYTE MATURATION AND LIVER FIBROSIS**

Jessica Mae Grindheim

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Kenneth S. Zaret, Ph.D.
Joseph Leidy Professor of Cell and Developmental Biology

Graduate Group Chairperson

Daniel S. Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee

Ben Z. Stanger, M.D., Ph.D., Professor of Medicine

Stephen A. Liebhaber, MD, Professor of Genetics

Paul J Gadue, Ph.D., Associate Professor of Pathology and Laboratory Medicine

Maya Capelson, Ph.D., Assistant Professor of Cell and Developmental Biology

ROLES OF EUCHROMATIN AND HETEROCHROMATIN IN HEPATOCYTE MATURATION
AND LIVER FIBROSIS

COPYRIGHT 2019

Jessica Mae Grindheim

ABSTRACT

REGULATION OF HEPATIC DEVELOPMENT AND MATURATION

BY TWO CLASSES OF HETEROCHROMATIN

Jessica Mae Grindheim

Kenneth S Zaret

Liver transplantation is the main treatment for acute liver failure patients; however, there is an insufficient supply of donor livers. Since transplanting hepatocytes, the main liver cell type, provides therapeutic effect and can be a bridge to transplant or recovery, scientists are working on generating replacement hepatocytes from stem cells and other cell types through reprogramming protocols. Currently, replacement hepatocytes recapitulate a subset of natural hepatocyte features, yet are still in an immature state, as they have not silenced all immature hepatocyte genes and activated all mature hepatocyte genes. Consequently, replacement hepatocytes do not perform as well as natural hepatocytes in transplant experiments. Despite these shortcomings, relatively little is known about how natural hepatic maturation is regulated, particularly at the chromatin level.

We discovered extensive chromatin dynamics during hepatic postnatal maturation, including changes in H3K9me3-marked and H3K27me3-marked heterochromatin, and transcription. Heterochromatin is of particular interest, as we found that it guards cell identity by repressing lineage-inappropriate or temporally-inappropriate genes. We further classified H3K9me3- and H3K27me3-marked chromatin by compaction state with a novel

assay, termed srHC-seq. In postnatal hepatocyte maturation H3K27me3-marked heterochromatin represses early maturation genes, late maturation genes, and alternative lineage genes to both regulate timing of hepatic maturation and repress alternate fates. Significantly, we identify a euchromatic H3K27me3⁺ promoter signature that predicts which H3K27me3-marked genes will derepress in response ablation of the enzymes that deposit H3K27me3. Disruption of either H3K9me3- or H3K27me3-marked chromatin leads to liver damage, and in the case of H3K27me3 this is likely due to the aberrant derepression of genes associated with fibrosis that normally have a euchromatic H3K27me3⁺ promoter signature. Our results emphasize the role of heterochromatin in regulating liver development, maturation, and fibrosis, and highlight the need to identify factors controlling heterochromatin formation and breakdown, both for the purposes of enhancing in vitro hepatic maturation and for understanding factors which predispose humans to disease.

Table of Contents

ABSTRACT	iii
Table of Contents	v
Chapter 1: Introduction	1
1.1. A clinical need for new sources of human hepatocytes	1
1.2. Lack of terminal maturation in artificially generated hepatocytes	2
1.3. Regulation of postnatal hepatic maturation	5
1.4. Guarding cell identity with Polycomb-marked chromatin	8
1.5. Polycomb repression in endoderm lineage development	19
1.6. Figures	21
Figure 1: Liver transplant survival and demand in the US	21
Figure 2: PRC1 and 2 complexes.....	22
Figure 3: Multiple mechanisms of PcG repression	24
1.7. Tables	25
1.8. References	26
Chapter 2: PRC2 proteins EZH1/2 regulate timely postnatal hepatocyte maturation and fibrosis by repression of euchromatic promoters.....	37
2.1. Preface	37
2.2. Respective Contributions	38
2.3. Abstract	39
2.4. Graphical abstract	40
2.5. Introduction	41
2.6. Results	42
Postnatal hepatic maturation involves differential expression of thousands of genes	42
P14 and M2 H3K27me3 states correlate with postnatal hepatic maturation	43
EZH1 and EZH2 restrain premature postnatal hepatic maturation	44
Non-hepatocyte lineage genes are derepressed in P14 <i>Ezh1/2</i> hepatocytes	45
A subset of maturation and alternative lineage genes are repressed by EZH1/2	45
srHC-seq reveals that <i>Ezh1/2</i> sensitive genes have euchromatic and bivalently-marked promoters	47
Chronic <i>Ezh1/2</i> loss leads to liver damage.....	49
Genes involved in liver fibrosis are primed by euchromatic H3K27me3+ promoters	50
2.7. Discussion	51
2.8. Main Figures	54
Figure 1. Postnatal hepatic maturation during the P14 to M2 transition involves differential expression of thousands of genes and H3K27me3 dynamics.	54
Figure 2. PRC2 proteins EZH1/2 restrain postnatal hepatic maturation.	55
Figure 3. Stable global chromatin compaction in maturation and P14 <i>Ezh1/2</i>	57
Figure 4. srHC-seq reveals that <i>Ezh1/2</i> sensitive genes have euchromatic and bivalently-marked promoters.....	59
Figure 5. <i>Ezh1/2</i> loss leads to chronic liver damage.....	61
Figure 6. PRC2 represses liver fibrosis signature genes.....	62

2.9. Supplementary Figures	63
2.10. Methods.....	74
2.11. Supplementary Tables.....	79
2.12. References.....	80
Chapter 3: H39me3-heterochromatin loss at protein-coding genes enables developmental lineage specification	86
3.1. Preface	86
3.2. Respective Contributions	87
3.3. Abstract	87
3.4. Main Text	88
3.5. Figures	94
Figure 1: Chromatin compaction and H3K9me3 landscape upon germ-layers differentiation.	94
Figure 2: Loss of srHC and H3K9me3 correlates with gene expression of hepatic-specific markers upon differentiation.	95
Figure 3: Setdb1 mutant hepatoblasts upregulate lineage non-specific genes.	96
Figure 4: TKO mutant cells lose hepatic identity and show developmental phenotypes associated with decreased H3K9me3 and srHC levels.....	97
3.6. Supplementary Figures	98
3.7. Methods	125
3.8. Tables	151
3.9. References.....	157
Chapter 4: Perspectives and Future Directions	161
4.1. Sonication-resistant heterochromatin sequencing as a method to map compacted heterochromatin and open euchromatin	161
4.2. PRC2/H3K27me3 function at euchromatic promoters	165
4.3. PRC2/H3K27me3 involvement in postnatal hepatocyte identity.....	166
4.4. Implications for artificially derived hepatocytes	168
4.5. PRC2/H3K27me3 involvement in liver fibrosis	171
4.5. References.....	173
Appendix A: Murine hepatocyte isolation by liver perfusion	180
I. Objective.....	180
II. Key Reagents	181
III. Protocol.....	183
IV. Results	186
V. Figures	187
Figure 1: Isolation of murine hepatocytes	187
VI. References.....	188
Appendix B: Sonication Resistant Heterochromatin Sequencing, a method for mapping sonication-resistant heterochromatin.....	189
I. Objective.....	189
II. Background.....	190
III. Key Reagents	191
IV. Protocol	193
V. Conclusions	203
VI. Figures.....	207

Figure 1: Fractionation of high and low molecular weight DNA for srHC-seq	207
Figure 2: High reproducibility of srHC-seq data	208
Figure 3: srHC-seq heterochromatin and euchromatin domain, promoter, and gene body calling	210
Figure 4: AB compartments	212
VII. Bioinformatics methods	212
VIII. References.....	213
Appendix C: Cut and Run, a method for mapping proteins on chromatin	215
I. Objective	215
II. Background.....	215
III. Key Reagents	218
IV. Protocol	220
V. Conclusions	224
VI. Figures.....	226
Figure 1: Cut and Run schematic	227
Figure 2: Successful Cut and Run.....	229
VII. References.....	230
Appendix D: Chromatin-associated RNA-seq, a method for isolating RNAs bound to chromatin	231
I. Objective	231
II. Protocol.....	232
III. Figures	234
Figure 1: Characterization of chromatin-associated RNA-seq samples.....	234
IV. References.....	235
Appendix E: qPCR primer design and testing	236
I. Objective	236
II. Background.....	236
III. RTqPCR.....	237
IV. ChIP-qPCR	239
V. Primer Assessment	240
VI. Figures.....	243
Figure 1: Assessing primer efficacy	243
VII. Verified <i>mus musculus</i> RTqPCR primers	244

Chapter 1: Introduction

1.1. A clinical need for new sources of human hepatocytes

Liver failure results in a constellation of complications including impaired blood coagulation, increased risk of multiple organ failure, sepsis, and, prior to the advent of liver transplantation in 1960s, was often fatal (Durand and Valla, 2005; Hurst, 2012). While initial mortality rates in liver transplant recipients were high, advances in surgical techniques, immunosuppressive treatments, and anti-microbial agents have made liver transplantation the standard clinical treatment for both acute and chronic liver failure. Modern 1 and 5 year survival rates in the US across age groups are above 88% and 67%, respectively (Neuberger, 2016; NIDDK, 2017; OPTN, 2019a) (Figure 1). However, these improvements in liver transplantation are tempered by an increasing demand for transplants that surpasses the pool of available donors. The vast majority of transplants originate from deceased donors and the median wait time to deceased donor transplant has increased 1.5- to 2-fold in adult age groups between 2003 and 2014 with only 46-73% of patients receiving deceased donor transplants 2 years post listing (OPTN, 2019c, 2019b, 2019d) (Figure 1).

Given the shortage of donor livers, there is ongoing research into alternative therapies for treatment or as a bridge until the patient's liver recovers or a transplant becomes available. One option that has been tested is hepatocyte transplantation, or the transplantation of only one of the main resident liver cells types instead of the whole liver or liver lobes, which is considered much less surgically invasive than whole liver

transplant. The main source of hepatocytes for hepatocyte transplantation is livers that were deemed unsuitable for whole organ transplantation for reasons such as steatosis-related viability complications or ischemic damage (Donato et al., 2006; Fox et al., 1998; Sagias et al., 2010; Strom et al., 1997). Hepatocyte transplantation has primarily been tested in patients with congenital metabolic disorders, as the defects are primarily limited to the hepatocyte cell type (Table 1). In many of these cases, there are variable levels of improvement in liver function for a limited time, which is generally attributed to engraftment level, persistence, lack of graft expansion, and the inability to track for graft rejection in real time, which is needed to adjust the immunosuppressant regimen. However, these methods are being improved, such as in one recent study showing that pre-operative liver radiation can increase engraftment and create an environment where transplanted cells can expand (Soltys et al., 2017). While in need of improvements, hepatocyte transplantation is a promising cell therapy alternative to liver transplant, for both congenital metabolic syndromes and liver failure.

1.2. Lack of terminal maturation in artificially generated hepatocytes

As hepatocyte transplantation proof-of-principle studies have proved useful in providing a bridge to transplant time (Table 1) and hepatocyte cell therapy protocols are likely to see improvement in engraftment and graft maintenance, there is great interest in generating hepatocytes from non-hepatocyte cell types. Autologously sourced cells avoid both the search for a suitable liver and one of the largest barriers to successful transplantation, graft rejection due to immunoincompatibility. In the case of patients with congenital metabolic disorders, the defect could be corrected by genome-editing technologies and still circumvent immune rejection problems. Alternatively, replacement hepatocytes could

be generated, classified with respect to a pre-assessed or broad acceptability to different immune system backgrounds, and banked for future use.

One source of artificially derived hepatocytes for transplantation is pluripotent stem cell-derived hepatocytes, or PSC-hepatocytes. Pluripotent stem cells can be obtained from either embryonic stem cells (Evans and Kaufman, 1981; Martin, 1981) or by direct reprogramming of somatic cells to a pluripotent state (Takahashi and Yamanaka, 2006). While protocols are constantly evolving, many of details are conserved. The first step involves derivation of definitive endoderm with BMP4 signaling and is highly efficient (Gouon-Evans et al., 2006; Yiangou et al., 2018), followed by generation of hepatoblast-like cells, or embryonic hepatic precursor cells, by blocking Activin and mimicking natural cardiac FGF and mesenchymal BMP signals to the endoderm (Jung et al., 1999; Rossi et al., 2001). To further differentiate hepatoblast-like cells to hepatocyte-like cells, many protocols treat with hepatic growth factor and oncostatin M and have extended growth of up to a month in these conditions to improve maturation (Agarwal et al., 2008; Brolén et al., 2010; Cai et al., 2007; Chen et al., 2012c; Hannan et al., 2013; Mallanna and Duncan, 2013; Siller et al., 2015; Si-Tayeb et al., 2010; Song et al., 2009; Touboul et al., 2010). A related technique involves partial pluripotency reprogramming in combination with treatment with small molecules to generate cells with hepatocyte phenotype, which is interesting for the capability for cellular expansion while theoretically reducing the tumor formation risk from undifferentiated cells (Zhu et al., 2014). The resulting hepatocyte-like cells can have binucleate morphology, express liver transcription factors and functional proteins such as HNF4 α , CEBP α , PROX1, HNF6, GATA4, ALB, AAT, and CYP3A4, and can perform liver functions such as glycogen synthesis, albumin secretion, urea production, LDL uptake, and some limited cytochrome P450 activity. However, these PSC-hepatocytes are not considered equivalent to native adult hepatocytes, as they often

still express immature liver genes, such as *Afp*, and fail to express mature genes to the appropriately high levels, such as *Alb* and CYP P450 enzymes (Agarwal et al., 2008; Chen et al., 2012c; Duan et al., 2010; Roelandt et al., 2013; Si-Tayeb et al., 2010; Song et al., 2009; Zhu et al., 2014). Additionally, hepatocyte-like cells can express CK19, a marker of the related cholangiocyte lineage (Cai et al., 2007) and secrete albumin, a key component of blood serum, at much lower levels than native hepatocytes. These results indicate that current PSC-hepatocytes fail to induce terminal differentiation.

Alternatively, hepatocytes for transplantation can also be generated from direct reprogramming of somatic cells to hepatocytes, here referred to as induced hepatocytes or iHeps. The Hui and Deng groups have described generation of human iHeps, either by ectopic expression of the transcription factors *Hnf4*, *Hnf1 α* , and *Foxa3* (Huang et al., 2014) or ectopic expression of transcription factors *Hnf1 α* , *Hnf4 α* , *Hnf6* and the maturation factors *Atf5*, *Prox1*, and *Cepb α* (Du et al., 2014). In both cases, the iHeps perform some liver metabolic functions, such as inducible expression and activity of CYP3A4 and have expression profiles similar to freshly isolated adult hepatocytes. Additionally, both groups are working on methods to make iHeps expandable, which is key for scaling up to clinical useful levels. However, the iHeps have a broad spectrum of genes differentially expressed from primary hepatocytes and in transplant studies cannot match primary hepatocytes for albumin secretion, liver colonization, and rescue of liver damage (Du et al., 2014; Huang et al., 2014). Additionally, the Deng study tested *in vivo* engraftment and expansion in cells that had tumorigenic modifications which prevent clinical use (Du et al., 2014). Before clinical use, hiHeps will need to be made more expandable and have more mature hepatocyte functionality.

In summary, both PSC-hepatocytes and iHeps fail to fully recapitulate features of adult hepatocytes including expansion in the liver environment, transcriptomic profiles, replacing liver function, and rescuing liver damage. Fine tuning of these cells to exhibit more mature phenotypes will be necessary before use in clinical settings.

1.3. Regulation of postnatal hepatic maturation

Embryonic liver development involves the concerted regulation by various transcription factors, such as HNF4 α , GATA and FOXA family members, and paracrine signaling events (Duncan et al., 2009; Zaret and Grompe, 2008; Zorn, 2008). Numerous changes occur in liver after birth but before fully mature adulthood, including egress of hematopoietic cells (Brauner et al., 2001), physiological changes from diet, a dramatic increase in liver size (Leibing et al., 2018; Septer et al., 2012; Suzuki et al., 2019), hepatocyte polyploidization, microbial colonization and consequential exposure to microbial metabolites (Avior et al., 2015; Morelli, 2008), changes in membrane fluidity (Devi et al., 1992), and sexual maturation. Postnatal hepatic maturation is also reflected in gene expression changes, where fetal liver-specific genes are silenced and mature liver genes are induced (Spear et al., 2006). Here I discuss some postnatal maturation processes and what is known of their regulation.

Diet dramatically changes from primarily placental glucose, to high lipid content breastmilk at birth, and returns to a higher carbohydrate diet post-weaning. Consequently, hepatic carbohydrate and lipid metabolism vary respectively, with postnatal induction of gluconeogenesis, fatty acid β -oxidation, and *de novo* lipogenesis (Decaux et al., 1988; Perez-Castillo et al., 1987; Périchon and Bourre, 1995; Sekine et al., 2007). Ligands in milk are thought to bind and activate PPAR α , as it is known that murine neonate fatty acid

oxidative function is induced partly through postnatal PPAR α -dependent DNA demethylation and increased mRNA expression, with the demethylation also shown at β -oxidation genes during the fetal to adult transition in humans (Ehara et al., 2015). However, how these metabolic pathways are activated during maturation remains poorly understood.

Hepatic polyploidization, including both multiple nuclei and polyploid nuclei, occurs postnatally. In humans, late teenagers have a mostly diploid with some tetraploid profile that increases from 20-30% polyploid to greater than 50% polyploid in the elderly (Liu et al., 2003; Watanabe and Tanaka, 1982). Similarly, murine polyploidization occurs around weaning and increases with age (Duncan et al., 2010). The co-occurrence of weaning and polyploidization from cytokinesis failure in rodents led to studies showing that weaning-driven increases in blood glucose leads to increased blood insulin, which stimulates hepatic insulin-PI3K-AKT signaling, and in turn AKT signaling is sufficient to decrease the frequency of cytokinesis events (Celton-Morizur et al., 2009, 2010; Duncan et al., 2010; Margall-Ducos et al., 2007). Deletion of various cell cycle factors or P53 perturbations effects postnatal hepatocyte polyploidization in mice (Chen et al., 2012a; Chipchase et al., 2003; Hsu et al., 2016; Li et al., 2013, 2013; Mayhew et al., 2005; Pandit et al., 2012). While many studies have found cell cycle-associated factors regulate hepatocyte ploidy (Wang et al., 2017), what other upstream hepatic maturation factors besides diet and insulin control these cell cycle factors remains unclear.

The control of liver zonation is one of the better characterized maturation phenomena (Perugorria et al., 2019). Liver zonation involves the unequal distribution of function and enzymes in hepatocytes across the liver lobule with respect to the pericentral and periportal blood vessels. Zonated functions include metabolism of carbohydrates,

ammonia, bile acids, and drugs, as well as transport and secretion enzymes (Jungermann and Kietzmann, 1996, 2000). While early studies showed that some zoned functions were related to oxygen, hormone, or nutrient gradients (Gebhardt and Gaunitz, 1997; Oinonen and Lindros, 1998), it was not until 2006 that a transcription factor was identified that controlled zonation. A WNT signaling gradient from liver sinusoidal endothelial cells (Leibing et al., 2018) across the portocentral axis leads to differential stabilization of β -catenin and TCF-LEF family transcription factors and consequential transcription of a subset of zoned genes, in particular ammonia and glutamine metabolism (Benhamouche et al., 2006). Stabilized TCF/LEF becomes a binding partner of the key liver transcription factor HNF4 α , which changes binding patterns to both activate and repress specific zoned genes, independent of CMYC, which is also known to be activated by Wnt/ β -catenin signaling (Burke et al., 2009; Colletti et al., 2009). Despite what is already known about Wnt signaling in the maturing liver, given the complexity of Wnt- β -catenin signaling, with 19 Wnt ligands, more than 15 receptors, non-canonical Wnt signaling, and pathway crosstalk, and given the failure of clinical trials involving Wnt in various liver diseases, a more in-depth understanding of liver Wnt- β -catenin signaling in liver maturation would likely prove useful in designing replacement hepatocytes

There are some known examples of postnatal hepatic maturation regulation that occur at the transcriptional and post-transcriptional levels. ZHX2 and ZBTB20 are direct transcriptional repressors in mice of *Afp*, which encodes a serum protein highly expressed in embryos that is repressed postnatally (Morford et al., 2007; Perincheri et al., 2005; Xie et al., 2008). ZHX2 may be used more generally, as it also represses mouse major urinary proteins and some sexually dimorphically expressed CYP P450 enzymes (Creasy et al., 2016; Jiang et al., 2017). Key liver transcription factor HNF4 α binding to some

transcriptional start sites of some maturation genes is promoted by TAF4-TAF12 two weeks after birth in mice (Alpern et al., 2014). HNF4 α is also one of several liver transcription factors that set up and maintain the liver program by creating a cross-regulatory and auto-regulatory network which increases in complexity between embryonic and adult stages, though how the network is refined postnatally is unknown (Kyrmizi et al., 2006; Odom et al., 2006). In mice and humans 20% of the neonatal-to-adult splice isoform switches are attributed to the induction and consequential function of the conserved splicing factor ESRP2, thus facilitating maturation of hepatocytes (Bhate et al., 2015). The scattering of known transcriptional and post-transcriptional repressors that have been identified underline how incomplete a picture we have a factors refining hepatocyte cell identity postnatally.

The wealth of knowledge of embryonic liver development and adult liver biology has informed protocols that aim to artificially generate hepatocytes for clinical use. In the case of PSC-hepatocytes, progenitor cell-like stages are generated with high efficiency but fail to achieve fully mature hepatocyte function. In the case of iHeps, which jump directly from one cell type to the hepatocyte fate, there is a similar failure to obtain a mature hepatocyte phenotype. The coincidence of the failure to obtain fully mature hepatocytes and the gaps in the postnatal hepatic maturation field indicate that more attention should be paid to hepatic maturation.

1.4. Guarding cell identity with Polycomb-marked chromatin

The 3 billion base pairs of DNA in human genome serves as an instruction manual for the development and maintenance of hundreds of different cell types with distinct functions and morphologies, which cooperate to form a functioning multicellular organism. This plethora of outcomes from the same DNA sequence is possible through differential activation and silencing of genes in each cell type in a temporally- and context-specific

manner. In order to achieve the differential gene regulation necessary, the genome of eukaryotic cells is organized into multiple hierarchical levels of chromatin (Probst et al., 2009). On the most basic level, DNA is organized into the nucleosome, a repeating unit in which approximately 147 bp of DNA is wrapped around a nucleosome, consisting of histone proteins, with linker DNA between adjacent nucleosomes. The series of nucleosomes form a chromatin fiber that coils into higher-order states, including chromatin loops and topologically associate domains (TADs) (Furlong and Levine, 2018). Histones have layers of possible regulation, including variants and posttranslational modifications (PTMs) on the tails that extend from the globular domain (Sitbon et al., 2017). PTM reader proteins, histone chaperones, chromatin remodelers and modifying enzymes, DNA-modifying enzymes, RNA, and chromatin structure shape a versatile chromatin landscape that regulates access of TFs to chromatin and thus allows the same DNA sequence to generate diverse cell types that respond to different inputs, which consequently defines the ability of a cell to have a particular phenotype and to respond to development, environmental changes, and aging Ardehali, M.B., Anselmo, A., Cochrane, J.C., Kundu, S., Sadreyev, R.I., and Kingston, R.E. (2017). Polycomb Repressive Complex 2 Methylates Elongin A to Regulate Transcription. *Mol. Cell* 68, 872-884.e6.

Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14926–14931.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* 32, 29–41.

Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* 68, 1023-1037.e15.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12, 123–135.

Bergmann, C., Brandt, A., Merlevede, B., Hallenberger, L., Dees, C., Wohlfahrt, T., Pötter, S., Zhang, Y., Chen, C.-W., Mallano, T., et al. (2018). The histone demethylase Jumonji domain-containing protein 3 (JMJD3) regulates fibroblast activation in systemic sclerosis. *Ann. Rheum. Dis.* 77, 150–158.

Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O’Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., Costello, J.F., et al. (2011). Characterization of the contradictory chromatin signatures at the 3’ exons of zinc finger genes. *PLoS ONE* 6, e17121.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Believeau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.

Borkham-Kamphorst, E., Herrmann, J., Stoll, D., Treptau, J., Gressner, A.M., and Weiskirchen, R. (2004). Dominant-negative soluble PDGF-beta receptor inhibits hepatic stellate cell activation and attenuates liver fibrosis. *Lab. Invest.* 84, 766–777.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes*

Dev. 20, 1123–1136.

Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* 412, 651–655.

Brown, S.W. (1966). Heterochromatin. *Science* 151, 417–425.

Cao, S., Yaqoob, U., Das, A., Shergill, U., Jagavelu, K., Huebert, R.C., Routray, C., Abdelmoneim, S., Vasdev, M., Leof, E., et al. (2010). Neuropilin-1 promotes cirrhosis of the rodent and human liver by enhancing PDGF/TGF-beta signaling in hepatic stellate cells. *J. Clin. Invest.* 120, 2379–2394.

Chen, H.-Z., Ouseph, M.M., Li, J., Pécot, T., Chokshi, V., Kent, L., Bae, S., Byrne, M., Duran, C., Comstock, G., et al. (2012). Canonical and atypical E2Fs regulate the mammalian endocycle. *Nat. Cell Biol.* 14, 1192–1202.

Chipchase, M.D., O'Neill, M., and Melton, D.W. (2003). Characterization of premature liver polyploidy in DNA repair (Ercc1)-deficient mice. *Hepatology* 38, 958–966.

Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. *Biochem. Soc. Trans.* 45, 193–205.

Di Bisceglie, A.M. (2000). Natural history of hepatitis C: its impact on clinical management. *Hepatology* 31, 1014–1018.

Fadloun, A., Eid, A., and Torres-Padilla, M.-E. (2013). Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions. *Curr. Top. Dev. Biol.* 104, 1–45.

Frenster, J.H., Allfrey, V.G., and Mirsky, A.E. (1963). REPRESSED AND ACTIVE CHROMATIN ISOLATED FROM INTERPHASE LYMPHOCYTES. *Proc. Natl. Acad. Sci. U.S.A.* 50, 1026–1032.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* 6, 784–791.

Ghirlando, R., and Felsenfeld, G. (2008). Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. *J. Mol. Biol.* 376, 1417–1425.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118, 555–566.

Grindheim, J.M., Nicetto, D., Donahue, G., and Zaret, K.S. (2019). PRC2 proteins EZH1 and EZH2 Regulate Timing of Postnatal Hepatocyte Maturation and Fibrosis by Repressing Gene Expression at Promoter Regions in Euchromatin in Mice. *Gastroenterology*.

Hashimoto, K., and Ogawa, Y. (2018). Epigenetic Switching and Neonatal Nutritional Environment. *Adv. Exp. Med. Biol.* 1012, 19–25.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491.

Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb Wiss Botanik* 762–818.

Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T.D., and Ge, K. (2007).

Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27

demethylases. *Proc. Natl. Acad. Sci. U.S.A.* *104*, 18439–18444.

Hsu, S.-H., Delgado, E.R., Otero, P.A., Teng, K.-Y., Kutay, H., Meehan, K.M., Moroney, J.B., Monga, J.K., Hand, N.J., Friedman, J.R., et al. (2016). MicroRNA-122 regulates polyploidization in the murine liver. *Hepatology* *64*, 599–615.

Innes, H.A., Hutchinson, S.J., Barclay, S., Cadzow, E., Dillon, J.F., Fraser, A., Goldberg, D.J., Mills, P.R., McDonald, S.A., Morris, J., et al. (2013). Quantifying the fraction of cirrhosis attributable to alcohol among chronic hepatitis C virus patients: implications for treatment cost-effectiveness. *Hepatology* *57*, 451–460.

Ishihara, S., Varma, R., and Schwartz, R.H. (2010). A new fractionation assay, based on the size of formaldehyde-crosslinked, mildly sheared chromatin, delineates the chromatin structure at promoter regions. *Nucleic Acids Res.* *38*, e124.

Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N.K., Pinello, L., Yuan, G.-C., Orkin, S.H., and Shivdasani, R.A. (2016). Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* *165*, 1389–1400.

Jeong, W.-I., Do, S.-H., Yun, H.-S., Song, B.-J., Kim, S.-J., Kwak, W.-J., Yoo, S.-E., Park, H.-Y., and Jeong, K.-S. (2004). Hypoxia potentiates transforming growth factor-beta expression of hepatocyte during the cirrhotic condition in rat liver. *Liver Int.* *24*, 658–668.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* *350*, aac4383.

Jung, Y.K., and Yim, H.J. (2017). Reversal of liver cirrhosis: current evidence and expectations. *Korean J. Intern. Med.* *32*, 213–228.

Kirmaz, C., Terzioglu, E., Topalak, O., Bayrak, P., Yilmaz, O., Ersoz, G., and Sebik, F. (2004). Serum transforming growth factor-beta1(TGF-beta1) in patients with cirrhosis, chronic hepatitis B and chronic hepatitis C [corrected]. *Eur. Cytokine Netw.* *15*, 112–116.

Kyrmizi, I., Hatzis, P., Katrakili, N., Tronche, F., Gonzalez, F.J., and Talianidis, I. (2006). Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev.* *20*, 2293–2305.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* *12*, 618–624.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* *13*, 1192–1200.

Lewis, P.H. (1947). New mutants. *Drosoph. Inf. Serv.*

Li, D., Cen, J., Chen, X., Conway, E.M., Ji, Y., and Hui, L. (2013). Hepatic loss of survivin impairs postnatal liver development and promotes expansion of hepatic progenitor cells in mice. *Hepatology* *58*, 2109–2121.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* *24*, 368–380.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG

islands. *Nature* 549, 287–291.

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.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13, 1196–1204.

Ma, W.-L., Lai, H.-C., Yeh, S., Cai, X., and Chang, C. (2014). Androgen receptor roles in hepatocellular carcinoma, fatty liver, cirrhosis and hepatitis. *Endocr. Relat. Cancer* 21, R165-182.

Mansour, A.A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., Krupalnik, V., Zerbib, M., Amann-Zalcenstein, D., Maza, I., et al. (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* 488, 409–413.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.

Marullo, F., Cesarini, E., Antonelli, L., Gregoret, F., Oliva, G., and Lanzuolo, C. (2016). Nucleoplasmic Lamin A/C and Polycomb group of proteins: An evolutionarily conserved interplay. *Nucleus* 7, 103–111.

Matsuoka, M., and Tsukamoto, H. (1990). Stimulation of hepatic lipocyte collagen production by Kupffer cell-derived transforming growth factor beta: implication for a pathogenetic role in alcoholic liver fibrogenesis. *Hepatology* 11, 599–605.

Mayhew, C.N., Bosco, E.E., Fox, S.R., Okaya, T., Tarapore, P., Schwemberger, S.J., Babcock, G.F., Lentsch, A.B., Fukasawa, K., and Knudsen, E.S. (2005). Liver-specific pRB loss results in ectopic cell cycle entry and aberrant ploidy. *Cancer Res.* 65, 4568–4577.

Nakatsukasa, H., Nagy, P., Evarts, R.P., Hsia, C.C., Marsden, E., and Thorgeirsson, S.S. (1990). Cellular distribution of transforming growth factor-beta 1 and procollagen types I, III, and IV transcripts in carbon tetrachloride-induced rat liver fibrosis. *J. Clin. Invest.* 85, 1833–1843.

Naveau, S., Perlemuter, G., and Balian, A. (2005). [Epidemiology and natural history of cirrhosis]. *Rev Prat* 55, 1527–1532.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* 2, 2006.0017.

Oksuz, O., Narendra, V., Lee, C.-H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* 70, 1149-1162.e5.

Pandit, S.K., Westendorp, B., Nantasanti, S., van Liere, E., Tooten, P.C.J., Cornelissen, P.W.A., Toussaint, M.J.M., Lamers, W.H., and de Bruin, A. (2012). E2F8 is essential for

polyploidization in mammalian cells. *Nat. Cell Biol.* *14*, 1181–1191.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* *464*, 306–310.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* *139*, 1290–1302.

Piacentini, L., Fanti, L., Berloco, M., Perrini, B., and Pimpinelli, S. (2003). Heterochromatin protein 1 (HP1) is associated with induced gene expression in *Drosophila* euchromatin. *J. Cell Biol.* *161*, 707–714.

Quiévryn, G., and Zhitkovich, A. (2000). Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteasome function. *Carcinogenesis* *21*, 1573–1580.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* *159*, 1665–1680.

Rao, V.K., Pal, A., and Taneja, R. (2017). A drive in SUVs: From development to disease. *Epigenetics* *12*, 177–186.

Riddle, N.C., Jung, Y.L., Gu, T., Alekseyenko, A.A., Asker, D., Gui, H., Kharchenko, P.V., Minoda, A., Plachetka, A., Schwartz, Y.B., et al. (2012). Enrichment of HP1a on *Drosophila* chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet.* *8*, e1002954.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* *42*, 330–341.

Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* *171*, 34–57.

Schuppan, D., Krebs, A., Bauer, M., and Hahn, E.G. (2003). Hepatitis C and liver fibrosis. *Cell Death Differ.* *10 Suppl 1*, S59-67.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* *139*, 1303–1314.

Slifer, E.H. (1942). A mutant stock of *Drosophila* with extra sex combs. *J. Exp. Zool.* *31–40*.

Snitow, M., Lu, M., Cheng, L., Zhou, S., and Morrissey, E.E. (2016). Ezh2 restricts the smooth muscle lineage during mouse lung mesothelial development. *Development* *143*, 3733–3741.

Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* *82*, 6470–6474.

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* *27*, 2663–2677.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* *151*, 994–1004.

Thieringer, F., Maass, T., Czochra, P., Klopčič, B., Conrad, I., Friebe, D., Schirmacher, P., Lohse, A.W., Blessing, M., Galle, P.R., et al. (2008). Spontaneous hepatic fibrosis in transgenic mice overexpressing PDGF-A. *Gene* *423*, 23–28.

Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad. Sci. U.S.A.* *113*, E744-753.

Torre, D., Lolli, F., Ciana, P., and Maggi, A. (2017). Sexual Dimorphism and Estrogen Action in Mouse Liver. *Adv. Exp. Med. Biol.* *1043*, 141–151.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* *28*, 1–13.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell* *19*, 381–391.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci. U.S.A.* *112*, 4672–4677.

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* *353*, 598–602.

Wells, R.G., Kruglov, E., and Dranoff, J.A. (2004). Autocrine release of TGF-beta by portal fibroblasts regulates cell growth. *FEBS Lett.* *559*, 107–110.

Xu, C.-R., Cole, P.A., Meyers, D.J., Kormish, J., Dent, S., and Zaret, K.S. (2011). Chromatin “prepattern” and histone modifiers in a fate choice for liver and pancreas. *Science* *332*, 963–966.

Xu, C.-R., Li, L.-C., Donahue, G., Ying, L., Zhang, Y.-W., Gadue, P., and Zaret, K.S. (2014). Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. *EMBO J.* *33*, 2157–2170.

Yanger, K., Zong, Y., Maggs, L.R., Shapira, S.N., Maddipati, R., Aiello, N.M., Thung, S.N., Wells, R.G., Greenbaum, L.E., and Stanger, B.Z. (2013). Robust cellular reprogramming occurs spontaneously during liver regeneration. *Genes Dev.* *27*, 719–724.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* *337*, 971–975.

Zhou, W.-C., Zhang, Q.-B., and Qiao, L. (2014). Pathogenesis of liver cirrhosis. *World J. Gastroenterol.* *20*, 7312–7324.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* *152*, 642–654. (Filipescu et al., 2014; Groth et al., 2007).

Physical compaction of the chromatin fiber is a major mechanism controlling transcription. Chromatin fiber compaction was originally defined cytologically by density of dye staining and light or dark appearance on electron micrographs, with light-staining open euchromatin and dark-staining compact heterochromatin compact (Brown, 1966; Heitz, 1928). Euchromatin is accessible to transcription factors and the transcriptional machinery and is associated with the majority of transcriptional activity (Li et al., 2007; Workman and Kingston, 1998). In comparison, physically compact chromatin, or heterochromatin, has more dense packing and is thus less accessible, and is associated with much less transcription. Constitutive heterochromatin is one class of heterochromatin that is found in most cell types at repetitive regions of the genome and functions to both repress the activity of transposable elements and to repress deleterious recombination between repetitive sequences (Becker et al., 2016). Constitutive heterochromatin is marked by the H3K9me3 histone modification, which does not have known inherent repressive properties, but instead is bound by several HP1 isoforms in mammals, which in turn oligomerize and recruit repressive histone modifiers, leading to chromatin compaction and heterochromatin spreading (Bannister et al., 2001; Becker et al., 2016; Canzio et al., 2011; Lachner et al., 2001). While more recent studies show that H3K9me3-marked heterochromatin can also be “facultative”, or highly dynamic during development, with lineage- and temporally-specific patterns and consequential repression of lineage-appropriate genes, facultative heterochromatin is more commonly associated with Polycomb group proteins (Becker et al., 2017; Nicetto et al., 2019; Trojer and Reinberg, 2007). Polycomb group (PcG) proteins were discovered in *Drosophila melanogaster* homeotic mutants as repressing Hox cluster genes and in mammals have been shown more generally to repress transcription of many developmental and lineage regulators (Boyer et al., 2006; Kassis et al., 2017; Lee et al., 2006; Lewis, 1978, 1947; Margueron

and Reinberg, 2011; Slifer, 1942). Of note, perturbation of PcG proteins and derepression of transcription factors frequently results in changes of cell fate (Bracken et al., 2006; Lewis, 1947; Slifer, 1942; Snitow et al., 2016; Xu et al., 2011), underlining the role of Polycomb-marked heterochromatin in cell fate control.

Biochemical studies show that PcG proteins function to maintain transcriptional repression by histone tail modification that promotes facultative heterochromatin formation. PcG proteins form 2 large complexes, Polycomb Repressive Complex 1 and 2 (PRC1 and 2) (Cao et al., 2002; Margueron and Reinberg, 2011), and the more recently identified Pho-repressive complex and Polycomb repressive deubiquitinase complex (Klymenko et al., 2006; Scheuermann et al., 2010). Understanding the complex recruitment, enzymatic activity, binding partners, and function in different chromatin contexts is required to understand how PcG proteins regulate cell identity

The core PRC2 complex is comprised of four proteins EED, SUZ12, RBBP46/48, and one of the two enhancer of zeste homologs, EZH1 or EZH2, with many additional proteins that generally have DNA and/or chromatin binding affinity and can promote enzymatic activity (Figure 2) (Holoch and Margueron, 2017). PRC2 is recruited to unmethylated CpG islands referred to as “nucleation sites” by PHF1, JARID2, and MTF2 (Landeira et al., 2010; Li et al., 2010, 2017; Oksuz et al., 2018; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009; Son et al., 2013) or to the PRC1-catalyzed histone mark, histone H2A lysine 119 monoubiquitin (H2AK119ub) (Blackledge et al., 2014; Cooper et al., 2014). Transcription or methylated cytosines can render CpG islands refractory to H3K27 trimethylation (Holoch and Margueron, 2017). After recruitment, the PRC2 complex catalyzes histone H3 lysine 27 trimethylation (H3K27me3) through the mutually exclusive histone methyltransferase subunits EZH1 or EZH2 (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002). Binding of PRC2 component EED to H3K27me3 through

its aromatic cage formed by WD-40 repeats leads to allosteric activation of H3K27me_{2/3} histone methyltransferase activity and spreading from nucleation sites (Margueron et al., 2009; Oksuz et al., 2018). This feed-forward property of PRC2/H3K27me₃ supports models where H3K27me₃ is a true epigenetic mark that persists through cell cycles and would explain at least partially for PRC2's ability to maintain transcriptional profiles (Hansen et al., 2008; Margueron et al., 2009).

Given the diversity of subunits that have been reported in various complexes, there is no agreed upon number of PRC1 complexes, but they have been generally subdivided into canonical PRC1 (cPRC1) and non-canonical PRC1 (ncPRC1) (Blackledge et al., 2015) (Figure 2). All PRC1 complexes share a core of a RING1A or RING1B and one of six Polycomb group ring-finger domain proteins (PCGF1-6). cPRC1 contains PCGF2/4, one CBX protein (CBX2/4/6/7/8) which recruits cPRC1 to the PRC2 mark H3K27me₃, and a Polyhomeotic homologous protein (PHC1/2/3) (Schuettengruber et al., 2017). cPRC1 catalyzes H2A119ub through RING1A/B subunits, can compact chromatin, and is generally associated with repression (Aranda et al., 2015; Cao et al., 2002; Ku et al., 2008). ncPRC1 contains RYBP or its paralog YAF2, and various PCGFs, and, depending on the complex, can have either activating and repressive functions. PRC1 proteins have roles in development, cell identity, senescence, and cancer (Gil and O'Loghlen, 2014).

PcG proteins repress chromatin by multiple mechanisms. One of the main mechanisms is by chromatin compaction, which makes chromatin inaccessible to some TFs, the transcriptional machinery, and chromatin remodelers (Schuettengruber et al., 2017; Shao et al., 1999) (Figure 3). Classically, PRC1 is considered the main effector of chromatin compaction (Cohen et al., 2018; Francis et al., 2004; Gil and O'Loghlen, 2014; Grau et al., 2011; Simon and Kingston, 2009), though there is also evidence of EZH1-containing PRC2 compacting *in vitro* chromatin templates and EZH2-PRC2 compacting

chromatin *in vivo* (Margueron et al., 2008; Terranova et al., 2008). While evidence remains to be seen in mammals, in flies a second mechanism of PcG repression involves PRC1 binding to the CBP, inhibiting CPB's enzymatic acetyltransferase activity, and thus inhibiting downstream transcriptional activation (Tie et al., 2016). A third mechanism of PcG repression is the looping of Polycomb domains to maintain chromatin domains in a repressed state (Ferrari et al., 2014; Lee et al., 2015). Another proposed mechanism of repression is the widespread distribution of H3K27me_{2/3}, which can suppress chromatin opening, enhancer activation, and transcription by directly preventing H3K27 acetylation from occurring on the same lysine residue (Ferrari et al., 2014; Lee et al., 2015).

While the PRC2 complex is classically thought to repress expression by eliciting chromatin compaction (Simon and Kingston, 2009), Polycomb-bound or -marked chromatin can be accessible to binding by some factors and transcribed, indicating that there are functions of Polycomb-regulated euchromatin (Becker et al., 2017; Beisel and Paro, 2011; Breiling et al., 2001; Dellino et al., 2004; Hawkins et al., 2010; Ku et al., 2008; Trojer and Reinberg, 2007). It is already known that the PRC2 complex repress various steps of transcription at promoters. The stepwise process of transcription involves recruitment of the transcriptional machinery to promoters, short transcription and then pausing of RNAPII, pausing release and productive elongation, and termination of transcription (Jonkers and Lis, 2015). The PRC2 histone mark inhibits transcription at multiple steps, including by inhibiting RNAPII recruitment by blocking acetylation of the same residue to H3K27ac (Chopra et al., 2011), inhibiting RNAPII release from pausing (Chen et al., 2012b), and inhibiting proper RNAPII elongation in gene bodies (Seenundun et al., 2010). Additionally, PRC2 can methylate non-histone proteins, including elongation factor A (ELOA), which seems to cause impaired ELOA ability to promote RNAPII release into the elongation stage (Ardehali et al., 2017). Despite identification of multiple methods

of PRC2-based repression at promoters, a remaining question in the field is why only a fraction of H3K27me3-marked genes derepress in response PRC2/H3K27me3 ablation (Bae et al., 2015; Ezhkova et al., 2011; Jadhav et al., 2016).

1.5. Polycomb repression in endoderm lineage development

Polycomb group proteins (PcG) are transcriptional repressor proteins that are documented as regulating the development and cell identity of many lineages, including endoderm (Margueron and Reinberg, 2011). In the intestine, Polycomb proteins are involved or implicated in intestinal stem cell proliferation, differentiation, and renewal, and for maintaining the appropriate balance of downstream cell types (Chiacchiera and Pasini, 2017; Koppens et al., 2016; López-Arribillaga et al., 2015; Vizán et al., 2016). In the pancreas, various PcG proteins are involved in regulating the number of endocrine progenitors and beta cells, are required for regeneration of the pancreatic exocrine compartment, and are required for beta cell regeneration and proliferation in diabetes mellitus (Cervantes et al., 2017; Chen et al., 2009; Fukuda et al., 2012; Mallen-St Clair et al., 2012; Xu et al., 2014). PRC1 proteins BMI1 and MEL18 contribute to colitis-associated cancer (Liu et al., 2017). In thymal epithelial cells, PcG proteins regulate growth (Guo et al., 2011; Liu et al., 2013). These examples of PcG proteins regulating proliferation, differentiation, and renewal in endoderm all suggest that PcG proteins may play a role in those processes in maturing and regenerating liver.

PcG proteins play known roles in multiple stages of liver development and homeostasis. In early liver development, EZH2, one of the two H3K27me3 histone methyltransferases, modulates the cell fate choice of embryonic endoderm to become

pancreatic or hepatic buds (Xu et al., 2011) and is required for hepatoblast proliferation (Koike et al., 2014). Loss of both EZH1 and EZH2 leads to chronic liver damage in adult mice and an inability to regenerate after liver toxin-induced damage (Bae et al., 2015). Significantly, genes marked by the PRC2 repressive histone modification H3K27me3 exhibit activation defects in iHeps, with genes in sonication-resistant heterochromatin being the most resistant to activation (Becker et al., 2017). Additionally, multiple PRC2 proteins have been implicated in repressing non-hepatocyte lineage genes in iHep reprogramming (Rastegar-Pouyani et al., 2016). These results not only leave open a role for Polycomb-based regulation in postnatal hepatic maturation but also suggest that manipulation of Polycomb proteins may be useful in the generation of artificial hepatocytes.

1.6. Figures

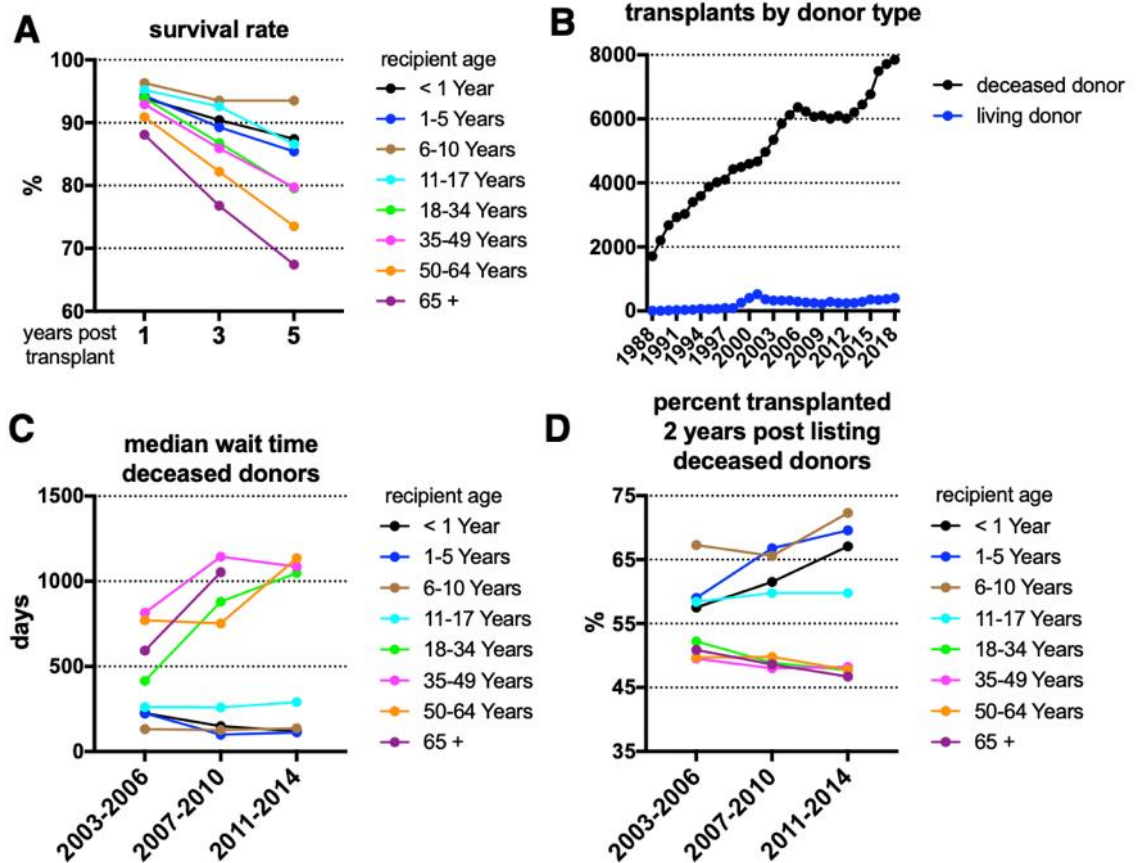


Figure 1: Liver transplant survival and demand in the US

All data in this figure was accessed from the U.S. Department of Health and Human Services Organ Procurement and Transplantation Net (<https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/>).

- A)** Liver Kaplan-Meier Patient Survival Rates For Transplants Performed: 2008 – 2015. 1-year survival based on 2012-2015 transplants, 3-year survival based on 2010-2013 transplants, 5-year survival based on 2006-2011 transplants. Based on OPTN data as of January 18, 2019.
- B)** Transplants by Donor Type. Based on OPTN data as of January 23, 2019.
- C)** Liver Competing Risk Median Waiting Time to Deceased Donor Transplant For Registrations Listed: 2003-2014. Missing data denotes n less than 10 or fewer than half of recipients having been transplanted. Data as of January 18, 2019
- D)** Liver Competing Risk Percentage with Deceased Donor Transplant at Specific time Points For Registrations Listed: 2003-2014. Based on OPTN data as of January 18, 2019

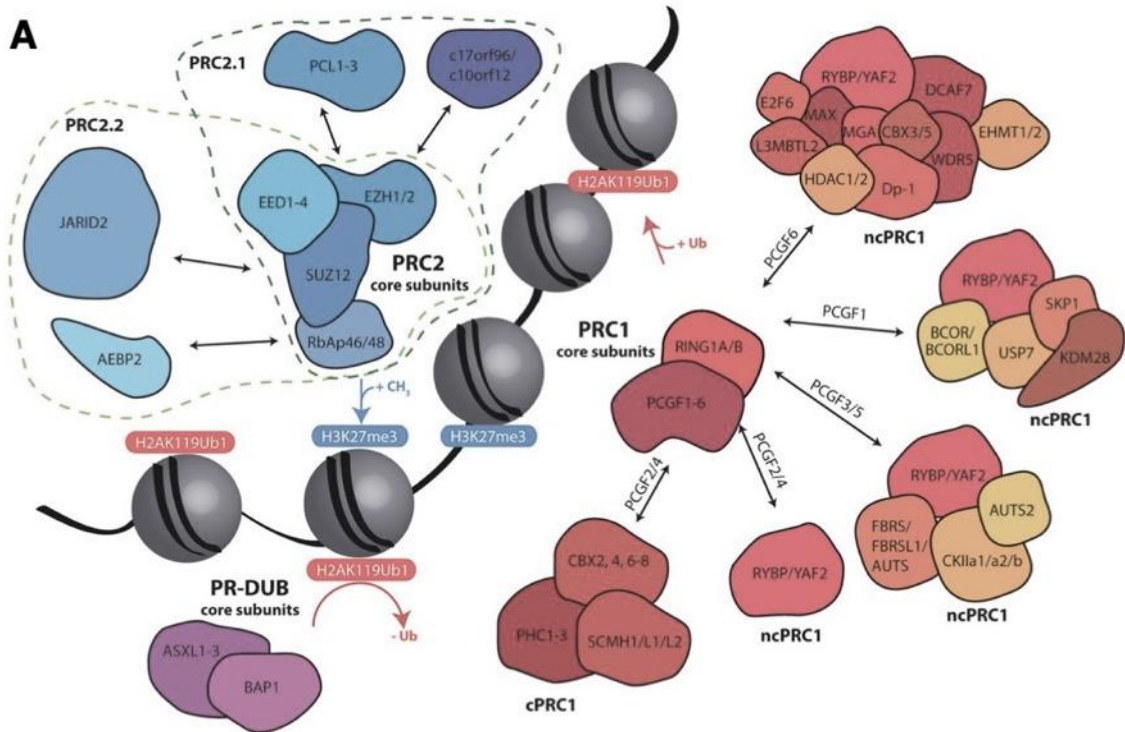


Figure 2: PRC1 and 2 complexes

A) A subset of PRC1 and 2 complexes. Adapted from (Chittock et al., 2017).

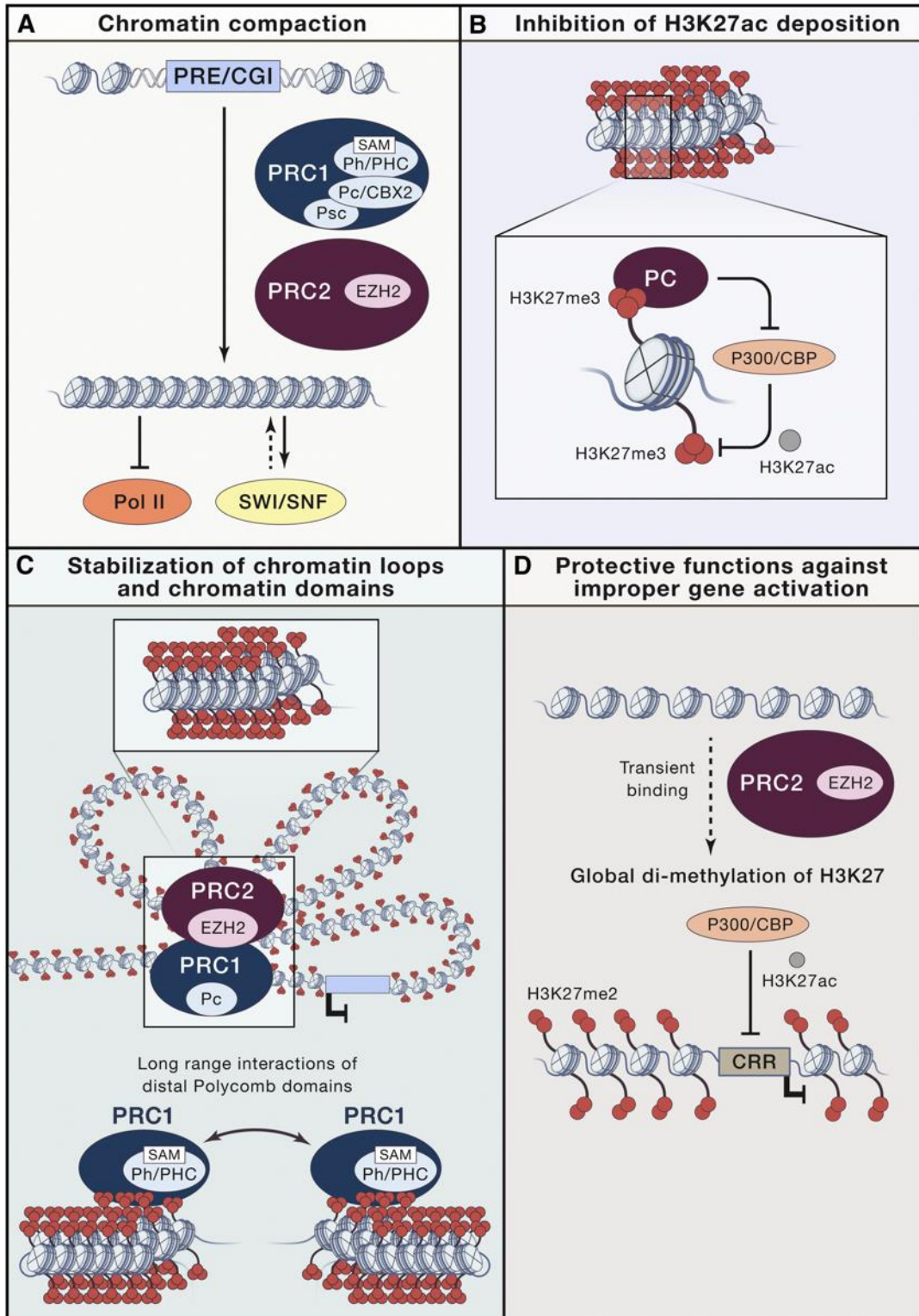


Figure 3: Multiple mechanisms of PcG repression

- A)** PRC1 can induce chromatin compaction via a positively charged compacting region present in mammalian CBX2 or fly Psc. In addition, Pc/CBX (via H3K27me₃-binding through its chromodomain) and Ph/PHC (via polymerization through its SAM domain) can contribute to chromatin compaction, which can interfere with SWI/SNF-mediated chromatin remodeling or Pol II recruitment. PRC2 can also contribute to the compaction of nucleosomal arrays.
- B)** Trimethylation of H3K27 (H3K27me₃) can directly block acetylation of H3K27 (H3K27ac), which is involved in gene activation. Further, Pc/CBX can inhibit the acetyltransferase activity of CBP, therefore favoring methylation of H3K27.
- C)** Distribution of the H3K27me₃ mark over large genomic regions might stabilize chromatin-looping interactions between PcG-binding sites and gene regulatory regions, thereby contributing to stably locking genes in a repressed state. Oligomerization of the SAM domain of Ph/PHC is essential for PcG-mediated repression and can mediate long-range interactions between distal Polycomb domains.
- D)** Dimethylation of H3K27 (H3K27me₂) exerts protective functions by preventing acetylation of cis regulatory regions (CRRs), such as enhancers or promoters, thereby inhibiting their inappropriate activation.

This figure comes from a review from the Cavalli group (Schuettengruber et al., 2017).

1.7. Tables

Table 1. Select reports of hepatocyte transplantation in humans

Indication	Number of patients	Age of patients	Outcome	Reference
Criggler-Najjar Type I	1	9 years	decreased bilirubin ~4 months, received OLT	(Ambrosino et al., 2005)
Criggler-Najjar Type I	2	1 and 9 years	decreased bilirubin ~4 and ~6 months received OLT	(Lysy et al., 2008)
Urea cycle disorder	4	1 day to 3 years	4-13 months stability, received OLT, 1 death	(Meyburg et al., 2009)
Familial hypercholesterolemia	5	7 to 41 years	3 patients with reduced LDL for ~4 months	(Grossman et al., 1995)
Glycogen storage disorders	1	8 years	ability to eat a normal diet for the 7-month follow-up	(Lee et al., 2007)
Glycogen storage disorder	1	47 years	ability to eat a normal diet for the 9-month follow-up	(Muraca et al., 2002)
Refsum disease	1	4 years	Cholestasis resumed, symptoms ameliorated for 18-month follow-up	(Sokal et al., 2003)
Hemophilia A	2	3 and 35 months	Reduced exogenous Factor VII required for 6 months	(Dhawan et al., 2004)
PKU	1	27 years	Ability to eat a normal diet for 7 months	(Soltys et al., 2017)

OLT: orthotopic liver transplantation

1.8. References

- Agarwal, S., Holton, K.L., and Lanza, R. (2008). Efficient differentiation of functional hepatocytes from human embryonic stem cells. *Stem Cells* 26, 1117–1127.
- Alpern, D., Langer, D., Ballester, B., Le Gras, S., Romier, C., Mengus, G., and Davidson, I. (2014). TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. *Elife* 3, e03613.
- Ambrosino, G., Varotto, S., Strom, S.C., Guariso, G., Franchin, E., Miotto, D., Caenazzo, L., Basso, S., Carraro, P., Valente, M.L., et al. (2005). Isolated hepatocyte transplantation for Crigler-Najjar syndrome type 1. *Cell Transplant* 14, 151–157.
- Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. *Sci Adv* 1, e1500737.
- Ardehali, M.B., Anselmo, A., Cochrane, J.C., Kundu, S., Sadreyev, R.I., and Kingston, R.E. (2017). Polycomb Repressive Complex 2 Methylates Elongin A to Regulate Transcription. *Mol. Cell* 68, 872-884.e6.
- Avior, Y., Levy, G., Zimerman, M., Kitsberg, D., Schwartz, R., Sadeh, R., Moussaieff, A., Cohen, M., Itskovitz-Eldor, J., and Nahmias, Y. (2015). Microbial-derived lithocholic acid and vitamin K2 drive the metabolic maturation of pluripotent stem cells-derived and fetal hepatocytes. *Hepatology* 62, 265–278.
- Bae, W.K., Kang, K., Yu, J.H., Yoo, K.H., Factor, V.M., Kaji, K., Matter, M., Thorgeirsson, S., and Hennighausen, L. (2015). The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration. *FASEB J.* 29, 1653–1662.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124.
- Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* 32, 29–41.
- Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* 68, 1023-1037.e15.
- Benhamouche, S., Decaens, T., Godard, C., Chambrey, R., Rickman, D.S., Moinard, C., Vasseur-Cognet, M., Kuo, C.J., Kahn, A., Perret, C., et al. (2006). Apc tumor suppressor gene is the “zonation-keeper” of mouse liver. *Dev. Cell* 10, 759–770.
- Bhate, A., Parker, D.J., Bebee, T.W., Ahn, J., Arif, W., Rashan, E.H., Chorghade, S., Chau, A., Lee, J.-H., Anakk, S., et al. (2015). ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat Commun* 6, 8768.
- Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 157, 1445–1459.
- Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* 16, 643–649.

- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20, 1123–1136.
- Brauner, P., Nibbelink, M., Flachs, P., Vítková, I., Kopecký, P., Mertelíková, I., Janderová, L., Pénicaud, L., Casteilla, L., Plavka, R., et al. (2001). Fast decline of hematopoiesis and uncoupling protein 2 content in human liver after birth: location of the protein in Kupffer cells. *Pediatr. Res.* 49, 440–447.
- Brolén, G., Sivertsson, L., Björquist, P., Eriksson, G., Ek, M., Semb, H., Johansson, I., Andersson, T.B., Ingelman-Sundberg, M., and Heins, N. (2010). Hepatocyte-like cells derived from human embryonic stem cells specifically via definitive endoderm and a progenitor stage. *J. Biotechnol.* 145, 284–294.
- Brown, S.W. (1966). Heterochromatin. *Science* 151, 417–425.
- Burke, Z.D., Reed, K.R., Pheese, T.J., Sansom, O.J., Clarke, A.R., and Tosh, D. (2009). Liver zonation occurs through a beta-catenin-dependent, c-Myc-independent mechanism. *Gastroenterology* 136, 2316-2324.e1-3.
- Cai, J., Zhao, Y., Liu, Y., Ye, F., Song, Z., Qin, H., Meng, S., Chen, Y., Zhou, R., Song, X., et al. (2007). Directed differentiation of human embryonic stem cells into functional hepatic cells. *Hepatology* 45, 1229–1239.
- Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell* 41, 67–81.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043.
- Celton-Morizur, S., Merlen, G., Couton, D., Margall-Ducos, G., and Desdouets, C. (2009). The insulin/Akt pathway controls a specific cell division program that leads to generation of binucleated tetraploid liver cells in rodents. *J. Clin. Invest.* 119, 1880–1887.
- Celton-Morizur, S., Merlen, G., Couton, D., and Desdouets, C. (2010). Polyploidy and liver proliferation: central role of insulin signaling. *Cell Cycle* 9, 460–466.
- Cervantes, S., Fontcuberta-PiSunyer, M., Servitja, J.-M., Fernandez-Ruiz, R., García, A., Sanchez, L., Lee, Y.-S., Gomis, R., and Gasa, R. (2017). Late-stage differentiation of embryonic pancreatic β -cells requires Jarid2. *Sci Rep* 7, 11643.
- Chen, H., Gu, X., Su, I.-hsi., Bottino, R., Contreras, J.L., Tarakhovsky, A., and Kim, S.K. (2009). Polycomb protein Ezh2 regulates pancreatic beta-cell Ink4a/Arf expression and regeneration in diabetes mellitus. *Genes Dev.* 23, 975–985.
- Chen, H.-Z., Ouseph, M.M., Li, J., Pécot, T., Chokshi, V., Kent, L., Bae, S., Byrne, M., Duran, C., Comstock, G., et al. (2012a). Canonical and atypical E2Fs regulate the mammalian endocycle. *Nat. Cell Biol.* 14, 1192–1202.
- Chen, S., Ma, J., Wu, F., Xiong, L.-J., Ma, H., Xu, W., Lv, R., Li, X., Villen, J., Gygi, S.P., et al. (2012b). The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes Dev.* 26, 1364–1375.

- Chen, Y.-F., Tseng, C.-Y., Wang, H.-W., Kuo, H.-C., Yang, V.W., and Lee, O.K. (2012c). Rapid generation of mature hepatocyte-like cells from human induced pluripotent stem cells by an efficient three-step protocol. *Hepatology* 55, 1193–1203.
- Chiacchiera, F., and Pasini, D. (2017). Control of adult intestinal identity by the Polycomb repressive machinery. *Cell Cycle* 16, 243–244.
- Chipchase, M.D., O'Neill, M., and Melton, D.W. (2003). Characterization of premature liver polyploidy in DNA repair (Erc1)-deficient mice. *Hepatology* 38, 958–966.
- Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. *Biochem. Soc. Trans.* 45, 193–205.
- Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The polycomb group mutant *esc* leads to augmented levels of paused Pol II in the *Drosophila* embryo. *Mol. Cell* 42, 837–844.
- Cohen, I., Zhao, D., Bar, C., Valdes, V.J., Dauber-Decker, K.L., Nguyen, M.B., Nakayama, M., Rendl, M., Bickmore, W.A., Koseki, H., et al. (2018). PRC1 Fine-tunes Gene Repression and Activation to Safeguard Skin Development and Stem Cell Specification. *Cell Stem Cell* 22, 726-739.e7.
- Colletti, M., Cicchini, C., Conigliaro, A., Santangelo, L., Alonzi, T., Pasquini, E., Tripodi, M., and Amicone, L. (2009). Convergence of Wnt signaling on the HNF4 α -driven transcription in controlling liver zonation. *Gastroenterology* 137, 660–672.
- Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep* 7, 1456–1470.
- Creasy, K.T., Jiang, J., Ren, H., Peterson, M.L., and Spear, B.T. (2016). Zinc Fingers and Homeoboxes 2 (Zhx2) Regulates Sexually Dimorphic Cyp Gene Expression in the Adult Mouse Liver. *Gene Expr.* 17, 7–17.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111, 185–196.
- Decaux, J.F., Ferré, P., Robin, D., Robin, P., and Girard, J. (1988). Decreased hepatic fatty acid oxidation at weaning in the rat is not linked to a variation of malonyl-CoA concentration. *J. Biol. Chem.* 263, 3284–3289.
- Devi, B.G., Gupta, P.D., and Habeebullah, C.M. (1992). Changes in membrane fluidity during human liver development. *Biochem. Int.* 28, 41–49.
- Dhawan, A., Mitry, R.R., Hughes, R.D., Lehec, S., Terry, C., Bansal, S., Arya, R., Wade, J.J., Verma, A., Heaton, N.D., et al. (2004). Hepatocyte transplantation for inherited factor VII deficiency. *Transplantation* 78, 1812–1814.
- Donato, M.T., Lahoz, A., Jiménez, N., Pérez, G., Serralta, A., Mir, J., Castell, J.V., and Gómez-Lechón, M.J. (2006). Potential impact of steatosis on cytochrome P450 enzymes of human hepatocytes isolated from fatty liver grafts. *Drug Metab. Dispos.* 34, 1556–1562.
- Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T., et al. (2014). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. *Cell Stem Cell* 14, 394–403.
- Duan, Y., Ma, X., Ma, X., Zou, W., Wang, C., Bahbahan, I.S., Ahuja, T.P., Tolstikov, V., and Zern, M.A. (2010). Differentiation and characterization of metabolically functioning hepatocytes from human embryonic stem cells. *Stem Cells* 28, 674–686.

- Duncan, A.W., Dorrell, C., and Grompe, M. (2009). Stem cells and liver regeneration. *Gastroenterology* 137, 466–481.
- Duncan, A.W., Taylor, M.H., Hickey, R.D., Hanlon Newell, A.E., Lenzi, M.L., Olson, S.B., Finegold, M.J., and Grompe, M. (2010). The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* 467, 707–710.
- Durand, F., and Valla, D. (2005). Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. *J. Hepatol.* 42 Suppl, S100-107.
- Ehara, T., Kamei, Y., Yuan, X., Takahashi, M., Kanai, S., Tamura, E., Tsujimoto, K., Tamiya, T., Nakagawa, Y., Shimano, H., et al. (2015). Ligand-activated PPAR α -dependent DNA demethylation regulates the fatty acid β -oxidation genes in the postnatal liver. *Diabetes* 64, 775–784.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156.
- Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62.
- Filipescu, D., Müller, S., and Almouzni, G. (2014). Histone H3 variants and their chaperones during development and disease: contributing to epigenetic control. *Annu. Rev. Cell Dev. Biol.* 30, 615–646.
- Fox, I.J., Chowdhury, J.R., Kaufman, S.S., Goertzen, T.C., Chowdhury, N.R., Warkentin, P.I., Dorko, K., Sauter, B.V., and Strom, S.C. (1998). Treatment of the Crigler-Najjar syndrome type I with hepatocyte transplantation. *N. Engl. J. Med.* 338, 1422–1426.
- Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574–1577.
- Fukuda, A., Morris, J.P., and Hebrok, M. (2012). Bmi1 is required for regeneration of the exocrine pancreas in mice. *Gastroenterology* 143, 821-831.e2.
- Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341–1345.
- Gebhardt, R., and Gaunitz, F. (1997). Cell-cell interactions in the regulation of the expression of hepatic enzymes. *Cell Biol. Toxicol.* 13, 263–273.
- Gil, J., and O’Loghlen, A. (2014). PRC1 complex diversity: where is it taking us? *Trends Cell Biol.* 24, 632–641.
- Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C.I., Kubo, A., Shafritz, D.A., and Keller, G. (2006). BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat. Biotechnol.* 24, 1402–1411.
- Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* 25, 2210–2221.
- Grossman, M., Rader, D.J., Muller, D.W., Kolansky, D.M., Kozarsky, K., Clark, B.J., Stein, E.A., Lupien, P.J., Brewer, H.B., and Raper, S.E. (1995). A pilot study of ex vivo gene therapy for homozygous familial hypercholesterolaemia. *Nat. Med.* 1, 1148–1154.
- Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007). Chromatin challenges during DNA replication and repair. *Cell* 128, 721–733.

- Guo, Y., Miyazaki, M., Itoi, M., Satoh, R., Iwama, A., Amagai, T., Kawamoto, H., and Kanno, M. (2011). Polycomb group gene *Bmi1* plays a role in the growth of thymic epithelial cells. *Eur. J. Immunol.* 41, 1098–1107.
- Hannan, N.R.F., Fordham, R.P., Syed, Y.A., Moignard, V., Berry, A., Bautista, R., Hanley, N.A., Jensen, K.B., and Vallier, L. (2013). Generation of multipotent foregut stem cells from human pluripotent stem cells. *Stem Cell Reports* 1, 293–306.
- Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* 10, 1291–1300.
- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb Wiss Botanik* 762–818.
- Holoch, D., and Margueron, R. (2017). Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem. Sci.* 42, 531–542.
- Hsu, S.-H., Delgado, E.R., Otero, P.A., Teng, K.-Y., Kutay, H., Meehan, K.M., Moroney, J.B., Monga, J.K., Hand, N.J., Friedman, J.R., et al. (2016). MicroRNA-122 regulates polyploidization in the murine liver. *Hepatology* 64, 599–615.
- Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell* 14, 370–384.
- Hurst, J. (2012). A modern Cosmas and Damian: Sir Roy Calne and Thomas Starzl receive the 2012 Lasker-DeBakey Clinical Medical Research Award. *J. Clin. Invest.* 122, 3378–3382.
- Jiang, J., Creasy, K.T., Purnell, J., Peterson, M.L., and Spear, B.T. (2017). *Zhx2* (zinc fingers and homeoboxes 2) regulates major urinary protein gene expression in the mouse liver. *J. Biol. Chem.* 292, 6765–6774.
- Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 167–177.
- Jung, J., Zheng, M., Goldfarb, M., and Zaret, K.S. (1999). Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* 284, 1998–2003.
- Jungermann, K., and Kietzmann, T. (1996). Zonation of parenchymal and nonparenchymal metabolism in liver. *Annu. Rev. Nutr.* 16, 179–203.
- Jungermann, K., and Kietzmann, T. (2000). Oxygen: modulator of metabolic zonation and disease of the liver. *Hepatology* 31, 255–260.
- Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics* 206, 1699–1725.
- Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* 20, 1110–1122.
- Koike, H., Ouchi, R., Ueno, Y., Nakata, S., Obana, Y., Sekine, K., Zheng, Y.-W., Takebe, T., Isono, K., Koseki, H., et al. (2014). Polycomb group protein *Ezh2* regulates hepatic progenitor cell proliferation and differentiation in murine embryonic liver. *PLoS ONE* 9, e104776.
- Koppens, M.A.J., Bounova, G., Gargiulo, G., Tanger, E., Janssen, H., Cornelissen-Steijger, P., Blom, M., Song, J.-Y., Wessels, L.F.A., and van Lohuizen, M. (2016). Deletion of Polycomb Repressive Complex 2 From Mouse Intestine Causes Loss of Stem Cells. *Gastroenterology* 151, 684-697.e12.

- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4, e1000242.
- Kyrmizi, I., Hatzis, P., Katrakili, N., Tronche, F., Gonzalez, F.J., and Talianidis, I. (2006). Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev.* 20, 2293–2305.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116–120.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.
- Lee, H.-G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res.* 25, 1170–1181.
- Lee, K.-W., Lee, J.-H., Shin, S.W., Kim, S.J., Joh, J.W., Lee, D.-H., Kim, J.-W., Park, H.-Y., Lee, S.-Y., Lee, H.H., et al. (2007). Hepatocyte transplantation for glycogen storage disease type Ib. *Cell Transplant* 16, 629–637.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313.
- Leibing, T., Géraud, C., Augustin, I., Boutros, M., Augustin, H.G., Okun, J.G., Langhans, C.-D., Zierow, J., Wohlfeil, S.A., Olsavszky, V., et al. (2018). Angiocrine Wnt signaling controls liver growth and metabolic maturation in mice. *Hepatology* 68, 707–722.
- Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570.
- Lewis, P.H. (1947). New mutants. *Drosoph. Inf. Serv.*
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719.
- Li, D., Cen, J., Chen, X., Conway, E.M., Ji, Y., and Hui, L. (2013). Hepatic loss of survivin impairs postnatal liver development and promotes expansion of hepatic progenitor cells in mice. *Hepatology* 58, 2109–2121.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.
- Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291.
- Liu, B., Liu, Y.-F., Du, Y.-R., Mardaryev, A.N., Yang, W., Chen, H., Xu, Z.-M., Xu, C.-Q., Zhang, X.-R., Botchkarev, V.A., et al. (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. *Development* 140, 780–788.
- Liu, H., Di Cunto, F., Imarisio, S., and Reid, L.M. (2003). Citron kinase is a cell cycle-dependent, nuclear protein required for G2/M transition of hepatocytes. *J. Biol. Chem.* 278, 2541–2548.
- López-Arribillaga, E., Rodilla, V., Pellegrinet, L., Guiu, J., Iglesias, M., Roman, A.C., Gutarra, S., González, S., Muñoz-Cánoves, P., Fernández-Salguero, P., et al. (2015). Bmi1 regulates murine intestinal stem cell proliferation and self-renewal downstream of Notch. *Development* 142, 41–50.

- Lysy, P.-A., Najimi, M., Stephenne, X., Bourgois, A., Smets, F., and Sokal, E.-M. (2008). Liver cell transplantation for Crigler-Najjar syndrome type I: update and perspectives. *World J. Gastroenterol.* 14, 3464–3470.
- Mallanna, S.K., and Duncan, S.A. (2013). Differentiation of hepatocytes from pluripotent stem cells. *Curr Protoc Stem Cell Biol* 26, Unit 1G.4.
- Mallen-St Clair, J., Soydaner-Azeloglu, R., Lee, K.E., Taylor, L., Livanos, A., Pylayeva-Gupta, Y., Miller, G., Margueron, R., Reinberg, D., and Bar-Sagi, D. (2012). EZH2 couples pancreatic regeneration to neoplastic progression. *Genes Dev.* 26, 439–444.
- Margall-Ducos, G., Celton-Morizur, S., Couton, D., Br erie, O., and Desdouets, C. (2007). Liver tetraploidization is controlled by a new process of incomplete cytokinesis. *J. Cell. Sci.* 120, 3633–3639.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* 32, 503–518.
- Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762–767.
- Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 78, 7634–7638.
- Mayhew, C.N., Bosco, E.E., Fox, S.R., Okaya, T., Tarapore, P., Schwemberger, S.J., Babcock, G.F., Lentsch, A.B., Fukasawa, K., and Knudsen, E.S. (2005). Liver-specific pRB loss results in ectopic cell cycle entry and aberrant ploidy. *Cancer Res.* 65, 4568–4577.
- Meyburg, J., Das, A.M., Hoerster, F., Lindner, M., Kriegbaum, H., Engelmann, G., Schmidt, J., Ott, M., Pettenazzo, A., Luecke, T., et al. (2009). One liver for four children: first clinical series of liver cell transplantation for severe neonatal urea cycle defects. *Transplantation* 87, 636–641.
- Morelli, L. (2008). Postnatal development of intestinal microflora as influenced by infant nutrition. *J. Nutr.* 138, 1791S-1795S.
- Morford, L.A., Davis, C., Jin, L., Dobierzewska, A., Peterson, M.L., and Spear, B.T. (2007). The oncofetal gene glypican 3 is regulated in the postnatal liver by zinc fingers and homeoboxes 2 and in the regenerating liver by alpha-fetoprotein regulator 2. *Hepatology* 46, 1541–1547.
- M ller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O’Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* 111, 197–208.
- Muraca, M., Gerunda, G., Neri, D., Vilei, M.-T., Granato, A., Feltracco, P., Meroni, M., Giron, G., and Burlina, A.B. (2002). Hepatocyte transplantation as a treatment for glycogen storage disease type 1a. *Lancet* 359, 317–318.
- Neuberger, J. (2016). An update on liver transplantation: A critical review. *J. Autoimmun.* 66, 51–59.
- Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*.

- NIDDK (2017). Definition & Facts of Liver Transplant (NIDDK).
- Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* 2, 2006.0017.
- Oinonen, T., and Lindros, K.O. (1998). Zonation of hepatic cytochrome P-450 expression and regulation. *Biochem. J.* 329 (Pt 1), 17–35.
- Oksuz, O., Narendra, V., Lee, C.-H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* 70, 1149-1162.e5.
- OPTN (2019a). Liver Kaplan-Meier Patient Survival Rates For Transplants Performed : 2008 - 2015 (U.S. Department of Health & Human Services, Organ Procurement and Transplantation Network).
- OPTN (2019b). Liver Competing Risk Percentage with Deceased Donor Transplant at Specific Time Points For Registrations Listed : 2003-2014 (U.S. Department of Health & Human Services, Organ Procurement and Transplantation Network).
- OPTN (2019c). Liver Competing Risk Median Waiting Time to Deceased Donor Transplant For Registrations Listed : 2003-2014 (U.S. Department of Health & Human Services, Organ Procurement and Transplantation Network).
- OPTN (2019d). Transplants by Donor Type U.S. Transplants Performed : January 1, 1988 - December 31, 2018 (U.S. Department of Health & Human Services, Organ Procurement and Transplantation Network).
- Pandit, S.K., Westendorp, B., Nantasanti, S., van Liere, E., Tooten, P.C.J., Cornelissen, P.W.A., Toussaint, M.J.M., Lamers, W.H., and de Bruin, A. (2012). E2F8 is essential for polyploidization in mammalian cells. *Nat. Cell Biol.* 14, 1181–1191.
- Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310.
- Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290–1302.
- Perez-Castillo, A., Schwartz, H.L., and Oppenheimer, J.H. (1987). Rat hepatic mRNA-S14 and lipogenic enzymes during weaning: role of S14 in lipogenesis. *Am. J. Physiol.* 253, E536-542.
- Périchon, R., and Bourre, J.M. (1995). Peroxisomal beta-oxidation activity and catalase activity during development and aging in mouse liver. *Biochimie* 77, 288–293.
- Perincheri, S., Dingle, R.W.C., Peterson, M.L., and Spear, B.T. (2005). Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the Zfx2 gene. *Proc. Natl. Acad. Sci. U.S.A.* 102, 396–401.
- Perugorria, M.J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J.J.G., Bujanda, L., and Banales, J.M. (2019). Wnt- β -catenin signalling in liver development, health and disease. *Nat Rev Gastroenterol Hepatol* 16, 121–136.
- Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* 10, 192–206.

- Rastegar-Pouyani, S., Khazaei, N., Wee, P., Mohammadnia, A., and Yaqubi, M. (2016). Role of Hepatic-Specific Transcription Factors and Polycomb Repressive Complex 2 during Induction of Fibroblasts to Hepatic Fate. *PLoS ONE* 11, e0167081.
- Roelandt, P., Vanhove, J., and Verfaillie, C. (2013). Directed differentiation of pluripotent stem cells to functional hepatocytes. *Methods Mol. Biol.* 997, 141–147.
- Rossi, J.M., Dunn, N.R., Hogan, B.L., and Zaret, K.S. (2001). Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev.* 15, 1998–2009.
- Sagias, F.G., Mitry, R.R., Hughes, R.D., Lehec, S.C., Patel, A.G., Rela, M., Mieli-Vergani, G., Heaton, N.D., and Dhawan, A. (2010). N-acetylcysteine improves the viability of human hepatocytes isolated from severely steatotic donor liver tissue. *Cell Transplant* 19, 1487–1492.
- Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* 465, 243–247.
- Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171, 34–57.
- Seenundun, S., Rampalli, S., Liu, Q.-C., Aziz, A., Pali, C., Hong, S., Blais, A., Brand, M., Ge, K., and Dilworth, F.J. (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *EMBO J.* 29, 1401–1411.
- Sekine, K., Chen, Y.-R., Kojima, N., Ogata, K., Fukamizu, A., and Miyajima, A. (2007). Foxo1 links insulin signaling to C/EBPalpha and regulates gluconeogenesis during liver development. *EMBO J.* 26, 3607–3615.
- Septer, S., Edwards, G., Gunewardena, S., Wolfe, A., Li, H., Daniel, J., and Apte, U. (2012). Yes-associated protein is involved in proliferation and differentiation during postnatal liver development. *Am. J. Physiol. Gastrointest. Liver Physiol.* 302, G493-503.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98, 37–46.
- Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139, 1303–1314.
- Siller, R., Greenhough, S., Naumovska, E., and Sullivan, G.J. (2015). Small-molecule-driven hepatocyte differentiation of human pluripotent stem cells. *Stem Cell Reports* 4, 939–952.
- Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* 10, 697–708.
- Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 51, 297–305.
- Sitbon, D., Podsypanina, K., Yadav, T., and Almouzni, G. (2017). Shaping Chromatin in the Nucleus: The Bricks and the Architects. *Cold Spring Harb. Symp. Quant. Biol.* 82, 1–14.
- Slifer, E.H. (1942). A mutant stock of *Drosophila* with extra sex combs. *J. Exp. Zool.* 31–40.
- Snitow, M., Lu, M., Cheng, L., Zhou, S., and Morrissey, E.E. (2016). Ezh2 restricts the smooth muscle lineage during mouse lung mesothelial development. *Development* 143, 3733–3741.

- Sokal, E.M., Smets, F., Bourgois, A., Van Maldergem, L., Buts, J.-P., Reding, R., Bernard Otte, J., Evrard, V., Latinne, D., Vincent, M.F., et al. (2003). Hepatocyte transplantation in a 4-year-old girl with peroxisomal biogenesis disease: technique, safety, and metabolic follow-up. *Transplantation* 76, 735–738.
- Soltys, K.A., Setoyama, K., Tafaleng, E.N., Soto Gutiérrez, A., Fong, J., Fukumitsu, K., Nishikawa, T., Nagaya, M., Sada, R., Haberman, K., et al. (2017). Host conditioning and rejection monitoring in hepatocyte transplantation in humans. *J. Hepatol.* 66, 987–1000.
- Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* 27, 2663–2677.
- Song, Z., Cai, J., Liu, Y., Zhao, D., Yong, J., Duo, S., Song, X., Guo, Y., Zhao, Y., Qin, H., et al. (2009). Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell Res.* 19, 1233–1242.
- Spear, B.T., Jin, L., Ramasamy, S., and Dobierzewska, A. (2006). Transcriptional control in the mammalian liver: liver development, perinatal repression, and zonal gene regulation. *Cell. Mol. Life Sci.* 63, 2922–2938.
- Strom, S.C., Fisher, R.A., Thompson, M.T., Sanyal, A.J., Cole, P.E., Ham, J.M., and Posner, M.P. (1997). Hepatocyte transplantation as a bridge to orthotopic liver transplantation in terminal liver failure. *Transplantation* 63, 559–569.
- Suzuki, T., Kikuguchi, C., Nishijima, S., Nagashima, T., Takahashi, A., Okada, M., and Yamamoto, T. (2019). Postnatal liver functional maturation requires Cnot complex-mediated decay of mRNAs encoding cell cycle and immature liver genes. *Development* 146.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679.
- Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad. Sci. U.S.A.* 113, E744-753.
- Touboul, T., Hannan, N.R.F., Corbinau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J., et al. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* 51, 1754–1765.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13.
- Vizán, P., Beringer, M., and Di Croce, L. (2016). Polycomb-dependent control of cell fate in adult tissue. *EMBO J.* 35, 2268–2269.
- Wang, M.-J., Chen, F., Lau, J.T.Y., and Hu, Y.-P. (2017). Hepatocyte polyploidization and its association with pathophysiological processes. *Cell Death Dis* 8, e2805.
- Watanabe, T., and Tanaka, Y. (1982). Age-related alterations in the size of human hepatocytes. A study of mononuclear and binucleate cells. *Virchows Arch., B, Cell Pathol.* 39, 9–20.
- Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* 67, 545–579.

- Xie, Z., Zhang, H., Tsai, W., Zhang, Y., Du, Y., Zhong, J., Szpirer, C., Zhu, M., Cao, X., Barton, M.C., et al. (2008). Zinc finger protein ZBTB20 is a key repressor of alpha-fetoprotein gene transcription in liver. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10859–10864.
- Xu, C.-R., Cole, P.A., Meyers, D.J., Kormish, J., Dent, S., and Zaret, K.S. (2011). Chromatin “prepattern” and histone modifiers in a fate choice for liver and pancreas. *Science* 332, 963–966.
- Xu, C.-R., Li, L.-C., Donahue, G., Ying, L., Zhang, Y.-W., Gadue, P., and Zaret, K.S. (2014). Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. *EMBO J.* 33, 2157–2170.
- Yiangou, L., Ross, A.D.B., Goh, K.J., and Vallier, L. (2018). Human Pluripotent Stem Cell-Derived Endoderm for Modeling Development and Clinical Applications. *Cell Stem Cell* 22, 485–499.
- Zaret, K.S., and Grompe, M. (2008). Generation and regeneration of cells of the liver and pancreas. *Science* 322, 1490–1494.
- Zhu, S., Rezvani, M., Harbell, J., Mattis, A.N., Wolfe, A.R., Benet, L.Z., Willenbring, H., and Ding, S. (2014). Mouse liver repopulation with hepatocytes generated from human fibroblasts. *Nature* 508, 93–97.
- Zorn, A.M. (2008). Liver development. In *StemBook*, (Cambridge (MA): Harvard Stem Cell Institute), p.

Chapter 2: PRC2 proteins EZH1/2 regulate timely postnatal hepatocyte maturation and fibrosis by repression of euchromatic promoters

2.1. Preface

The manuscript presented in this chapter as originally published in *Gastroenterology* (Grindheim et al., 2019). It has been reformatted here in accordance with the University of Pennsylvania dissertation formatting guidelines. Extended methods are available in Appendices. All sequencing data is available under GEO accession number GSE119219. The references for this chapter are at the end of this chapter.

Authors:

Jessica Mae Grindheim¹⁻⁵, Dario Nicetto¹⁻³, Greg Donahue^{1-3,5}, Kenneth S Zaret^{1-3,5,6,#}

Affiliations:

¹Institute for Regenerative Medicine,

²Penn Epigenetics Institute,

³Dept. Cell and Developmental Biology,

⁴Dept. of Cancer Biology,

⁵Perelman School of Medicine, University of Pennsylvania, Smilow Center for Translational Research, 3400 Civic Center Blvd, Bldg. 421, Philadelphia, PA 19104-5157, USA.

⁶Lead contact

#Corresponding author: zaret@penmedicine.upenn.edu

Grant support: This work was supported by grants by University of Pennsylvania DSRB Training Grant NIH T32HD083185-01 to J.M.G and NIH R01GM036477 to K.S.Z.

Correspondence: Kenneth S Zaret, Ph.D. 3400 Civic Center Blvd., Rm. 9-132 Philadelphia, PA 19104-5157. zaret@penmedicine.upenn.edu. Office: 215-573-5813.

2.2. Respective Contributions

The majority of the experiments and analysis presented in this chapter were designed and performed by myself, with the guidance of my thesis advisor, Kenneth S. Zaret, PhD. For H3K27me3 chromatin immunoprecipitation experiments, I collected chromatin and Dario Nicetto, PhD performed the chromatin immunoprecipitation portion. For the P14 hepatocyte H3K27me3 samples, I prepared libraries and sequenced. For the M2 hepatocyte H3K27me3 samples, Dario Nicetto prepared libraries and sequenced. Greg Donahue provided custom scripts, advice for computational analysis, and performed most complex statistical analyses. I performed the majority of computational analysis, including mapping of high-throughput sequencing data, the differential gene expression analysis for RNA-seq, calling of marked compaction and H3K27me3 domains, and visualization of sequencing data. The manuscript was written by myself, with the assistance of Dr. Zaret.

2.3. Abstract

BACKGROUND & AIMS: The inability to derive fully functional hepatocytes from stem cells for transplantation, disease modeling, and drug testing may emanate from the lack of knowledge about mechanisms that underlie postnatal cell maturation. Additionally, the chromatin-based regulation of fibrosis genes remains incompletely understood. We investigated transcriptional dynamics, the roles of chromatin compaction, and the Polycomb Repressive Complex 2 (PRC2) in postnatal hepatocytes and in repression of fibrosis genes. **METHODS:** Mouse hepatocytes from postnatal day 14 (P14) and 2-month-old (M2) wild-type (*Wt*) and liver-specific deletion of PRC2 components *Ezh1* and *Ezh2* (*Ezh1/2*) mice were characterized using RNA-seq, H3K27me3 ChIP-seq, and sonication-resistant heterochromatin-seq, a newly presented method to map heterochromatin. Liver damage was characterized by histological analysis. **RESULTS:** We discovered a resource of more than 3000 genes differentially expressed in hepatocytes during P14 to M2 liver maturation. Genetic ablation of both PRC2 histone methyltransferases in perinatal livers causes hepatocytes to prematurely differentiate, expressing genes at P14 that would normally be induced by M2, along with a fibrotic phenotype. Genes with euchromatic H3K27me3+/H3K4me3+ promoter marking were found to be sensitive to *Ezh1/2* loss, which included maturation genes, non-liver lineage genes, and fibrosis genes. **CONCLUSIONS:** Polycomb repression is used to restrain expression of hepatocyte maturation, fibrosis, and lineage inappropriate genes at euchromatic promoters, thereby promoting liver homeostasis and preventing liver damage. Manipulation of Polycomb proteins may be used to improve hepatocyte derivation protocols and be a clinical target for liver fibrosis patients.

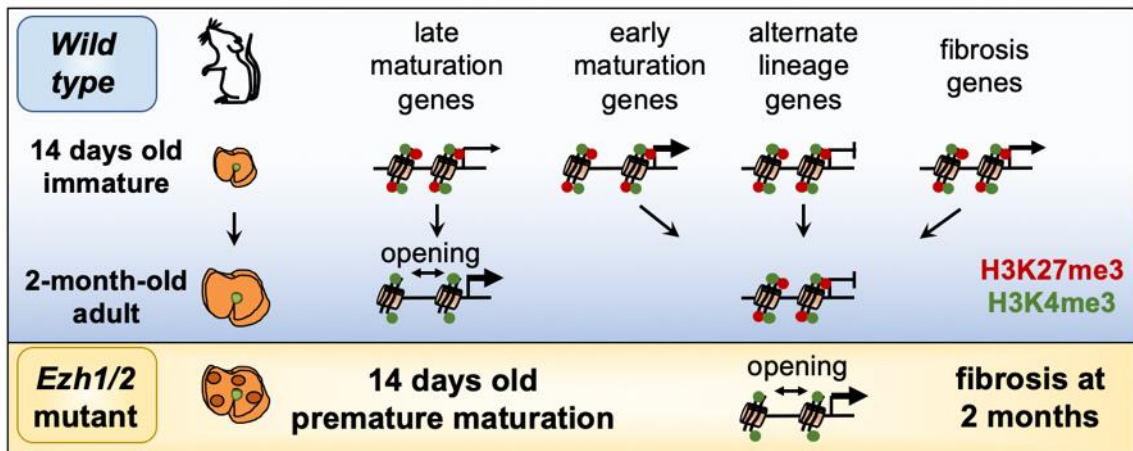
Abbreviations: ChIP, chromatin immunoprecipitation; DHS, DNase hypersensitive sites; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steohepatitis; PRC2, Polycomb Repressive Complex 2; RNA-seq, RNA-sequencing; RNAP2, RNA polymerase 2; srHC-seq; sonication-resistant heterochromatin sequencing; TSS, transcriptional start site.

Transcript profiling: RNA-seq, srHC-seq, and H3K27me3 ChIP-seq data is available on NCBI GEO GSE119219. Results are additionally summarized in Supplementary Tables as a resource datasets. Further information, code, and requests may be directed to and will be fulfilled by the lead contact, Ken Zaret (zaret@penmedicine.upenn.edu).

Disclosures: No competing interests to declare.

Author contributions: Conceptualization, J.M.G. and K.S.Z.; Investigation, J.M.G, D.N.; Formal Analysis, J.M.G, G.D.; Data Curation, J.M.G. and G.D.; Writing, J.M.G. and K.S.Z.; Resources, K.S.Z.; Supervision, K.S.Z.; Funding Acquisition, J.M.G. and K.S.Z.

2.4. Graphical abstract



2.5. Introduction

Due to the shortage of fully differentiated cells for transplantation and disease modeling of various tissue types, there is interest in generating replacement cells. Directed differentiation and cell reprogramming can generate early stage cells, but with a failure to activate a terminally differentiated transcriptional program and a failure to repress genes of the starting cell type (Horisawa and Suzuki, 2015; Ieda, 2013; Johannesson et al., 2015; Patel et al., 2016). In directed differentiation of hepatocyte-like cells, many fetal genes can be activated, but often there is failure to induce mature CYP P450 enzymes and repress fetal markers such as *Afp* (Chen et al., 2012; Duan et al., 2010; Roelandt et al., 2013; Si-Tayeb et al., 2010; Song et al., 2009). When fibroblasts are reprogrammed to hepatic cells, the cells express liver markers and perform some liver metabolic functions but fail to activate various mature liver genes and cannot consistently rescue liver damage (Huang et al., 2011, 2014; Sekiya and Suzuki, 2011). The literature on embryonic liver development (Duncan et al., 2009; Zaret and Grompe, 2008) rarely covers postnatal hepatic maturation (Bhate et al., 2015; Morford et al., 2007; Perincheri et al., 2005), though there are major physiological changes in liver size, diet (Shearer et al., 1982), microbiota and microbial metabolites (Avior et al., 2015; Morelli, 2008), sexual maturation, and polyploidization (Duncan, 2013). Thus, there is a need to understand postnatal hepatic maturation.

Components of Polycomb Repressive Complexes 1 or 2 (PRC1 or PRC2) regulate multiple gastrointestinal cell types. The PRC2 protein EED promotes intestinal stem cell proliferation and inhibits differentiation (Koppens et al., 2016), while the PRC1 protein BMI1 regulates intestinal stem cell proliferation and renewal (López-Arribillaga et al., 2015). PRC1 proteins BMI1 and MEL18 contribute to colitis-associated cancer (Liu et al.,

2017). PRC2 protein JARID2 is needed for late differentiation of pancreatic beta-cells (Cervantes et al., 2017). In liver development, EZH2, one of the two H3K27me3 histone methyltransferases, modulates the cell fate choice of embryonic endoderm to become pancreatic or hepatic buds (Xu et al., 2011) and is required for hepatoblast proliferation (Koike et al., 2014). Loss of both EZH1 and EZH2 leads to chronic liver damage in adult mice (Bae et al., 2015). Genes marked by the PRC2 repressive histone modification H3K27me3 exhibit activation defects in human fibroblast-to-hepatocyte reprogramming protocols, with genes in heterochromatin being the most resistant to activation (Becker et al., 2017; Matoba et al., 2014; Onder et al., 2012; Soufi et al., 2012; Sridharan et al., 2013). This leaves open a role for Polycomb-based regulation in postnatal hepatic maturation.

By genetically ablating both *Ezh1* and *Ezh2*, we investigated the relationship between postnatal transcriptional, H3K27me3, and chromatin compaction dynamics in postnatal hepatocyte maturation and homeostasis. We found that PRC2 represses three main classes of genes in postnatal hepatocytes; hepatic maturation genes, non-liver lineage genes, and fibrosis genes. Genes that derepress in response to *Ezh1/2* loss have a unique chromatin signature. These findings impact our understanding of the means by which hepatocytes mature postnatally, how hepatocyte lineage fidelity is maintained, and how fibrosis genes are primed for a transcriptional response.

2.6. Results

Postnatal hepatic maturation involves differential expression of thousands of genes

Using RNA-seq on hepatocytes isolated by liver perfusion, we found 1215 upregulated and 2011 downregulated genes in hepatic maturation between P14 and M2 ($\alpha \geq 0.05$,

FC ≥ 2) (Figure 1A, Supplementary Figure 1A-D, Supplementary Table 2, 3, 4). Upregulated genes include the key liver metabolic enzymes *Ces1f* and *Ces3b*, and are enriched by Gene Ontology for the Uniprot liver tissue expression category, P450 enzymes, and genes for metabolism of retinol, xenobiotics, and bile acids (Figure 1AB, Supplementary Table 5). As expected, fetal liver genes *Afp* and *Gpc3* are among the top downregulated genes (Figure 1A) (Belayew and Tilghman, 1982; Morford et al., 2007). Downregulated maturation genes enrich cell cycle, cell adhesion, differentiation, and signaling categories (Figure 1B). The cell cycle-associated categories reflect diminished proliferation in adult hepatocytes, as assayed by staining for proliferation marker PCNA (Supplementary Figure 1E). PubMed searches for a random sampling of genes in the cell adhesion and signaling categories and “liver” or “hepatocyte” often return no results, indicating that these may be new targets to study in liver maturation. Thus, P14 hepatocytes are transitioning extensively from a fetal to a mature transcriptional program.

P14 and M2 H3K27me3 states correlate with postnatal hepatic maturation

We profiled H3K27me3 in P14 and M2 hepatocytes (Supplementary Table 2, 3). As expected, H3K27me3 was absent from the expressed liver gene *Alb* and present at the silent *Hoxd* cluster, and this anticorrelation of expression and H3K27me3 was observed genome-wide (Supplementary Figure 1FG). H3K27me3 genomic coverage increased from 608 MB at P14 to 827 MB at M2 (Supplementary Figure 1H).

We called promoters and gene bodies as “lacking”, “losing”, “gaining”, or “retaining” H3K27me3 from P14 to M2. Acquisition of H3K27me3 is characteristic of many genes that are downregulated in the maturation transition from P14 to M2, with 44% of such promoters gaining (n=179) or retaining (n=909) H3K27me3 and 40% gene bodies gaining (n=172) or retaining (n=821) H3K27me3 (Figure 1C). In the case of gene bodies that retain

H3K27me3, the percent of coverage increases from a median of 62% at P14 to 78% at M2; thus increasing H3K27me3 presence at these “retain H3K27me3 genes” was not discernible in categorical lack/lose/gain/retain calls (Figure 1D). This increase was not solely due to increased H3K27me3 coverage in M2 (p-value at 0.001, Monte Carlo simulation). About 9.4% (140) of promoters and 8.6% (129) of gene bodies belonging to maturation upregulated genes lose H3K27me3 during the P14 to M2 transition (Figure 1C). The H3K27me3 and transcriptional dynamics between P14 and M2 indicate a role for PRC2 proteins in regulating maturation gene expression.

EZH1 and EZH2 restrain premature postnatal hepatic maturation

To assess the role of PRC2 repression in postnatal hepatocytes, we performed RNA-seq on *Alb-Cre/Alb-Cre; Ezh1^{-/-}; Ezh2^{flox/flox}* (“*Ezh1/2*”) hepatocytes, in which both histone methyltransferases for H3K27me3 are ablated. Loss of *Ezh1* or *Ezh2* alone does not result in H3K27me3 loss (Supplementary Figure 2A). *Ezh1* single knockout have transcriptomes highly similar to *Wt*, with only 203 differentially expressed genes (Supplementary Figure 2B, Supplementary Tables 2, 3). With a homozygous *Alb-Cre* transgene, whose expression starts around birth, P14 is the earliest age we observed quantitative H3K27me3 loss in *Ezh1/2* hepatocytes; hence this time was used to minimize secondary effects (Supplementary Fig 2CD). Notably, P14 *Ezh1/2* hepatocytes have expression profiles between *Wt* hepatocytes P14 and M2 hepatocytes (Supplementary Figure 1D). Strikingly, of the 1215 genes upregulated during hepatocyte maturation at M2, 263 (22%) are prematurely upregulated in P14 *Ezh1/2* hepatocytes (Figure 3A, Supplementary Table 3). Of the 2011 genes downregulated in hepatocyte maturation, 128 (6.3%) are prematurely downregulated in P14 *Ezh1/2* hepatocytes. Expression changes do not preferentially affect zoned genes (Supplementary Figure S2E). Taken together, the

results indicate that the PRC2 proteins EZH1/2 normally restrain maturation of postnatal hepatocytes until the appropriate postnatal time.

Non-hepatocyte lineage genes are derepressed in P14 *Ezh1/2* hepatocytes

There are 665 genes upregulated and 90 genes downregulated in P14 *Ezh1/2* hepatocytes that are not normally changed during P14 to M2 maturation (Figure 3A, Supplementary Table 3). Upregulated genes include many transcriptional regulators, DNA-binding factors, and genes expressed in non-liver tissues, but GO analysis did not reveal specific biological pathways that were deregulated (Supplementary Table 5). These data are consistent with work from others that Polycomb proteins broadly repress alternate lineage-specific identity regulators (Boyer et al., 2006; Ezhkova et al., 2011; Margueron et al., 2008; Shen et al., 2008), but are insufficient upon deletion to result in outright hepatocyte identity change at this postnatal maturation stage.

A subset of maturation and alternative lineage genes are repressed by EZH1/2

As expected, given the repressive functions of PRC2, the 263 prematurely upregulated “late maturation” genes have high H3K27me3 at promoters and gene bodies at P14, but not at M2, as compared to genes only upregulated in maturation (Figure 2B). There are 148 genes upregulated in P14 *Ezh1/2* hepatocytes (Figure 2A), although they are normally downregulated in maturation and H3K27me3-marked, indicating that these are “early maturation” genes that are downregulated in the course of maturation by PRC2. The 665 “alternative lineage” genes upregulated in P14 *Ezh1/2* mutants have high P14 and M2 H3K27me3 levels and the majority are lowly expressed in *Wt* hepatocytes (Supplementary Figure 2H). These data support a model where PRC2/H3K27me3 repression is used to

repress early and late maturation genes at the appropriate times and non-hepatocyte alternative lineage genes.

P14 and M2 H3K27me3 was low at liver enhancers, regardless of whether the enhancers were centered by DNase hypersensitivity or if windows were varied around enhancer centers from 200 bp to 5 kb (Supplementary Figure 2G). Specifically, H3K27me3 levels were low at enhancers associated with genes upregulated in P14 *Ezh1/2*, except for the early maturation genes, which had only 33 associated enhancers (Figure 2B). From this general lack of H3K27me3 at liver enhancers, we conclude that the *Ezh1/2* premature maturation phenotype is not functioning primarily through altered repression of enhancers.

In contrast, of the 263 prematurely upregulated maturation genes, 99 have P14 H3K27me3 at promoters and 85 have P14 H3K27me3 on gene bodies, for a total of 117 genes with at least one type of H3K27me3 marking. These genes include *Slc13a5*, which plays a key role in importing citrate into liver cells (Gopal et al., 2007), *Pcsk9*, which is associated with liver cholesterol (LDL) uptake (Ruscica et al., 2016), and *Cyp26a1*, a key enzyme in the clearance of retinoic acid from the liver (Thatcher et al., 2010) (Figure 2C, Supplementary Figure 2F). In comparison, *Pax3*, a silent muscle gene, retains H3K27me3 marking at the promoter and was not upregulated in maturation. GO analysis on the 117 H3K27me3-marked genes did not reveal grouped biological process (data not shown). Taken together, we conclude H3K27me3 is used to restrain expression of genes in postnatal hepatocytes specifically at promoters and gene bodies, and minimally so at enhancers.

srHC-seq reveals that *Ezh1/2* sensitive genes have euchromatic and bivalently-marked promoters

To assess whether there are global changes in chromatin compaction in hepatic maturation, we characterized P14 and M2 *Wt* hepatocytes using an enzyme- and antibody-independent assay termed sonication-resistant heterochromatin sequencing (Supplementary Figure 3A). srHC-seq utilizes the physical property of crosslinked chromatin to be differentially sensitive to sonication; that is, structurally compact, heterochromatic regions are sonication-resistant while structurally open, euchromatic regions are sonication-sensitive (Becker et al., 2017). srHC-seq involves fractionation of large and small DNA fragments and analyzes the ratio of the two to identify both heterochromatic and euchromatic regions.

srHC-seq was highly reproducible in replicates (Supplementary Table 2, Supplementary Figure 3BC). We plotted srHC data as $\log_2(\text{large fragments}/\text{small fragments})$, with heterochromatic regions as $y > 0$ (Figure 3A, red) and euchromatic regions as $y < 0$ (Figure 3A, green). As expected, the silent, non-hepatocyte gene *Zfp936* has a heterochromatic profile in liver and genes at the highly expressed, hepatocyte-specific *Alb/Afp/Afm* locus have euchromatic profiles. Importantly, srHC-seq scores for promoters and gene bodies shows the inverse correlation between expression and chromatin compaction holds true genomically (Figure 3B).

In contrast to silent and highly expressed loci, *Cux2* expression is temporally- and sex-specific (Conforto et al., 2012) and srHC profiles reflect these dynamics (Figure 3A). While sex-specific differences were ascertainable, the differences represent only a fraction of the genome. We therefore merged male and female srHC data for the remaining analyses.

Both P14 and M2 *Wt* hepatocytes have about 1200 Mb of srHC heterochromatin and 1100 Mb of euchromatin, with H3K27me3 occurring in both domains (Figure 3CD, Supplementary Figure 3DE). We did not observe a global increase of euchromatin in P14 *Ezh1/2* hepatocytes or by confocal microscopy of DAPI staining of 6-week-old *Ezh1/2* samples (data not shown). We conclude that large scale euchromatic and heterochromatic domains are generally stable in postnatal hepatic maturation, while a subset of genes are dynamic.

While the PRC2 complex is classically thought to repress expression by eliciting chromatin compaction (Simon and Kingston, 2009), Polycomb-bound or -marked chromatin can be accessible to binding by some factors and transcribed (Becker et al., 2017; Beisel and Paro, 2011; Breiling et al., 2001; Dellino et al., 2004; Hawkins et al., 2010; Ku et al., 2008; Trojer and Reinberg, 2007). PRC2 loss leads to derepression of only a fraction of H3K27me3-marked genes in diverse tissues, with promoter H3K4me2/3 predicting derepression (Bae et al., 2015; Ezhkova et al., 2011; Jadhav et al., 2016). To investigate the basis by which a subset of H3K27me3-marked promoters genetically respond to *Ezh1/2* loss, we plotted srHC-seq profiles at 8121 silent and 2221 highly expressed in P14 and M2 *Wt* animals (Figure 4A). Significantly, promoters of genes from all three classes of genes upregulated in P14 *Ezh1/2* hepatocytes are already euchromatic in P14 *Wt* hepatocytes, including those with promoter H3K27me3 signal (Figure 4A, note extensive green in P14 *Wt*). To focus on apparent direct PRC2 targets, we plotted srHC signal of genes with H3K27me3-marked promoters in P14 *Wt* samples (Figure 4B). These promoters become further euchromatic in *Ezh1/2* hepatocytes and the open regions increase in width (Figure 4B). Using published P12 liver ChIP data (Alpern et al., 2014), we found that the three classes of P14 *Ezh1/2* upregulated genes have promoters with

low levels of H3K4me3, RNA Polymerase II, and general transcriptional factors (Figure 4A, Supplementary Figure 4B).

By plotting signal at promoters of genes that are not differentially expressed in maturation or in P14 *Ezh1/2* hepatocytes and do have P14 promoter H3K27me3 (Figure 4A, “remaining P14 H3K27me3+ promoters”), we found that H3K4me3 does indeed help predict genes that become upregulated in *Ezh1/2* mutants, as does RNAP2, general TFs, and a pre-existing euchromatic chromatin state (Figure 4A, Supplementary Figure 4B). Thus, promoters that are open and poised yet restrained by PRC2 are regulated during hepatocyte maturation.

Chronic *Ezh1/2* loss leads to liver damage

It has previously been reported that *Ezh1/2* in liver leads to chronic liver damage and fibrosis in mice by 8 months and increased sensitivity to liver damaging agents (Bae et al., 2015). We observe that apoptosis occurs 4.6-fold more in *Ezh1/2* hepatocytes by 1 month postnatal with very rare ductular reactions (Figure 5A). Yet by M2, but not at M1, liver fibrosis can be detected by Sirius Green/Fast Red staining (Figure 5B, Supplementary Figure 5A). Additionally, ductular responses can occur by H&E and CK19 staining by M2, with some larger diameter, malformed ducts than were reported previously (Figure 5B, Supplementary Figure 6A). Macroscopic regenerative nodules occur in nearly half of *Ezh1/2* livers by M2, visible in H&E sections or in the most severe cases, by the naked eye as lumps on the liver (Figure 5C). While the *Ezh1/2* model loses H3K27me3 in nearly 100% of hepatocytes at P14 (Supplementary Figure 2C), by M2 these nodules stain for H3K27me3 similar to *Wt* and also stain for EZH2 (Figure 5DE, Supplementary Figure 6AC), which was not observed previously in Bae et al. (Bae et al., 2015). As the liver is a regenerative organ and nodular hepatocytes in *Ezh1/2* livers have increased proliferation

(Supplementary Figure 6C), we conclude that the EZH2⁺ nodules originate by selection for rare cells which escape *Alb-Cre* recombination. Thus, loss of PRC2-based repression results in loss of hepatic homeostasis, chronic liver damage, and fibrosis.

Genes involved in liver fibrosis are primed by euchromatic H3K27me3⁺ promoters

A 232 gene signature predicting fibrosis before histopathological detection was identified using a non-alcoholic steohepatitis (NASH) mouse model, and the relevance to humans was confirmed by finding 71 of 123 human NASH genes in the mouse 232 gene datasets (van Koppen et al., 2018; Teufel et al., 2016). RNA-seq on liver biopsies from a mixed cohort of chronic liver disease patients with Hepatitis C and/or fatty liver disease identified 121 genes upregulated in advanced human liver fibrosis as compared to early fibrosis (Ramnath et al., 2018). We found that genes from both datasets are normally downregulated in postnatal hepatic maturation, are upregulated in P14 and M2 *Ezh1/2* hepatocytes, and have the predictive euchromatic H3K27me3⁺/H3K4me3⁺ promoter chromatin state that we identified as sensitizing genes to *Ezh1/2* loss (Figure 6A). Genes significantly upregulated in P14 *Ezh1/2* samples in mouse and human datasets were each 2.6-fold more likely to have P14 H3K27me3⁺ promoters (permutation test, $p=0.003$ and $p=0.009$). Included in fibrosis gene sets with H3K27me3 promoter-marked genes are *Fbn1*, an inflammation and chemotaxis gene that is also upregulated in response to damaging chemicals (Ippolito et al., 2016), *Fstl1*, which is upregulated in humans with HCV-induced fibrosis and steatosis (Murphy et al., 2016), and *Col1a1*, which is upregulated in mice and humans with liver fibrosis (Dattaroy et al., 2015) (Figure 6B, Supplementary Figure 7BC). Interestingly, genes proposed to be involved hepatocyte epithelial-mesenchymal transition (EMT) in response to liver damage include *Tgfb1*, *Vim*,

and *S100a4* (Choi and Diehl, 2009) also have H3K27me3-marked promoters and may be upregulated in M2 *Ezh1/2* hepatocytes (Figure 6B, Supplementary Figure 7BC). Together, the failure to repress fibrosis-related genes starting at two weeks postnatal is associated with the 2-month-old fibrotic phenotype in *Ezh1/2* livers.

2.7. Discussion

To be effective therapeutically, newly generated hepatocyte-like cells must perform the complex metabolic functions of native hepatocytes that are naturally induced postnatally (Cui et al., 2012; Fouts and Adamson, 1959; Treluyer et al., 1996), but these late maturational functions are limited in many hepatocyte-like cells (Chen et al., 2012; Duan et al., 2010; Roelandt et al., 2013; Si-Tayeb et al., 2010; Song et al., 2009). Here we define a P14 to M2 transitional hepatocyte profile with more than 3000 transcripts exhibiting differential expression. The RNA maturation dataset can be a useful resource for assessing the stage of hepatocytes generated from stem cells or by directed reprogramming. Perinatal loss of EZH1/2 in hepatocytes leads to premature differentiation, suggesting that modulation of Polycomb components may enhance maturation in protocols for generating new hepatocytes.

In our study, altering PRC2 repression led to impaired liver function and fibrosis, and previous reports have shown impaired ability to respond to liver damaging agents (Bae et al., 2015). Interfering with splicing factors alters liver maturation in mice and also leads to susceptibility to liver damage (Bhate et al., 2015). Considering the rising worldwide prevalence of NASH and hepatitis B (Wong and Huang, 2018) and that fibrosis is a key indicator of chronic liver injury of any etiology (Jung and Yim, 2017), the common thread of impaired maturation predisposing the adult liver to damage suggests that

studying mechanisms of hepatic maturation may help us identify environmental factors that alter maturation in humans and predispose them to liver diseases. Here we have identified a H3K27me3⁺/H3K4me3⁺ euchromatic promoter signature that primes many genes to be upregulated in response to maturation or other signaling events, and we propose that this is a chromatin-level mechanism leading to liver fibrosis in the *Ezh1/2* model. Given the euchromatic state of these promoters, it may be that compaction-independent repressive activities of Polycomb proteins, such as PRC2 inhibiting elongation (Ardehali et al., 2017) and PRC1 interfering with RNAPII recruitment (Chopra et al., 2009; Tie et al., 2016), are key to regulating transcription in postnatal hepatocytes, though a full catalog of precise mechanisms by which Polycomb proteins repress transcription is still being worked out. It will be interesting to see how agonists or antagonists of different aspects of Polycomb Protein function may help reverse fibrosis, considering that there is growing evidence of fibrotic reversal after treatment for hepatitis B, C, and autoimmune hepatitis (Jung and Yim, 2017).

There is conflicting evidence suggesting that hepatocytes can acquire a fibroblastic phenotype and expression of mesenchymal markers through EMT during liver fibrosis or in response to TGFβ treatment (Choi and Diehl, 2009; Dooley et al., 2008; Nitta et al., 2008; Taura et al., 2010). We find that expression of these mesenchymal markers is slightly upregulated in M2 *Ezh1/2* hepatocytes, but expression is highly variable (Supplementary Figure 6). This may reflect variability of multiple different sources, including variable levels of liver damage and fibrosis, variable contributions of H3K27me3⁺ nodular hepatocytes, and variable numbers of cells undergoing EMT at possibly different stages of EMT. The mesenchymal genes tend to be marked by H3K27me3 and have euchromatic promoters (Figure 6, Supplementary Figure 7). While we cannot conclude from these results whether EMT is occurring, it will be interesting to

assess whether PRC2 repression maintains an epithelial state in hepatocytes by repressing mesenchymal genes and whether liver damage elicits their derepression.

Our results emphasize the role of PRC2 proteins in regulating maturation and how that may affect fibrosis, but not all maturation genes in our dataset are prematurely up- or downregulated in P14 *Ezh1/2* mutants. There must be other mechanisms that regulation maturation genes in response to changing diet, microbiota, and sexual maturation during the two week to two month period. These results highlight the need to identify maturation factors, both for the purposes of enhancing in vitro hepatic maturation and for understanding factors which predispose humans to disease.

2.8. Main Figures

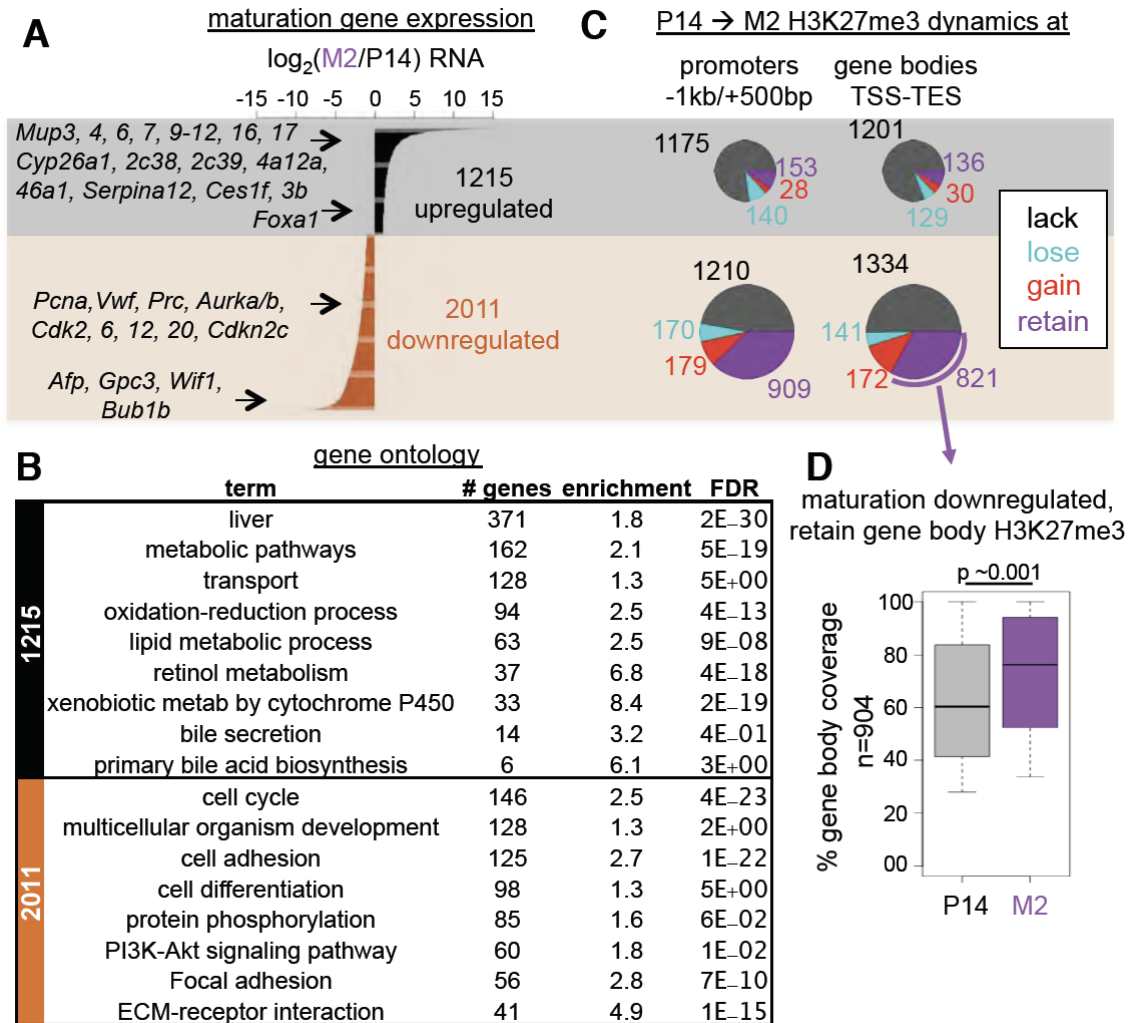


Figure 1. Postnatal hepatic maturation during the P14 to M2 transition involves differential expression of thousands of genes and H3K27me3 dynamics.

- A)** $\log_2(\text{M2/P14})$ fold change) for genes differentially expressed in maturing hepatocytes.
- B)** Gene Ontology for genes up and downregulated from P14 to M2 in *Wt* hepatocytes.
- C)** H3K27me3 dynamics at genes up- (top, grey shading) and downregulated (bottom, orange shading) in maturation
- D)** Percent H3K27me3 gene body coverage of “retain” H3K27me3 genes. Monte Carlo simulation and estimated p-value. Whiskers: 5th-95th percentiles.

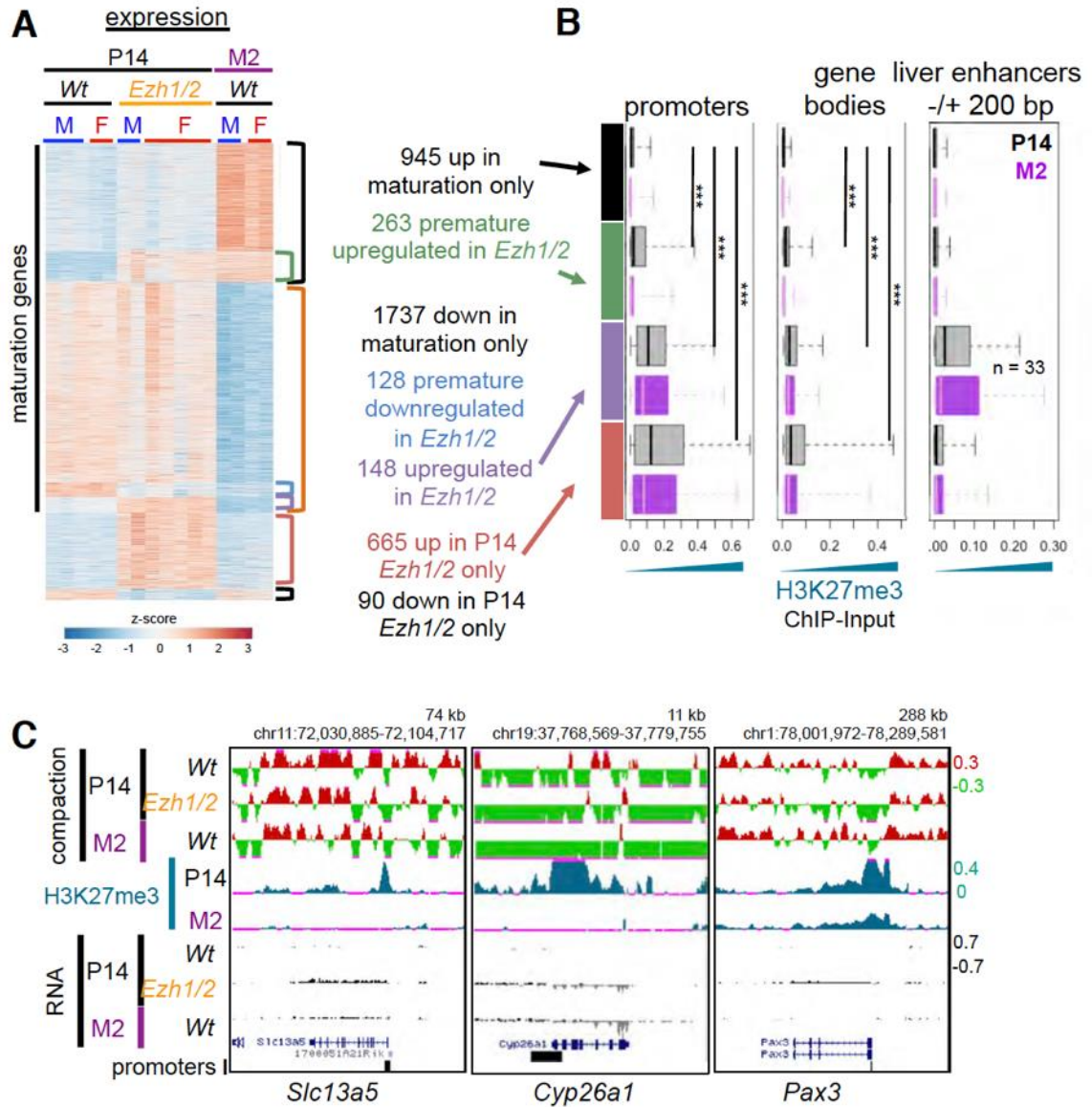


Figure 2. PRC2 proteins EZH1/2 restrain postnatal hepatic maturation.

- A)** Relative expression of genes differentially expressed in maturation or in P14 *Ezh1/2* hepatocytes.
- B)** P14 and M2 H3K27me3 density at promoters, genes, or enhancers. Whiskers: 5th-95th percentiles. Wilcoxon rank-sum test.
- C)** srHC-seq (heterochromatic-enriched in red, euchromatic-enriched in green), H3K27me3, and RNA signal at genes prematurely upregulated in P14 *Ezh1/2* hepatocytes (*Slc13a5*, *Cyp26a1*) and a control gene (*Pax3*).

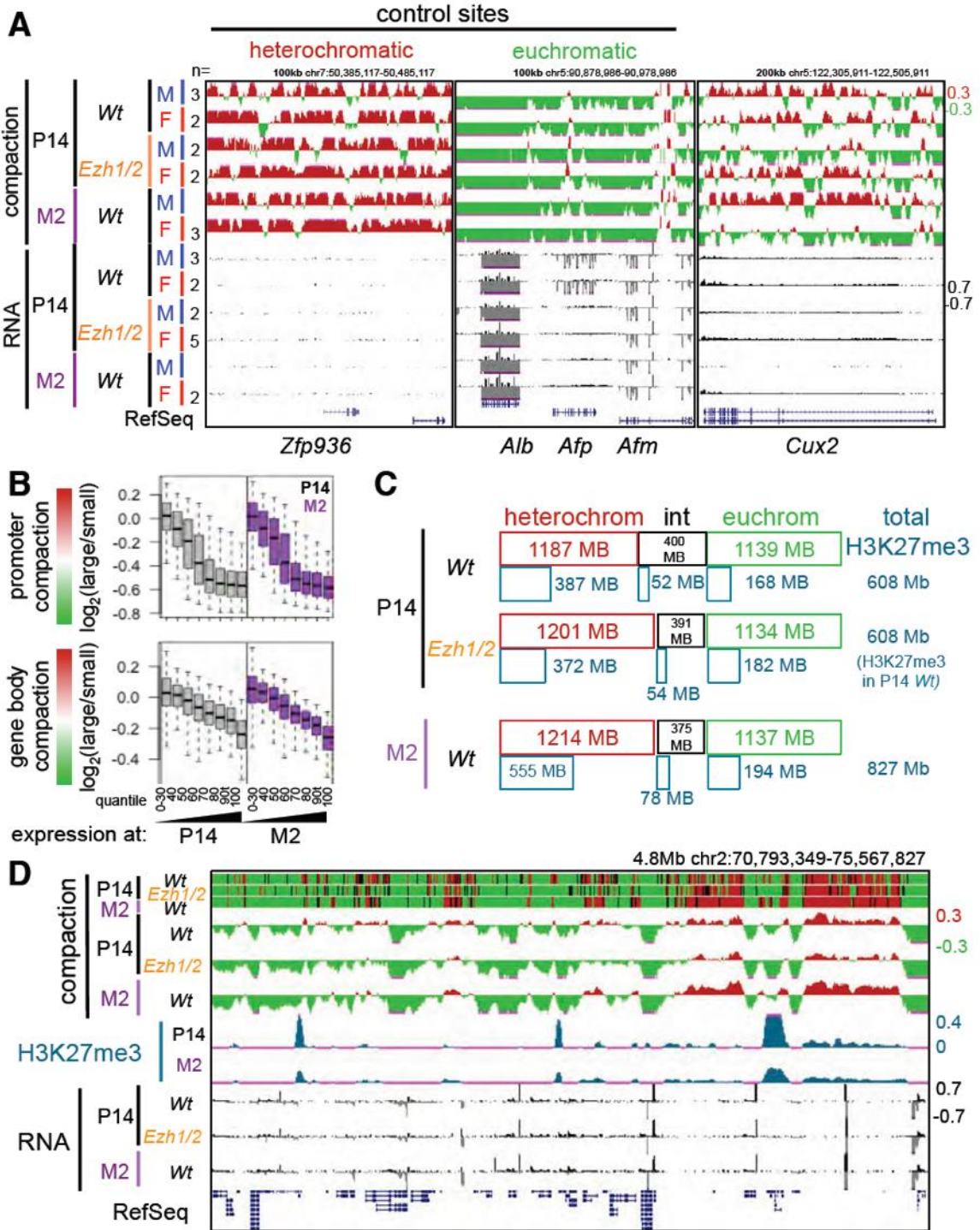


Figure 3. Stable global chromatin compaction in maturation and P14 *Ezh1/2*

- A)** srHC-seq and RNA signal for positive controls sites for heterochromatin, *Zfp936*, euchromatin, *Alb* locus, or a dynamic gene, *Cux2*.
- B)** srHC-seq scores at promoters and gene bodies for genes binned into expression quantiles. 0-30 represents genes without any RNA signal. Note that with increasing expression, srHC scores become more euchromatic. Whiskers: 5th-95th percentiles.
- C)** Megabases of the genome called as heterochromatic, intermediate, or euchromatic and the overlap with H3K27me3 domains.
- D)** srHC-seq domains and signal, H3K27me3, and RNA. Note that H3K27me3 occurs in both heterochromatic (red) and euchromatic (green) regions.

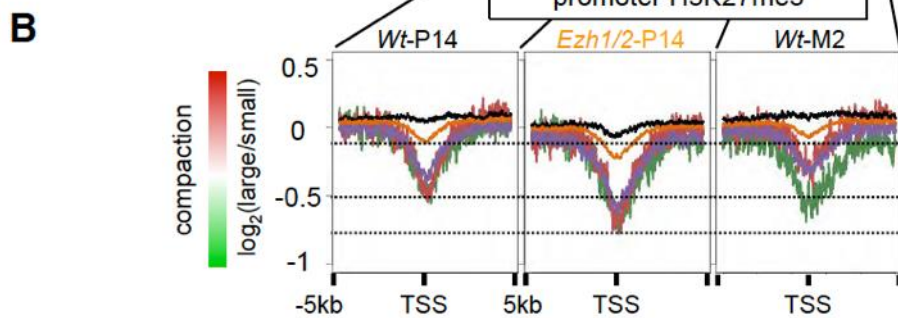
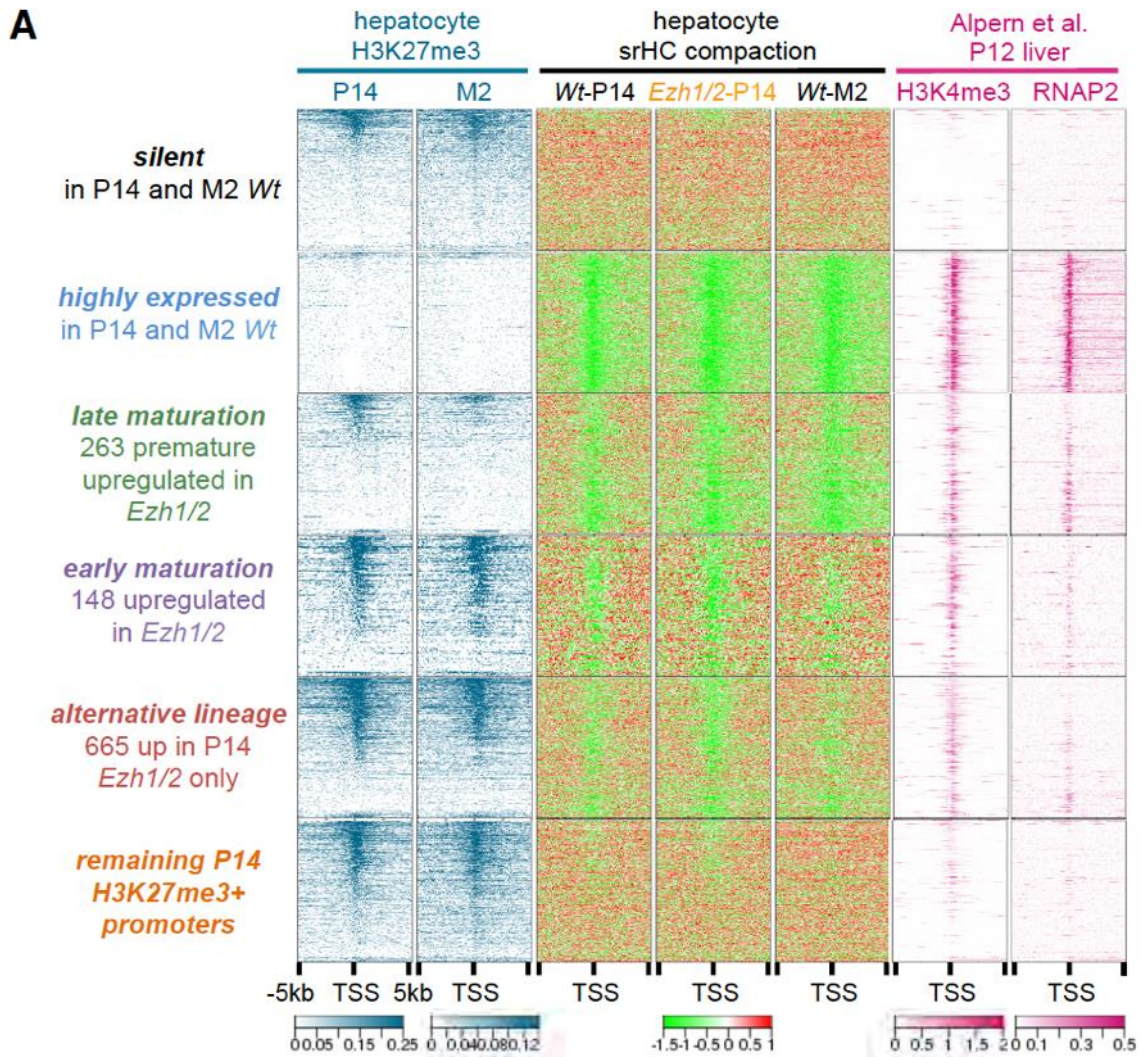


Figure 4. srHC-seq reveals that *Ezh1/2* sensitive genes have euchromatic and bivalently-marked promoters

- A)** ChIP and srHC-Seq signal around transcriptional start sites (-/+ 5kb) for silent genes, genes highly expressed in P14 and M2 (top 10% in expression in P14 and M2), three classes genes upregulated in P14 *Ezh1/2*, and the remaining P14 H3K27me3+ promoters. Note that the three classes of genes upregulated in P14 *Ezh1/2* hepatocytes have euchromatic promoters (green srHC signal) in P14 *Wt* hepatocytes. In comparison promoters that have H3K27me3 but are not upregulated (bottom, orange) have heterochromatic signals (red srHC signal) in P14 *Wt* hepatocytes.
- B)** srHC-seq metaplots for panel (A) groups except highly expressed genes. Filtered for genes with P14 H3K27me3-marked promoters. Note that late maturation genes, early maturation genes, and alternative lineage genes (green, purple, and red lines) have more euchromatic signal in P14 *Wt* hepatocytes than the genes which are not upregulated in P14 *Ezh1/2* and have H3K27me3+ promoters (orange line).

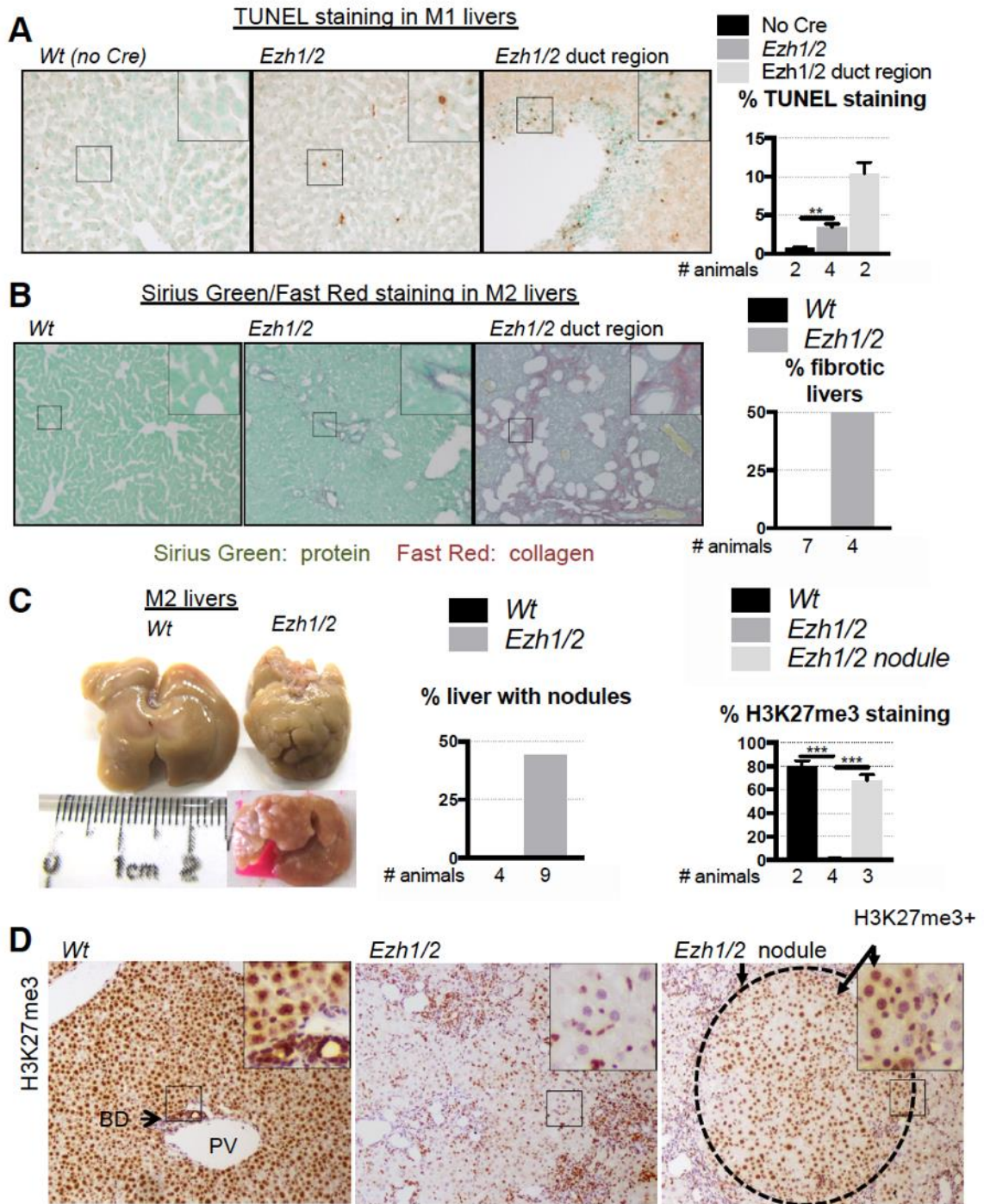


Figure 5. *Ezh1/2* loss leads to chronic liver damage.

- A)** Apoptosis as assayed by TUNEL staining.
- B)** Fibrosis as assayed Sirius Green (total protein) and Fast Red (collagen) staining.
- C)** Examples of a normal *Wt* liver and 2 *Ezh1/2* livers with macroscopic liver nodules. Quantification of animals with liver nodules as assessed by histology.
- D)** H3K27me3 immunohistochemistry in M2 livers. Regions inside and outside of regenerative nodules are outlined by the dotted circle. *Wt*: 2034 hepatocytes counted. *Ezh1/2*: 2440 hepatocytes counted. *Ezh1/2* nodule: 1056 cells counted in 7 nodules. Two-sided student's t-test with standard error.

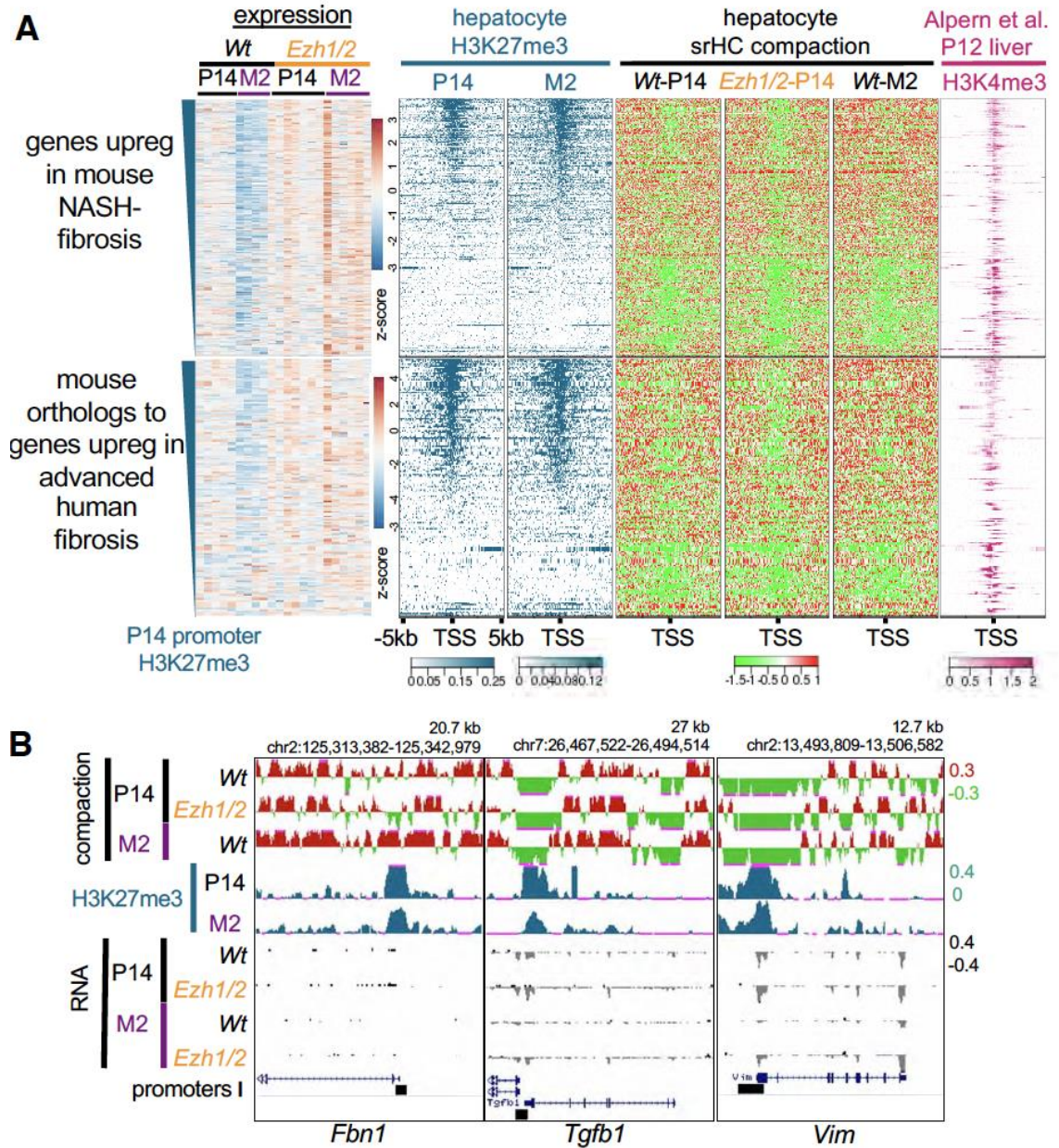
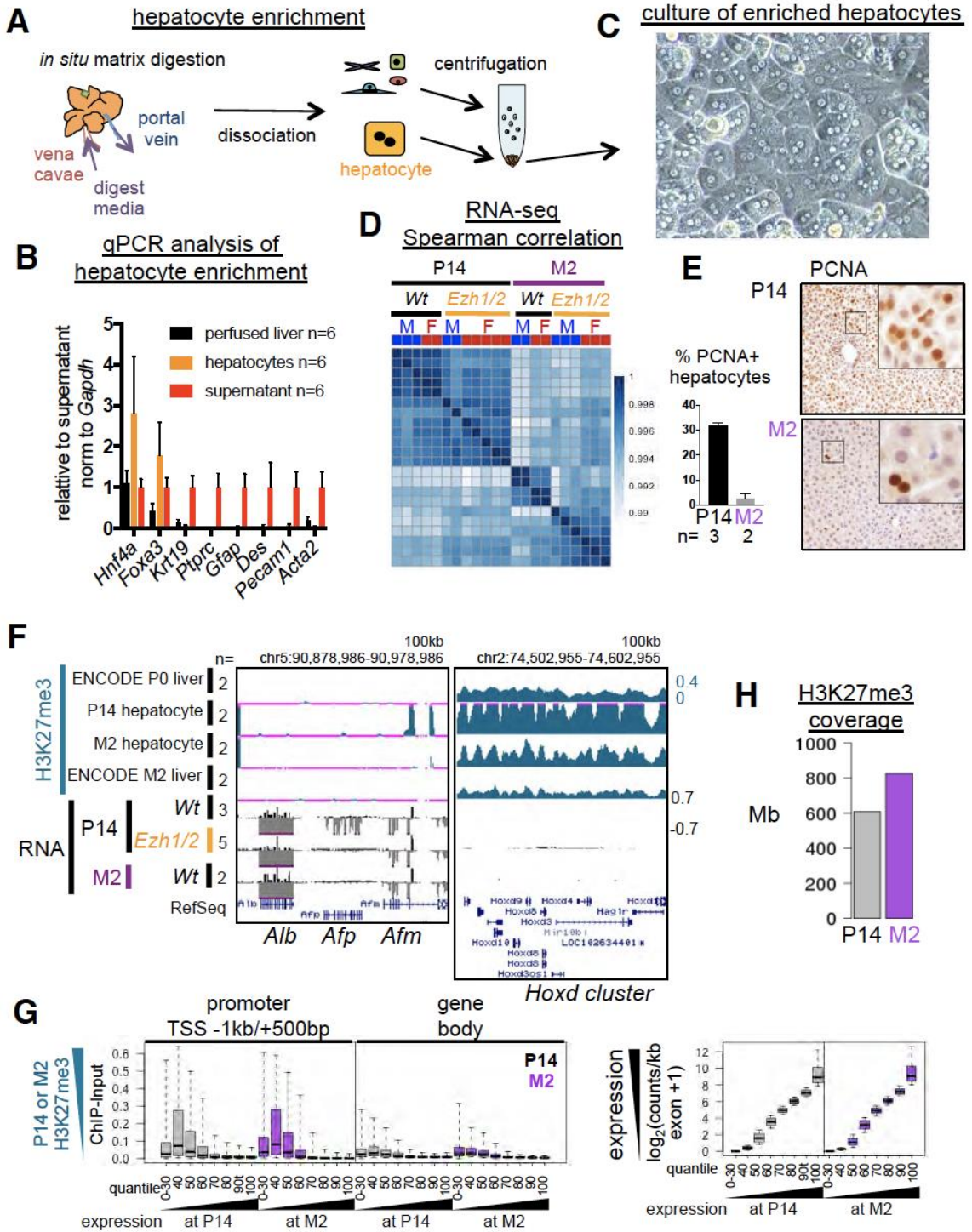


Figure 6. PRC2 represses liver fibrosis signature genes

A) Genes upregulated in a murine model of NASH-related fibrosis (top) and genes upregulated in human advanced fibrosis patients as compared to low level fibrosis (bottom). Murine hepatocyte expression, H3K27me3 and srHC-seq, and H3K4me3 signal around transcriptional start sites \pm 5 kb. Note that many fibrosis genes have H3K27me3+ promoters that are also euchromatic (green srHC signal) and have promoter H3K4me3. M2 *Ezh1/2* RNA-seq represents a mix of hepatocytes from inside and outside of nodules.

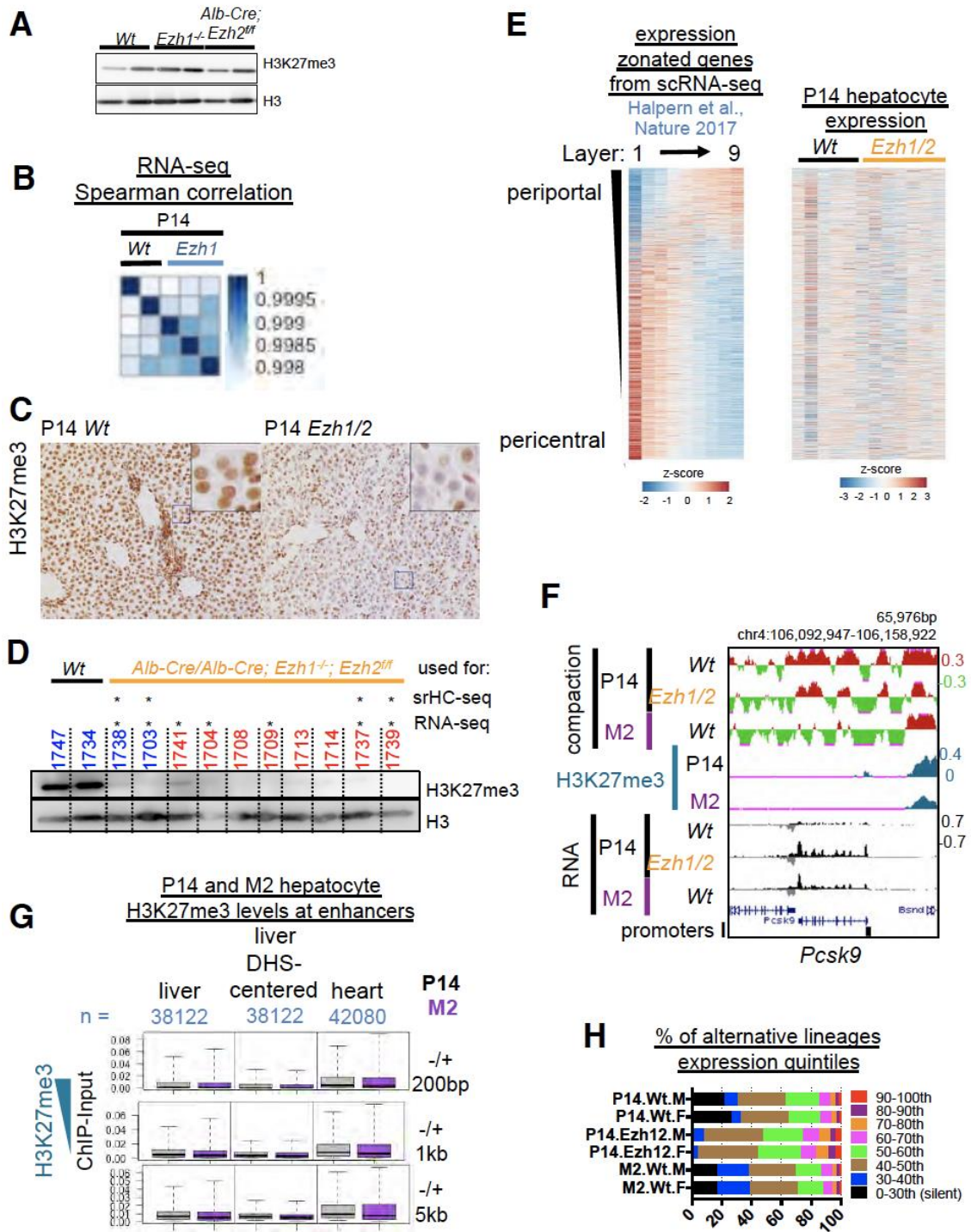
B) srHC-seq, H3K27me3, and RNA signal at fibrosis-related or EMT-related genes.

2.9. Supplementary Figures



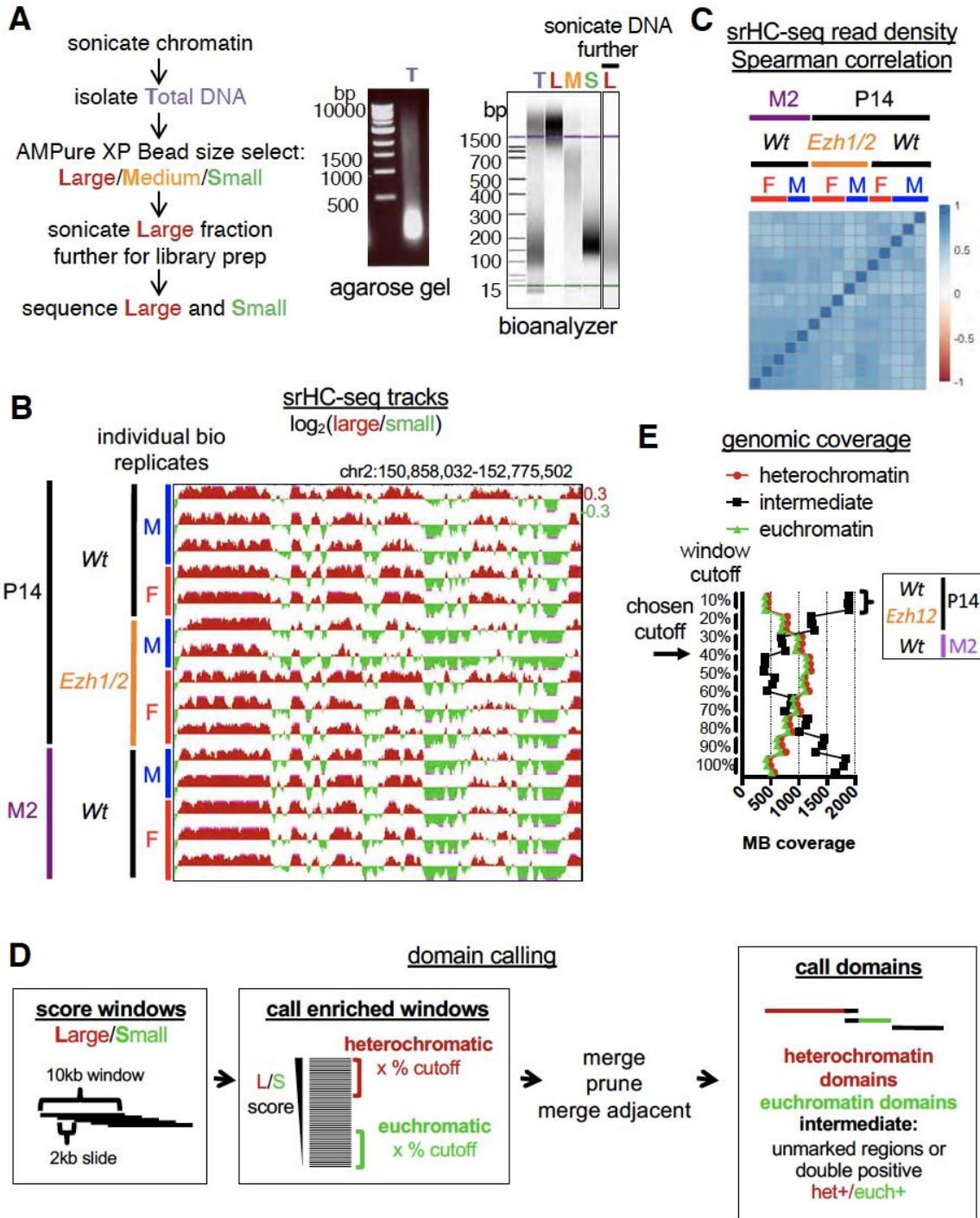
Supplementary Figure 1: Hepatocyte enrichment strategy and RNA-seq and ChIP-seq quality control

- A)** Hepatocyte enrichment strategy. The liver ECM is digested *in situ* by perfusion, the liver is dissociated, filtered through a 100 μ M filter, and hepatocytes pelleted by centrifugation. Remaining cells are pelleted from the supernatant.
- B)** Expression of markers of hepatocytes (*Hnf4a*, *Foxa1*), biliary cells (*Krt19*), blood (*Ptprc*, not red blood cells or platelets), stellate cells (*Gfap*, *Desmin*), and endothelial cells (*Pecam1*, *Acta2*) by RTqPCR from whole perfused liver, pelleted hepatocytes, or pelleted supernatant.
- C)** Cells from the hepatocyte enrichment strategy have hepatocyte morphology when plated on collagen-coated plates in William's E with penn/strep, 10% FBS, 2 mM L-glutamine at 37°C.
- D)** Spearman correlation for DESeq2-normalized genic expression.
- E)** Proliferation as assayed PCNA staining. Error bars with standard error.
- F)** P14 and M2 hepatocyte H3K27me3 at negative (*Alb/Afp/Afp*) and positive (*Hoxd* cluster) control sites compared to ENCODE consortium P0 and M2 H3K27me3 liver ChIP data and expression in hepatocytes
- G)** P14 and M2 hepatocyte H3K27me3 density at promoters and gene bodies, and expression for genes binned into expression quantiles (female). 0-30th represent silent genes. Whiskers: 5th-95th percentiles.
- H)** Genomic coverage of H3K27me3 domains



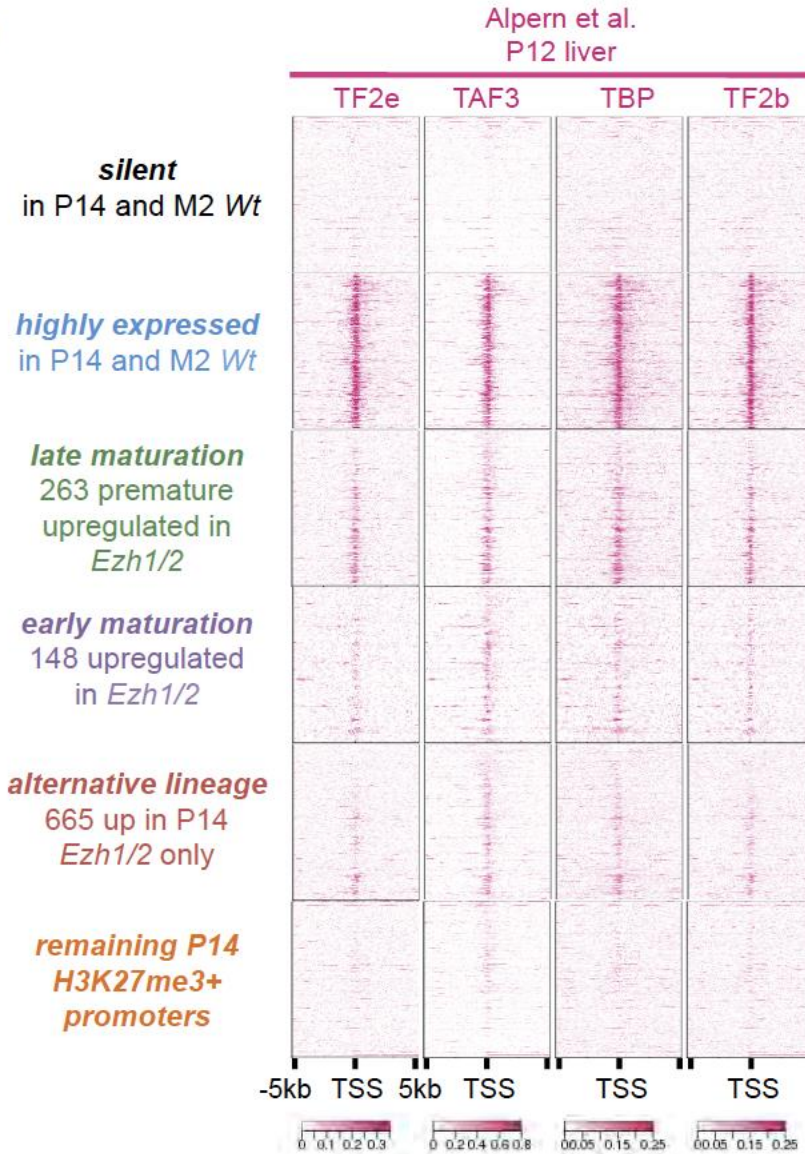
Supplementary Figure 2. Confirmation of H3K27me3 loss in hepatocytes, confirmation of minimal *Ezh1* germline knockout effect, and RNA expression analysis of zoned and alternative lineage genes

- A)** H3 and H3K27me3 Western on M2 hepatic nuclear extracts. Note that single KO of *Ezh1* or *Ezh2* does not cause H3K27me3 loss.
- B)** Spearman correlation for DESeq2-normalized genic expression comparing P14 *Wt* and *Ezh1* single knockout hepatocytes.
- C)** H3K27me3 staining in P14 *Wt* and *Ezh1/2* hepatocytes. 1/500 AM39155
- D)** Screening for H3K27me3 loss by Western on P14 hepatocyte whole cell extracts for H3 (Millipore 05-928) and H3K27me3 (Active Motif 39155). Samples used for RNA-seq and srHC-seq are starred.
- E)** Zoned genes and expression in layers 1-7 from Halpern et al., 2017 (left). Expression of the same genes in the same order in P14 hepatocytes (right).
- F)** srHC-seq (heterochromatic-enriched in red, euchromatic-enriched in green), H3K27me3, and RNA signal at genes prematurely upregulated in P14 *Ezh1/2* hepatocytes (*Pcsk9*).
- G)** Hepatocyte H3K27me3 density at liver enhancers, liver enhancers centered on a DNase hypersensitive site, and heart enhancers. The defined size of an enhancer from its center is varied from 200 bp to 1 kb to show that the results are reproducible. Whiskers: 5th-95th percentiles.
- H)** Percentage of the 665 alternative lineage genes in each expression quintile. Note that they are mostly silent or very lowly expressed.



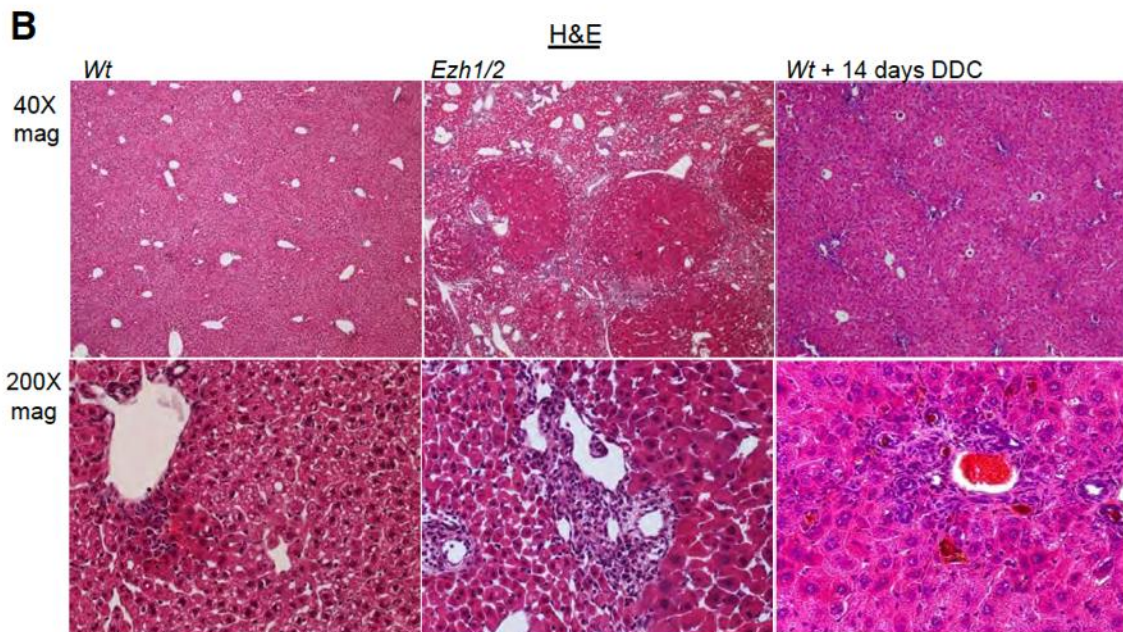
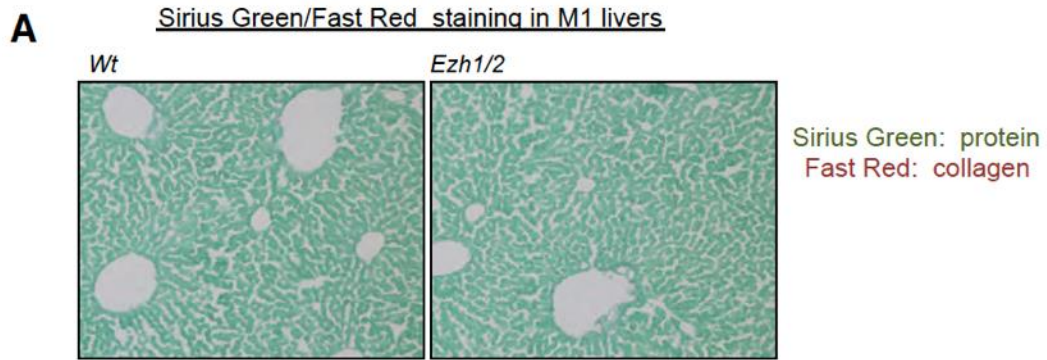
Supplementary Figure 3. srHC-seq overview and method validation

- A)** High cell number srHC-seq overview. Total DNA isolated from crosslinked, sonicated hepatocyte chromatin is shown on an agarose gel for comparison to Total DNA on a Bioanalyzer DNA 1000 chip. Representative profiles for Large, Medium, or Small DNA fractions and Large DNA that has been sonicated to a smaller size for library prep considerations.
- B)** srHC-seq (heterochromatic-enriched in red, euchromatic-enriched in green) for all biological replicates.
- C)** Spearman correlation of srHC-seq signal for non-overlapping 10kb windows across the genome.
- D)** Compaction domain calling overview. Conceptually, calling srHC-seq domains involved scoring the average large/small fragment score for 10 kb windows with a two kb slide, then using high and low score cutoffs to get heterochromatic- and euchromatic-enriched windows, merging windows and pruning steps. With increasingly lax enriched window cutoffs, an increasing portion of the genome is called as heterochromatic or euchromatic, until the 50% cutoff, where het+/euch+ double positive regions that are reclassified as intermediate takeover.

A

Supplementary Figure 4. Analysis of general transcription factors at transcriptional start sites at *Ezh1/2*-sensitive gene classes and control gene classes

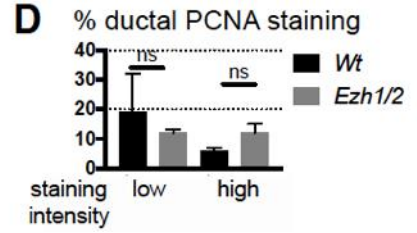
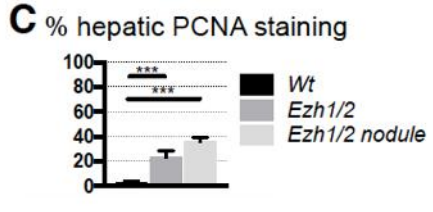
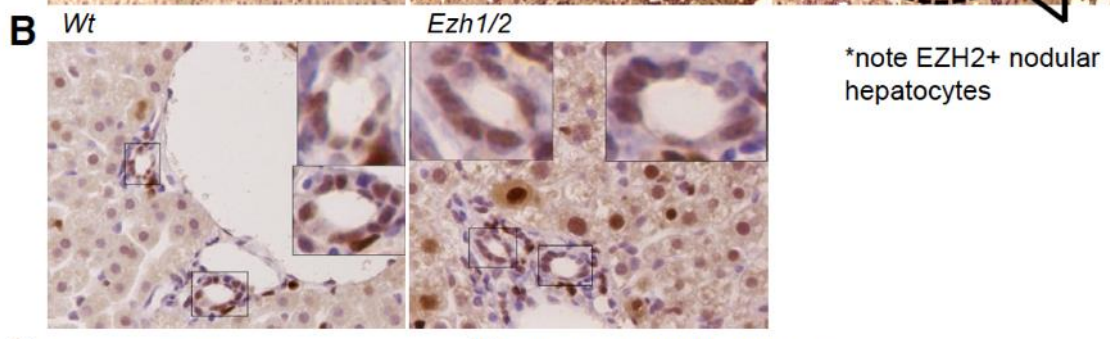
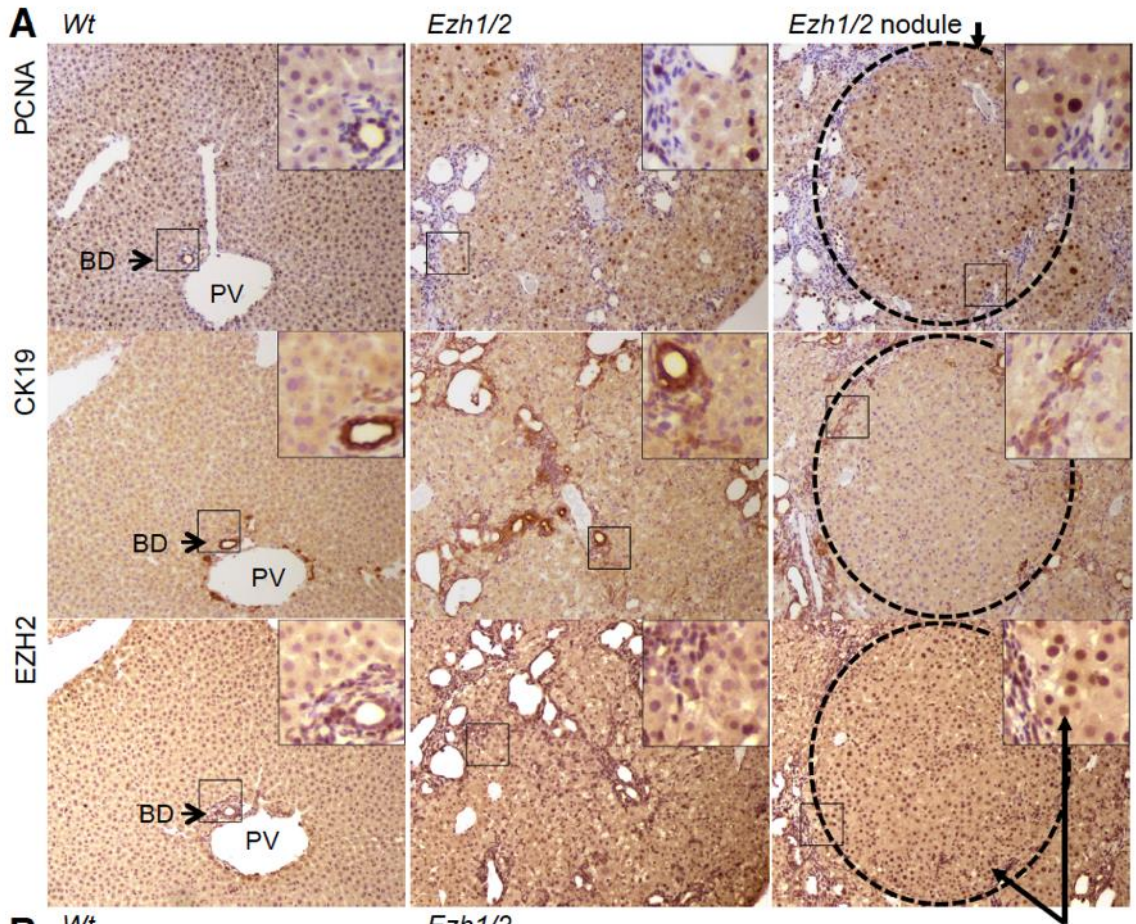
A) P12 liver ChIP signal for general transcription factors around transcriptional start sites (-/+ 5kb) for silent genes, genes highly expressed in P14 and M2 (top 10% in expression in P14 and M2, male and female), three classes of *Ezh1/2* upregulated genes, and other P14 H3K27me3+ promoters (have P14 H3K27me3, but not significantly differentially expressed in maturation or in P14 *Ezh1/2*).



Supplementary Figure 5. Histological analysis of fibrosis in M1 and M2 *Wt* and *Ezh1/2* livers

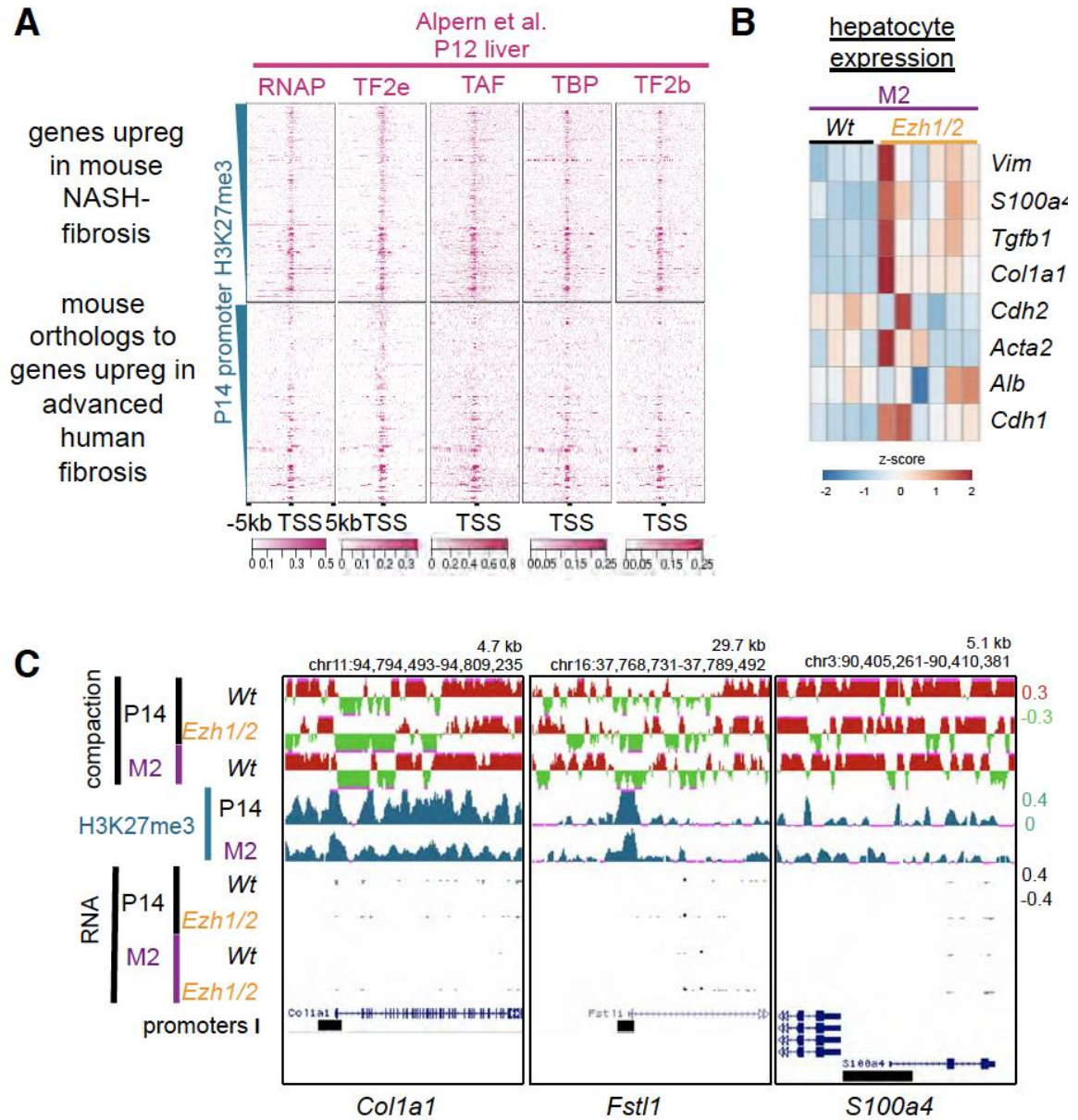
B) Fibrosis as assayed Sirius Green (total protein) and Fast Red (collagen) staining.

C) HE on *Wt* and *Ezh1/2* livers and adult *Wt* mouse liver treated with 14 days DDC. Regenerative nodules in 40X magnification *Ezh1/2* samples present as pink circles.



Supplementary Figure 6. Histological analysis of proliferation, ductal markers, and EZH2 in *Wt* and *Ezh1/2* livers

- A)** Immunohistochemistry for the proliferation marker PCNA, the biliary marker CK19, and H3K27me3 histone methyltransferase EZH2 on M2 *Wt* and *Ezh1/2* livers. Regions inside and outside of regenerative nodules are denoted and (C) quantification of hepatocyte PCNA staining in *Wt* (2063 cells counted from 2 animals), *Ezh1/2* outside of nodules (2624 cells counted from 4 animals), and *Ezh1/2* nodular areas (1299 cells counted in 7 nodules from 2 animals). PV=portal vein. BD=bile duct. Error bars: SEM.
- B)** Immunohistochemistry for proliferation marker PCNA on ducts and (D) Quantification of biliary PCNA staining in *Wt* (447 cells counted in 65 ducts from 2 animals) and *Ezh1/2* (1063 cells counted from 68 ducts from 3 animals).



Supplementary Figure 7. Chromatin state analysis of fibrosis- and EMT-related genes

- A)** Genes upregulated in a murine model of NASH-related fibrosis (top) and genes upregulated in human advanced fibrosis patients as compared to low level fibrosis (bottom). P12 liver ChIP signal for general transcription factors around transcriptional start sites (-/+ 5kb).
- B)** Relative expression of genes involved in putative hepatocyte EMT (*Vim*, *S100a4*, *Tgfb1*, *Acta2*), a marker of mesenchymal cells (*Cdh2*), *Alb*, and a maker of epithelial cells (*Cdh1*).
- C)** srHC-seq (heterochromatin-enriched in red, euchromatin-enriched in green), H3K27me3, and RNA signal at fibrosis-related or EMT-related genes.

2.10. Methods

Liver perfusion and hepatocyte isolation

Mice are anesthetized using isoflurane, the abdominal cavity exposed, venae cavae cannulated, and the portal vein severed. 37°C liver perfusion media (Invitrogen 17701-038) and then liver digest media (Invitrogen 17703-034) are perfused (25 mL for P14, 45 mL for M2). Dissociated livers in William's E are strained through a 100 µm filter (Figure S1A) and pelleted at 50g for 5 min at 4°C. Hepatocyte enrichment was confirmed by depletion of RNAs from contaminating cells types by comparing whole perfused liver, isolated hepatocytes, and supernatant (pelleted at 500g) fractions.

RTqPCR

RNA was isolated from TRIzol, cDNA generated (Biorad 170-8891), and expression analyzed with Power SYBR Green (Thermo 4368577).

Western blotting

Whole cells or nuclei (isolated by douncing in RSB) were resuspended in Buffer C (200 mM Tris pH 7.9, 400 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.8% SDS, PIC (Roche #11873580001), 1mM PMSF). Nuclear extracts were sonicated to reduce the viscosity from high DNA content. Extracts were denatured for 30 min at 99°C in 100mM DTT/Sample Buffer (Thermo NP0007) and run with the NuPAGE system (NP0335,NP0002) at 80V. Wet transfer to PVDF membranes (100V for 3 hr, transfer buffer NP0006) and membranes were blocked for an hour in 5% NFDM-TBST. Primary antibodies were incubated overnight in 5% NFDM-TBST. Secondary antibodies (Santa Cruz Biotechnology sc-2004, sc-2005) were incubated in 5% NFDM-TBST for 1 hour.

Blots were developed using Thermo #34080. Primary antibodies: H3: 1/5000 Millipore 05-928. H3K27me3: 5ug/mL Abcam 6147. H3K27me3: 1/1000 Active Motif 39155.

Chromatin preparation and immunoprecipitation

Hepatocytes were fixed in 25 mL 1% formaldehyde in PBS for 10 min at room temperature, quenched with 2.3 mL 2.5 M glycine for 5 min, pelleted at 4°C at 50g for 5 min, resuspended in 10 (P14) or 20 mL (M2) ice-cold RSB (10 mM Tris pH7.4, 10 mM NaCl, 3 mM MgCl₂, 0.5% NP40, PIC, 0.1 mM PMSF, 1% Triton-X 100), dounced, pelleted for 10 min at 4°C at 100g, and resuspended in 2 mL ice-cold AS sonication-lysis buffer (10 mM Tris-HCl, pH8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine, 1 mM DTT, PIC, 0.1 mM PMSF).

For srHC experiments, samples were sonicated using a Diagenode Biorupter UCD-200 (settings: 30 sec on high, 30 sec off for 30 minutes), Triton-X 100 added to 1%, debris pelleted for 15min at 4°C, and supernatant collected. DNA was extracted from 50 µL chromatin by adding 150 µL TE/1% SDS and decrosslinked overnight at 65°C. Next, 200 µL TE and 8 uL 10 mg/mL RNase A added for 2 hours at 37°C shaking. Next, 4 µL 20 mg/mL Proteinase K was added for 2 hours at 55°C shaking. DNA was extracted by phenol:chloroform:isoamyl separation and ethanol precipitation.

For ChIP experiments, hepatocytes in AS-sonication lysis buffer proceeded to the extensive sonication and ChIP protocol in Nicetto et al., (manuscript accepted, *Science*). P14 and M2 hepatocyte H3K27me3 (Millipore 07-449) libraries were then generated using the ThruPlex DNA-seq kit (Rubicon #R400428). The M2 H3K27me3 ChIP was previously published Nicetto et al. (GSE114198).

RNA-seq

RNA was isolated using TRIzol, polyA-selected (Invitrogen (dT)25-61002), libraries prepped (NEB 7420), and sequenced with 75 bp single-end reads. For P14 *Wt* versus *Ezh1* libraries, libraries were prepped with the NEBNext E6110.

Immunohistochemistry

Mice were prepped as for liver cannulation as described above and blood blanching from the liver with liver perfusion media. Livers were rinsed in PBS, fixed in 4% PFA-PBS for 1 hour at 4°C, washed in PBS, dehydrated, embedded in paraffin, 12 µm sections taken, dried, rehydrated in H₂O, and washed. Antigen retrieval was performed in citrate buffer (10 mM Na Citrate, 0.05% Tween20, pH 6) by microwaving for 15 min. Slides were rinsed in water and PBS, quenched in 3% H₂O₂ in PBS for 15 min, washed, blocked for 15 min in avidin, washed, blocked for 15 min with biotin, washed, and serum blocked (10% FBS in PBS) for 30 min at room temperature. Primary antibodies were incubated in 10% FBS in PBS overnight at 4°C, washed in PBST (PBS, 0.1% Tween 20). Secondary antibodies (1/200 Santa Cruz-2004, -2005) were incubated in 10% FBS in PBS for 45 min at 37°C. Slides were washed in PBST and developed with DAB. Slides were then dehydrated and mounted with Cytoseal. PCNA antibody: 1/50 Santa Cruz 7907 Lot0402. H3K27me3 antibody: 1/500 Active Motif 39155 Lot 01613015. CK19 antibody: 1/100 from Ben Stanger's Lab. For statistics, the two-sided student's ttest with ftest were used on the average percent staining per animal.

TUNEL staining: Trevigen TACS TdT-DAB kit (Cat #4801-30-K) after avidin/biotin blocking.

H3K27me3 gene body coverage increase statistical analysis

Monte Carlo simulation was employed to assess the significance of the gene body coverage by H3K27me3 domains (Figure 1D): the selected genes were measured for a

median P14-M2 increase in gene body domain coverage, a random set of genes of equal size was sampled, and the median difference in gene body domain coverage was measured 1,000 times, with the p-value estimated as the number of samples in which the median difference met or exceeded the observed difference divided by 1,000.

srHC

10 µg DNA from sonicated chromatin were resuspended in 50 µL TE. For large fragments: 25 µL beads (0.5 volumes) (Beckman Coulter A63881) were added to the 50µL of DNA, incubated, and beads were removed and large DNA isolated from them as described by the manufacturer. For medium fragments: 10 µL beads (0.2 volumes) were added to the supernatant from the large beads/DNA slurry, incubated, and beads were removed and medium DNA isolated from the beads. For small fragments: 35 µL beads (0.7 volumes) were added to the supernatant from the medium beads/DNA slurry, incubated, and beads were removed and small DNA isolated from the beads. Size selection efficacy was confirmed (Agilent 5067-4626) (Figure S3a).

For library preparation considerations, large DNA was sonicated after size selection with a Covaris S220 with the following settings PP-175 W, DF-10, CB-200, 4-9°C, 5 minutes, then ethanol precipitated.

Libraries were prepped (NEB E7370) per the manufacturer's recommendations. For size selection of small libraries: 55 µL beads for the first step and 25 µL beads for the second step. Libraries were sequenced with 75 bp single-end reads.

Enhancer-promoter unit and DHS processing

EPU's were downloaded from Shen et al, 2007(Shen et al., 2012). 1 bp was added to enhancers that loop to multiple genes. Enhancers were centered by intersecting with

concatenated liver DHS sites (GSM1014195 replicates 1-14)(Yue et al., 2014). The first DHS event in each enhancer was used.

Domain calling

We adapted a previously described algorithm(Becker et al., 2017) to call domains. H3K27me3 parameters: 2kb windows, 1kb slide, top 30% cutoff. srHC parameters: 10kb windows, 2kb slide, and 40% cutoff. Intermediate domains include regions not called as hetero/euchromatic and heterochromatin/euchromatin double positive regions. 50% or 75% coverage by domains were used for marked gene and promoter calls, respectively.

DNA-sequencing and processing

Reads were aligned to the NCBI v37/mm9 genome using STAR2.4.2a with the following arguments: `--outFilterMultimapNmax 20, --alignIntronMax 1`, then filtered for unique alignments to avoid PCR duplication artifacts. RPM-normalized bedgraphs of alignments were generated, then values $\log_2(\text{large}/\text{small})$ or `ChIP minus Input` subtracted calculated for every block. To avoid dividing by zero, a small addend was added to every block. Biological replicate values were merged using `bedtools unionbedg` and averaged.

Read density (or srHC-seq scores) were calculated using the Bioconductor Genomation v1.6.0 package using `ChIP minus Input` (negative values converted to 0) or $\log_2(\text{large}/\text{small})$ bigwigs. The resulting values were quantile-normalized across all P14 and M2 individual biological replicates. On box and whisker plots, whiskers indicate 5th and 95th percentiles and Wilcoxon statistical testing was used.

RNA-sequencing and processing

Reads were aligned to the mm9 genome using STAR2.4.2a with the following arguments: `--outFilterType BySJout, --outFilterMultimapNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --alignIntronMin 20, --`

alignIntronMax 1000000. HTSeq version 0.6.1p1 count was used to assign reads to genes with the NCBI v37/mm9 genome file (RefSeq genes, refFlat), then normalized and DE genes called using DESeq2 ($\alpha \leq 0.05$, $FC \geq 2$).

For browser views, alignments greater than 75 bp were filtered out to avoid showing spurious intronic signal, and converted to RPM-normalized strand-specific bigWigs. Replicates were averaged at the RPM-normalized bedGraph stage with the tool bedtools unionbedg.

Zonated genes were kindly provided by Dr. Itzkovitz (Halpern et al., 2017).

Experimental model and subject details

All animal studies were performed with the University of Pennsylvania IACUC approval. Genetics include the *Alb-Cre* transgene (Postic et al., 1999), and *Ezh1*^{-/-} (Ezhkova et al., 2011) and *Ezh2*^{ff} (*Ezh2tm1Tara*) (Su et al., 2003) alleles in a mixed C57BL/6J and C3H background.

Acknowledgements

We thank the Elaine Fuchs lab for the *Ezh1* mutant mice, Ryan McCarthy and Tony Hsieh for comments on the manuscript, and Bomyi Lim for help with confocal microscopy.

2.11. Supplementary Tables

The supplementary tables are large files. They are available online on the *Gastroenterology* journal website, on NCBI GEO, or from the corresponding author.

Supplementary _Table1: *mus musculus* RTqPCR primer sequences

Supplementary _Table2: Millions aligned reads for RNA-seq, ChIP-seq, srHC-seq

Supplementary_Table3: Differential expression calls. H3K27me3- and heterochromatin/intermediate/euchromatin-marked gene body and promoter calls.

Supplementary_Table4: “Liver-enriched” maturation genes. Genes upregulated in maturation that enrich for the UniProt Liver Tissue Expression Category and genes called as liver specific in Li et al., Nature 2017

Supplementary_Table5: Gene Ontology

Supplementary_Table6: Genes differentially expressed in P14 *Ezh1* single knockout hepatocytes as compared to P14 *Wt*.

Supplementary_Table7: mm9 gene loci, human and mouse orthologs, and H3K27me3-marking status for fibrosis gene sets

2.12. References

- Alpern, D., Langer, D., Ballester, B., Le Gras, S., Romier, C., Mengus, G., and Davidson, I. (2014). TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. *Elife* 3, e03613.
- Ardehali, M.B., Anselmo, A., Cochrane, J.C., Kundu, S., Sadreyev, R.I., and Kingston, R.E. (2017). Polycomb Repressive Complex 2 Methylates Elongin A to Regulate Transcription. *Mol. Cell* 68, 872-884.e6.
- Avior, Y., Levy, G., Zimmerman, M., Kitsberg, D., Schwartz, R., Sadeh, R., Moussaieff, A., Cohen, M., Itskovitz-Eldor, J., and Nahmias, Y. (2015). Microbial-derived lithocholic acid and vitamin K2 drive the metabolic maturation of pluripotent stem cells-derived and fetal hepatocytes. *Hepatology* 62, 265–278.
- Bae, W.K., Kang, K., Yu, J.H., Yoo, K.H., Factor, V.M., Kaji, K., Matter, M., Thorgeirsson, S., and Hennighausen, L. (2015). The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration. *FASEB J.* 29, 1653–1662.
- Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* 68, 1023-1037.e15.
- Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12, 123–135.
- Belayew, A., and Tilghman, S.M. (1982). Genetic analysis of alpha-fetoprotein synthesis in mice. *Mol. Cell. Biol.* 2, 1427–1435.

- Bhate, A., Parker, D.J., Bebee, T.W., Ahn, J., Arif, W., Rshan, E.H., Chorghade, S., Chau, A., Lee, J.-H., Anakk, S., et al. (2015). ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat Commun* 6, 8768.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353.
- Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* 412, 651–655.
- Cervantes, S., Fontcuberta-PiSunyer, M., Servitja, J.-M., Fernandez-Ruiz, R., García, A., Sanchez, L., Lee, Y.-S., Gomis, R., and Gasa, R. (2017). Late-stage differentiation of embryonic pancreatic β -cells requires Jarid2. *Sci Rep* 7, 11643.
- Chen, Y.-F., Tseng, C.-Y., Wang, H.-W., Kuo, H.-C., Yang, V.W., and Lee, O.K. (2012). Rapid generation of mature hepatocyte-like cells from human induced pluripotent stem cells by an efficient three-step protocol. *Hepatology* 55, 1193–1203.
- Choi, S.S., and Diehl, A.M. (2009). Epithelial-to-mesenchymal transitions in the liver. *Hepatology* 50, 2007–2013.
- Chopra, V.S., Hong, J.-W., and Levine, M. (2009). Regulation of Hox gene activity by transcriptional elongation in *Drosophila*. *Curr. Biol.* 19, 688–693.
- Conforto, T.L., Zhang, Y., Sherman, J., and Waxman, D.J. (2012). Impact of CUX2 on the female mouse liver transcriptome: activation of female-biased genes and repression of male-biased genes. *Mol. Cell. Biol.* 32, 4611–4627.
- Cui, J.Y., Renaud, H.J., and Klaassen, C.D. (2012). Ontogeny of novel cytochrome P450 gene isoforms during postnatal liver maturation in mice. *Drug Metab. Dispos.* 40, 1226–1237.
- Dattaroy, D., Pourhoseini, S., Das, S., Alhasson, F., Seth, R.K., Nagarkatti, M., Michelotti, G.A., Diehl, A.M., and Chatterjee, S. (2015). Micro-RNA 21 inhibition of SMAD7 enhances fibrogenesis via leptin-mediated NADPH oxidase in experimental and human nonalcoholic steatohepatitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* 308, G298-312.
- Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. *Mol. Cell* 13, 887–893.
- Dooley, S., Hamzavi, J., Ciucian, L., Godoy, P., Ilkavets, I., Ehnert, S., Ueberham, E., Gebhardt, R., Kanzler, S., Geier, A., et al. (2008). Hepatocyte-specific Smad7 expression attenuates TGF-beta-mediated fibrogenesis and protects against liver damage. *Gastroenterology* 135, 642–659.
- Duan, Y., Ma, X., Ma, X., Zou, W., Wang, C., Bahbahan, I.S., Ahuja, T.P., Tolstikov, V., and Zern, M.A. (2010). Differentiation and characterization of metabolically functioning hepatocytes from human embryonic stem cells. *Stem Cells* 28, 674–686.
- Duncan, A.W. (2013). Aneuploidy, polyploidy and ploidy reversal in the liver. *Semin. Cell Dev. Biol.* 24, 347–356.
- Duncan, A.W., Dorrell, C., and Grompe, M. (2009). Stem cells and liver regeneration. *Gastroenterology* 137, 466–481.
- Ezhkova, E., Lien, W.-H., Stokes, N., Pasolli, H.A., Silva, J.M., and Fuchs, E. (2011). EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev.* 25, 485–498.
- Fouts, J.R., and Adamson, R.H. (1959). Drug metabolism in the newborn rabbit. *Science* 129, 897–898.

- Gopal, E., Miyauchi, S., Martin, P.M., Ananth, S., Srinivas, S.R., Smith, S.B., Prasad, P.D., and Ganapathy, V. (2007). Expression and functional features of NaCT, a sodium-coupled citrate transporter, in human and rat livers and cell lines. *Am. J. Physiol. Gastrointest. Liver Physiol.* 292, G402-408.
- Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, 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.
- Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491.
- Horisawa, K., and Suzuki, A. (2015). Cell-Based Regenerative Therapy for Liver Disease. In *Innovative Medicine: Basic Research and Development*, K. Nakao, N. Minato, and S. Uemoto, eds. (Tokyo: Springer), p.
- Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 475, 386–389.
- Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell* 14, 370–384.
- Ieda, M. (2013). Direct reprogramming into desired cell types by defined factors. *Keio J Med* 62, 74–82.
- Ippolito, D.L., AbdulHameed, M.D.M., Tawa, G.J., Baer, C.E., Permenter, M.G., McDyre, B.C., Dennis, W.E., Boyle, M.H., Hobbs, C.A., Streicker, M.A., et al. (2016). Gene Expression Patterns Associated With Histopathology in Toxic Liver Fibrosis. *Toxicol. Sci.* 149, 67–88.
- Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N.K., Pinello, L., Yuan, G.-C., Orkin, S.H., and Shivdasani, R.A. (2016). Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* 165, 1389–1400.
- Johannesson, B., Sui, L., Freytes, D.O., Creusot, R.J., and Egli, D. (2015). Toward beta cell replacement for diabetes. *EMBO J.* 34, 841–855.
- Jung, Y.K., and Yim, H.J. (2017). Reversal of liver cirrhosis: current evidence and expectations. *Korean J. Intern. Med.* 32, 213–228.
- Koike, H., Ouchi, R., Ueno, Y., Nakata, S., Obana, Y., Sekine, K., Zheng, Y.-W., Takebe, T., Isono, K., Koseki, H., et al. (2014). Polycomb group protein Ezh2 regulates hepatic progenitor cell proliferation and differentiation in murine embryonic liver. *PLoS ONE* 9, e104776.
- van Koppen, A., Verschuren, L., van den Hoek, A.M., Verheij, J., Morrison, M.C., Li, K., Nagabukuro, H., Costessi, A., Caspers, M.P.M., van den Broek, T.J., et al. (2018). Uncovering a Predictive Molecular Signature for the Onset of NASH-Related Fibrosis in a Translational NASH Mouse Model. *Cell Mol Gastroenterol Hepatol* 5, 83-98.e10.
- Koppens, M.A.J., Bounova, G., Gargiulo, G., Tanger, E., Janssen, H., Cornelissen-Steijger, P., Blom, M., Song, J.-Y., Wessels, L.F.A., and van Lohuizen, M. (2016). Deletion of Polycomb Repressive Complex 2 From Mouse Intestine Causes Loss of Stem Cells. *Gastroenterology* 151, 684-697.e12.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4, e1000242.

- Liu, X., Wei, W., Li, X., Shen, P., Ju, D., Wang, Z., Zhang, R., Yang, F., Chen, C., Cao, K., et al. (2017). BMI1 and MEL18 Promote Colitis-Associated Cancer in Mice via REG3B and STAT3. *Gastroenterology* 153, 1607–1620.
- López-Arribillaga, E., Rodilla, V., Pellegrinet, L., Guiu, J., Iglesias, M., Roman, A.C., Gutarra, S., González, S., Muñoz-Cánoves, P., Fernández-Salguero, P., et al. (2015). Bmi1 regulates murine intestinal stem cell proliferation and self-renewal downstream of Notch. *Development* 142, 41–50.
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* 32, 503–518.
- Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell* 159, 884–895.
- Morelli, L. (2008). Postnatal development of intestinal microflora as influenced by infant nutrition. *J. Nutr.* 138, 1791S-1795S.
- Morford, L.A., Davis, C., Jin, L., Dobierzewska, A., Peterson, M.L., and Spear, B.T. (2007). The oncofetal gene glypican 3 is regulated in the postnatal liver by zinc fingers and homeoboxes 2 and in the regenerating liver by alpha-fetoprotein regulator 2. *Hepatology* 46, 1541–1547.
- Murphy, N., Gaynor, K.U., Rowan, S.C., Walsh, S.M., Fabre, A., Boylan, J., Keane, M.P., and McLoughlin, P. (2016). Altered Expression of Bone Morphogenetic Protein Accessory Proteins in Murine and Human Pulmonary Fibrosis. *Am. J. Pathol.* 186, 600–615.
- Nitta, T., Kim, J.-S., Mohuczy, D., and Behrns, K.E. (2008). Murine cirrhosis induces hepatocyte epithelial mesenchymal transition and alterations in survival signaling pathways. *Hepatology* 48, 909–919.
- Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483, 598–602.
- Patel, V., Mathison, M., Singh, V.P., Yang, J., and Rosengart, T.K. (2016). Direct Cardiac Cellular Reprogramming for Cardiac Regeneration. *Curr Treat Options Cardiovasc Med* 18, 58.
- Perincheri, S., Dingle, R.W.C., Peterson, M.L., and Spear, B.T. (2005). Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the Zfx2 gene. *Proc. Natl. Acad. Sci. U.S.A.* 102, 396–401.
- Postic, C., Shiota, M., Niswender, K.D., Jetton, T.L., Chen, Y., Moates, J.M., Shelton, K.D., Lindner, J., Cherrington, A.D., and Magnuson, M.A. (1999). Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* 274, 305–315.
- Ramnath, D., Irvine, K.M., Lukowski, S.W., Horsfall, L.U., Loh, Z., Clouston, A.D., Patel, P.J., Fagan, K.J., Iyer, A., Lampe, G., et al. (2018). Hepatic expression profiling identifies steatosis-independent and steatosis-driven advanced fibrosis genes. *JCI Insight* 3.
- Roelandt, P., Vanhove, J., and Verfaillie, C. (2013). Directed differentiation of pluripotent stem cells to functional hepatocytes. *Methods Mol. Biol.* 997, 141–147.
- Ruscica, M., Ferri, N., Macchi, C., Meroni, M., Lanti, C., Ricci, C., Maggioni, M., Fracanzani, A.L., Badiali, S., Fargion, S., et al. (2016). Liver fat accumulation is associated with circulating PCSK9. *Ann. Med.* 48, 384–391.

- Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* 475, 390–393.
- Shearer, M.J., Rahim, S., Barkhan, P., and Stimmler, L. (1982). Plasma vitamin K1 in mothers and their newborn babies. *Lancet* 2, 460–463.
- Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120.
- Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* 10, 697–708.
- Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 51, 297–305.
- Song, Z., Cai, J., Liu, Y., Zhao, D., Yong, J., Duo, S., Song, X., Guo, Y., Zhao, Y., Qin, H., et al. (2009). Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell Res.* 19, 1233–1242.
- Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151, 994–1004.
- Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 γ in reprogramming to pluripotency. *Nat. Cell Biol.* 15, 872–882.
- Su, I.-H., Basavaraj, A., Krutchinsky, A.N., Hobert, O., Ullrich, A., Chait, B.T., and Tarakhovskiy, A. (2003). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* 4, 124–131.
- Taura, K., Miura, K., Iwaisako, K., Osterreicher, C.H., Kodama, Y., Penz-Osterreicher, M., and Brenner, D.A. (2010). Hepatocytes do not undergo epithelial-mesenchymal transition in liver fibrosis in mice. *Hepatology* 51, 1027–1036.
- 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.
- Thatcher, J.E., Zelter, A., and Isoherranen, N. (2010). The relative importance of CYP26A1 in hepatic clearance of all-trans retinoic acid. *Biochem. Pharmacol.* 80, 903–912.
- Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad. Sci. U.S.A.* 113, E744-753.
- Treluyer, J.M., Cheron, G., Sonnier, M., and Cresteil, T. (1996). Cytochrome P-450 expression in sudden infant death syndrome. *Biochem. Pharmacol.* 52, 497–504.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13.
- Wong, M.C.S., and Huang, J. (2018). The growing burden of liver cirrhosis: implications for preventive measures. *Hepatol Int* 12, 201–203.

- Xu, C.-R., Cole, P.A., Meyers, D.J., Kormish, J., Dent, S., and Zaret, K.S. (2011). Chromatin “patterning” and histone modifiers in a fate choice for liver and pancreas. *Science* 332, 963–966.
- Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364.
- Zaret, K.S., and Grompe, M. (2008). Generation and regeneration of cells of the liver and pancreas. *Science* 322, 1490–1494.

Chapter 3: H39me3-heterochromatin loss at protein-coding genes enables developmental lineage specification

3.1. Preface

The manuscript presented in this chapter was originally published on January 3, 2019 (Nicetto et al., 2019). It has been reformatted here in accordance with University of Pennsylvania dissertation formatting guidelines. All genomic data are accessible at the Gene Expression Omnibus (GEO) database repository GSE114198. The references for this chapter are at the end of this chapter.

One Sentence Summary: H3K9me3-marked heterochromatin is transiently deployed during early development and removed later to allow expression of cell type-specific genes.

Authors:

Dario Nicetto¹⁻³, Greg Donahue¹⁻³, Tanya Jain¹⁻³, Tao Peng^{4,5}, Simone Sidoli^{2,6}, Lihong Sheng^{2,3}, Thomas Montavon⁷, Justin S. Becker¹⁻³, Jessica M. Grindheim¹⁻³, Kimberly Blahnik¹⁻³, Benjamin A. Garcia^{2,6}, Kai Tan^{3-5,8}, Roberto Bonasio^{2,3}, Thomas Jenuwein⁷, Kenneth S. Zaret^{1-3,*}

Affiliations:

¹Institute for Regenerative Medicine

²Penn Epigenetics Institute

³Dept. Cell and Developmental Biology

⁴Dept. of Biomedical and Health Informatics

⁵Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, USA

⁶Dept. of Biochemistry and Molecular Biophysics

⁷Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany

⁸Dept. of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

*Corresponding author: zaret@upenn.edu

3.2. Respective Contributions

D.N. and K.S.Z. designed this study and wrote the manuscript; D.N., T.J., L.S., T.M., and J.M.G. conducted the experiments; G.D., T.P., and J.S.B. performed the bioinformatic analysis; S.S. performed the K-mean clustering analysis of RNA-seq data; K.B. helped with small cell number ChIP protocol; and T.M. and T.J. provided Suv39h dn samples. B.A.G., K.T., and R.B. helped with proteomic and single-cell RNA seq analysis.

3.3. Abstract

Gene silencing by chromatin compaction is integral to establishing and maintaining cell fates. Trimethylated histone 3 lysine 9 (H3K9me3)-marked heterochromatin is reduced in embryonic stem cells compared to differentiated cells. However, the establishment and dynamics of closed regions of chromatin at protein-coding genes, in embryologic development, remain elusive. We developed an antibody-independent method to isolate

and map compacted heterochromatin from low–cell number samples. We discovered high levels of compacted heterochromatin, H3K9me3-decorated, at protein-coding genes in early, uncommitted cells at the germ-layer stage, undergoing profound rearrangements and reduction upon differentiation, concomitant with cell type–specific gene expression. Perturbation of the three H3K9me3-related methyltransferases revealed a pivotal role for H3K9me3 heterochromatin during lineage commitment at the onset of organogenesis and for lineage fidelity maintenance.

3.4. Main Text

The phylotypic period of embryologic development occurs at the onset of organogenesis, in which morphological development is most conserved between different species (1–3). The “hourglass” model suggests that cell fate decisions are restricted during the phylotypic period by evolutionarily conserved transcription factor and signaling activities (1–3). Limited assay sensitivity and small numbers cells have made it difficult to investigate chromatin dynamics during the phylotypic period, when cell differentiation initiates extensively in embryos. Current thinking from the embryonic stem (ES) cell model (4) suggests that compacted heterochromatic domains expand as cells mature, helping to establish cell identity (5–11). However, these studies did not examine the dynamic events occurring during natural lineage commitment.

Regions of H3K9me3-marked heterochromatin can have a physically condensed structure (12–14) that serves to repress repeat-rich regions of the genome (7, 15–17), including centromeric and telomeric regions (18, 19), and silence protein-coding genes at facultative heterochromatin (20, 21). The early lethal *in vivo* developmental phenotypes associated with the depletion of H3K9me3-related histone methyltransferases (HMTases)

(15, 22, 23) support the idea that H3K9me3 controls genome stability and differentiation. Recently, H3K9me3 dynamics at repetitive elements and promoters have been characterized at pre-gastrula stages (24). The global heterochromatin reorganization at germ layer stages and during lineage commitment in vivo has not been addressed, and prior studies did not distinguish H3K9me3-decorated regions that are euchromatic from those that are heterochromatic (25). H3K9me3-enriched domains also impede cell reprogramming and somatic cell nuclear transfer (17, 25–27), underscoring the significance of understanding the natural dynamics by which heterochromatic domains restrict cell fates during normal development.

We globally assessed the dynamics of compacted, sonication-resistant heterochromatin (srHC) (25) and H3K9me3 deposition at critical developmental time points in the murine endoderm germ layer and in cells along the descendent hepatic and pancreatic lineages (Fig. 1A, and fig. S1A and B, fig. S2A to H, fig. S3A to F). Since the embryonic starting material has low cell numbers, we developed a sonication-resistant heterochromatin sequencing (srHC-seq) method that is sucrose gradient-independent (25) to detect regions of srHC (fig. S4 A to E). We performed srHC-seq in definitive endodermal cells, hepatocytes, and mature beta cells, and found similar fractions of the genome in srHC in the three cell types (fig. S4 F and G). In all stages, gene expression was anti-correlated with sonication resistance (fig. S4H). Analysis of Hi-C-identified closed compartments revealed 40% overlap with both adult hepatocytes and mature beta cells srHC (fig. S4I), whereas no significant correlation with open compartments was detected. We observed extensive dynamics of srHC upon definitive endoderm differentiation (Fig. 1B, Table S7), including 5,979 and 4,879 genes that lose compaction, whereas 1,630 and 5,632 genes gain srHC, during hepatocyte and mature beta cell development,

respectively. GO analysis revealed that srHC is removed at adult function genes (Table S7).

We mapped H3K9me3 in cells sorted from embryos at different developmental stages (fig. S5A to D). Unsupervised hierarchical clustering revealed a high correlation between individual replicates, with definitive endoderm cells clustering separately from the hepatic and pancreatic lineages (fig. S5E). To compare H3K9me3 landscapes across the three germ layers, we included mesoderm progenitors and ectoderm-derived, already specified midbrain neuroepithelial cells isolated at e8.25 (fig. S6 A to H) and compared their H3K9me3 profiles to definitive endodermal cells, as well as to P0 heart and adult nucleus accumbens (Figure 1A). Concordant with the heterochromatin analysis, H3K9me3 marked more gene bodies, promoters, and termination transcription sites (TTS) in endoderm and mesoderm germ layer than in pregastrula stages or differentiating cells (Fig. 1C, fig. S7A-E, Tables S8, S9). A step-wise developmental transition analysis of H3K9me3 revealed a substantial loss of H3K9me3 when definitive endodermal cells differentiate into hepatic and pancreatic progenitors (Fig. 1D, Tables S10, S11). A similar process is detected in the mesoderm lineage, but not upon differentiation of midbrain neuroepithelium, which is already past the ectoderm stage, into neurons (fig. S7F, Tables S10, S11).

H3K9me3 and H3K27me3 reside both in srHC and open chromatin, where they decorate regions independently or in combination (25) (fig. S8A to C). However, unlike H3K9me3, heterochromatin marked by H3K27me3 was similarly distributed over genes and intergenic regions in definitive endoderm, hepatocytes, and mature beta cells (fig. S9A to E).

We assessed the acquisition of stage-specific transcriptional signatures along the hepatic and pancreatic lineages (fig. S10A to C, Table S12). Combined analysis of srHC,

H3K9me3, and transcriptional profiles revealed that gene bodies, TSSs, or TTSs marked by H3K9me3 are more repressed when present in srHC than in open chromatin (fig. S11A and B). K-means cluster analysis of six developmental stages (fig. S12, Table S13) identified 15 patterns of gene expression. Notably, cell type-specific genes that acquire expression in terminally differentiated cells showed a net loss of srHC and H3K9me3 along both the hepatic and pancreatic lineages (Fig. 2A, C to E, fig. S13A, fig. S14A to D, Table S14), whereas constitutively expressed or repressed genes were depleted or decorated by the mark, respectively (Fig. 2B and fig. S13B and C). H3K27me3 dynamics at both hepatic and pancreatic-specific genes were also detected, showing loss in development (fig. S14E). Of the 1,008 and 1,249 genes in adult hepatocytes and mature beta cells that fail to be expressed at a higher level in differentiated versus uncommitted cells, but lose H3K9me3, 71% and 74% respectively, showed increased H3K27me3 levels compared to definitive endoderm (fig S15A), indicating a compensatory mechanism for maintaining heterochromatin at a subset of genes that remain developmentally silent. Overall, the results show that H3K9me3-marked heterochromatin is transiently deployed in germ layer cells to repress genes associated with mature cell function, and is removed at many sites during differentiation to allow tissue-specific gene expression.

H3K9me3 is established by three main HMTases, *Setdb1*, *Suv39h1*, and *Suv39h2* (15, 22, 23). *Setdb1* single and *Suv39h1/h2* double germ-line knock-outs are associated with early lethal phenotypes (15, 17, 22). We used *FoxA3-Cre* to generate endoderm-specific (28, 29), conditional knockout (KO) mice for *Setdb1* (30) (fig. S16A) and analyzed e11.5 livers. H3K9me3 was modestly reduced in mutant embryos (fig. S16B), which showed bleeding in different body regions, but no gross morphological differences in the liver structure and cell composition (fig. S16C). Single cell RNA-seq on wild-type e11.5 hepatoblasts revealed three clusters of cell types (cluster 1-3, Table S15), whose

differentially expressed genes were associated with developmental processes, hepatic metabolism, and hematopoiesis, respectively (Fig. 3A and fig. S17A to E). Expressed genes in *Setdb1* mutant cells more than doubled in cluster 1 compared to wt cells, but were reduced in cluster 2 (Fig. 3A). *Setdb1* mutant *albumin* positive (*Alb*⁺) cells from cluster 2 fail to induce hepatic markers and separate into a distinct sub-cluster from wt *Alb*⁺ cells (Fig 3B and Table S15). Adult, conditional *Setdb1* mutant livers show occasional hypertrophic hepatocytes (fig. S16D and E) that maintain nuclear *Setdb1* and H3K9me3 levels as well as expression of major urinary proteins (MUPs) (fig. S16D). However, the bulk of *Setdb1*-negative cells show lower levels of H3K9me3 and no expression of MUPs, in stark contrast to wt livers that uniformly express pericentral MUPs (fig. S16D). Thus, *Setdb1* modulates hepatocyte differentiation.

The persistence of low-level H3K9me3 in the conditional *Setdb1* mutants and *Suv39h1* and *Suv39h2* double mutants (fig. S16B and S18A), the appearance of escaper cells (fig. S16D, arrows), and the expression of H3K9me3-related HMTases being higher in definitive endoderm, compared to more specified cells (fig. S18B), prompted us to generate an endoderm-specific conditional triple knock-out mutant (TKO) murine strain of all 3 H3K9me3-related HMTases (fig. S19A to F). Protein analysis showed a clear reduction in *Setdb1*, *Suv39h1* (fig. S20A), and *Suv39h2* (fig. S20B), leading to a dramatic decrease in H3K9me3, but not in H3K27me3 and H3K9me2 (fig. S20A and C). Single cell RNA-seq on E11.5 *Liv2*⁺ cells revealed that TKO hepatoblasts clustered into a separate group compared to wt and *Setdb1* mutant cells (Fig. 4A), with an overlap of only 43 genes, of non-liver types, upregulated in common between TKO and *Setdb1* mutant cells (Table S15). Indeed, despite expressing *albumin* (fig. S20D), TKO cells never gained a clear hepatic transcriptional profile (fig. S20E and Table S15). One-month old triple mutant animals (n=5) appeared smaller in size compared to control littermates (Fig. 4B), showing

up to a 3-fold reduction in body weight (fig. S21A). TKO livers display inflammatory phenotypes, characterized by a ductular reaction (Fig. 4C). Genomic analysis (fig. S21B and C) confirmed a dramatic loss in srHC and H3K9me3 (Fig. 4D), which was validated by a global loss of condensed chromatin as seen by electron microscopy (Fig. 4E). RNA-seq data on 1-month old livers (fig. S21D and E) revealed a marked derepression of non-hepatic genes in TKO livers and a failure to induce mature hepatocyte genes such as MUPs (fig. S22A and B, Table S16). The latter phenotype, seen also in adult *Setdb1* KO livers (fig. S16D), indicates secondary effects upon depletion of H3K9me3-related HMTases. Importantly, markers associated with chromosomal instability were mostly unaffected (fig. S22C to E). Thus, failure to establish H3K9me3-marked heterochromatin during early development leads to a failure of hepatocyte maturation, even 1 month after birth, and results in expression of inappropriate lineage genes.

Heterochromatin has been defined by biophysical properties more than by repressive histone modifications (25). We employed an approach whereby srHC is isolated and characterized independently and in correlation to H3K9me3. We found higher levels of heterochromatin at gene bodies in early, uncommitted endodermal and mesodermal cells, and observed a developmental loss of H3K9me3 and srHC during cell differentiation in vivo (fig. S23A). Genetics of H3K9me3-related HMTase mutant mice highlighted the importance of proper heterochromatin establishment to promote cell differentiation. These findings underscore how epigenetic regulation of chromatin structure controls cell identity in embryogenesis. We propose a role for H3K9me3-marked heterochromatin as an epigenetic contributor to the hourglass model (1–3), working in concert with homeobox proteins (1) and signaling (2) influences, to constrain gene activity during the phylotypic period of embryonic development and guarantee establishment of cell identity.

3.5. Figures

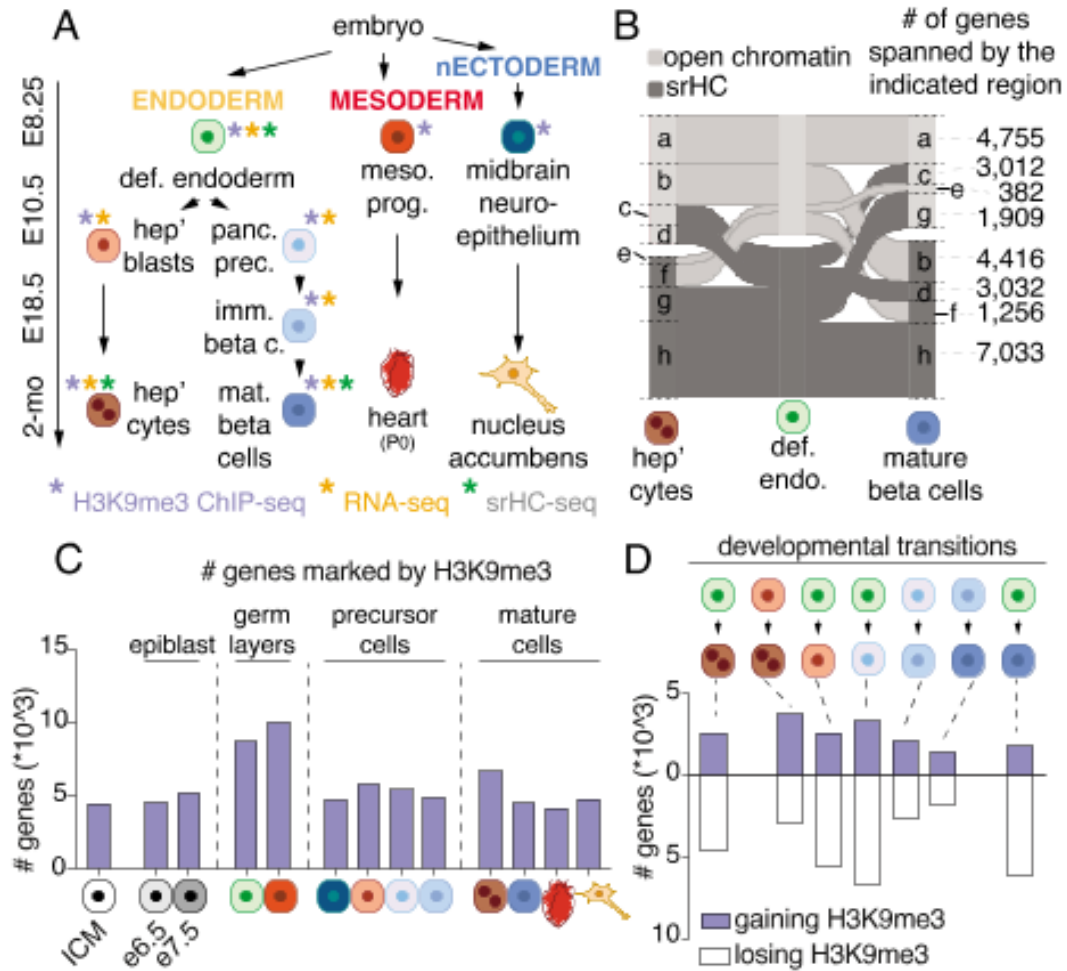


Figure 1: Chromatin compaction and H3K9me3 landscape upon germ-layers differentiation.

(A) Schematic of the cell types and embryonic developmental stages considered in this study. Purple, orange and green asterisks beside cell the represented cell types indicate samples processed for H3K9me3 ChIP-seq, RNA-seq and srHC-seq, respectively (B) Alluvial plot showing dynamics in definitive endoderm srHC (dark grey) and open chromatin (light grey) patches upon differentiation into adult hepatocytes and insulin producing cells; a-h indicate distinct categories/alluvia, characterized by a specific dynamic in srHC and open chromatin; the number of genes in each alluvium is indicated. (C) Number of genes marked by H3K9me3 in each indicated stage, along the hepatic and pancreatic lineages. ICM, e6.5, and e7.5 data from (25). (D) Number of genes gaining (purple) or losing (white) H3K9me3 upon step-wise transition in successive developmental stages. Each transition is indicated above the corresponding gene number bar.

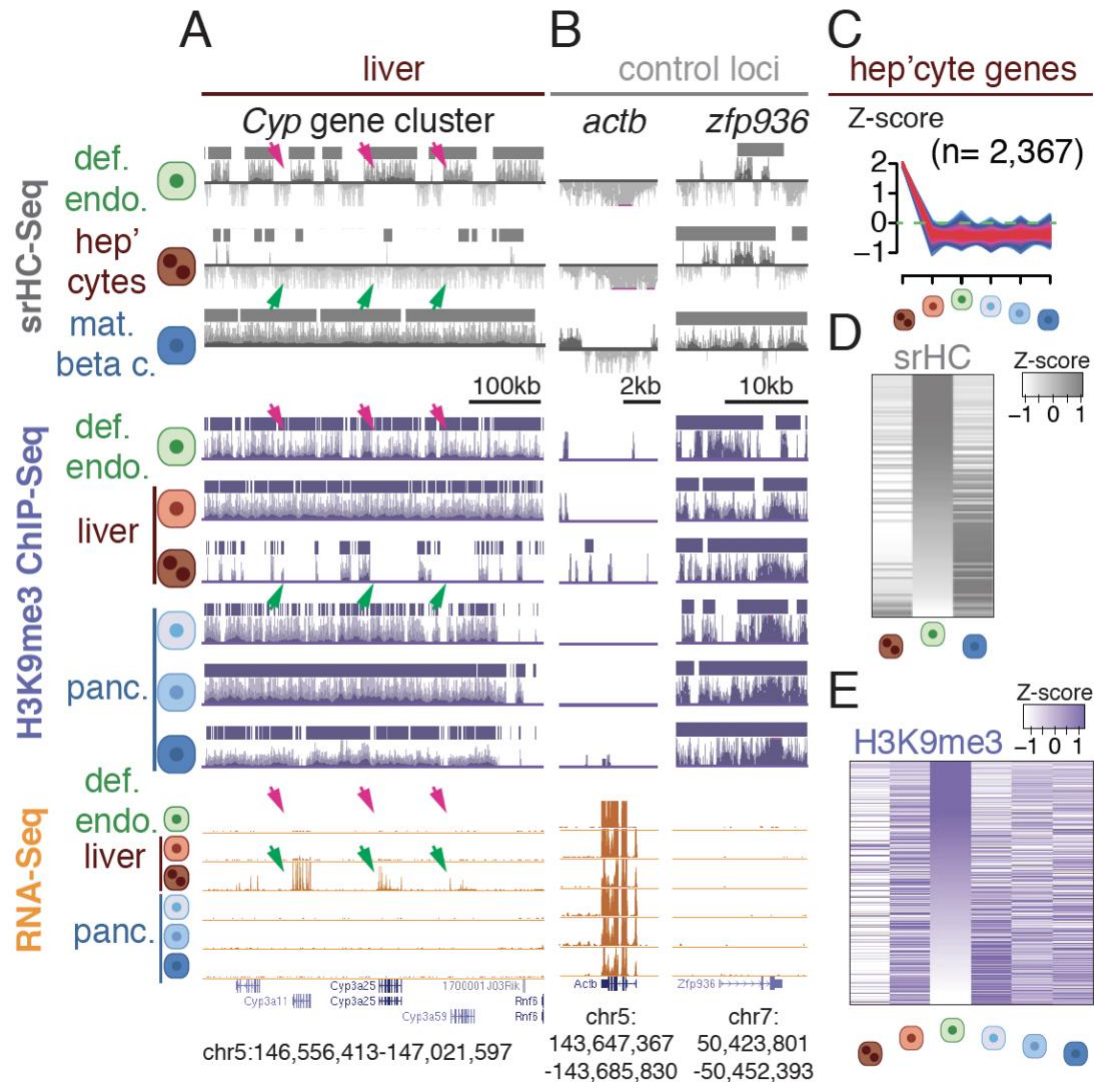
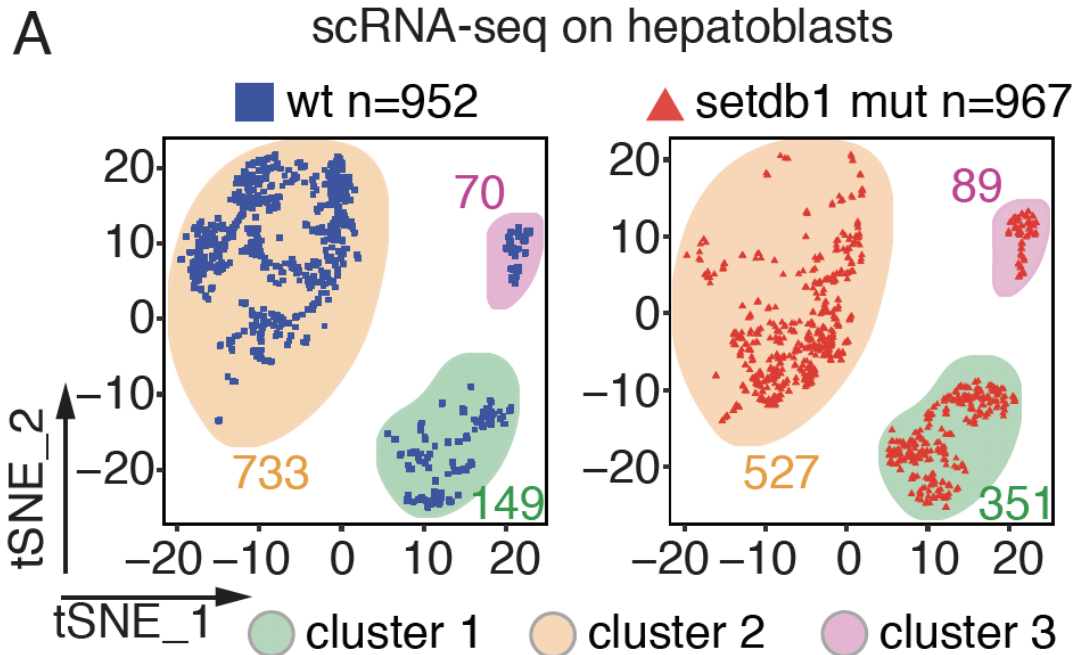


Figure 2: Loss of srHC and H3K9me3 correlates with gene expression of hepatic-specific markers upon differentiation.

(A) Representative UCSC genome browser tracks of srHC-Seq (gray), input-divided H3K9me3 (purple) and RNA-Seq profiles (orange) upon definitive endoderm differentiation into adult hepatocytes. SrHC and H3K9me3 patches are shown as grey and purple bars above each profile, respectively. Cytochrome P450 (*Cyp*) genes on chromosome 5 (A) are shown. The constitutive H3K9me3-undecorated and active *Actin b* and the permanently H3K9me3-enriched and silenced *zinc finger protein (Zfp) 936* (B) have been included as examples of genes whose expression inversely correlate to H3K9me3 presence. Magenta arrows indicate presence of srHC and H3K9me3 and absence of expression. Green arrows indicate absence of srHC, H3K9me3, and gene expression. (C) Z-score cluster representations for genes expressed in adult hepatocytes. (D and E) heatmaps showing levels of srHC (D) and H3K9me3 (E) in the indicated stages. For both srHC and H3K9me3 heatmaps, definitive endodermal cells values have been ordered in a descendent manner.



B scRNA-seq analysis on *Alb*⁺ cells from cluster 2

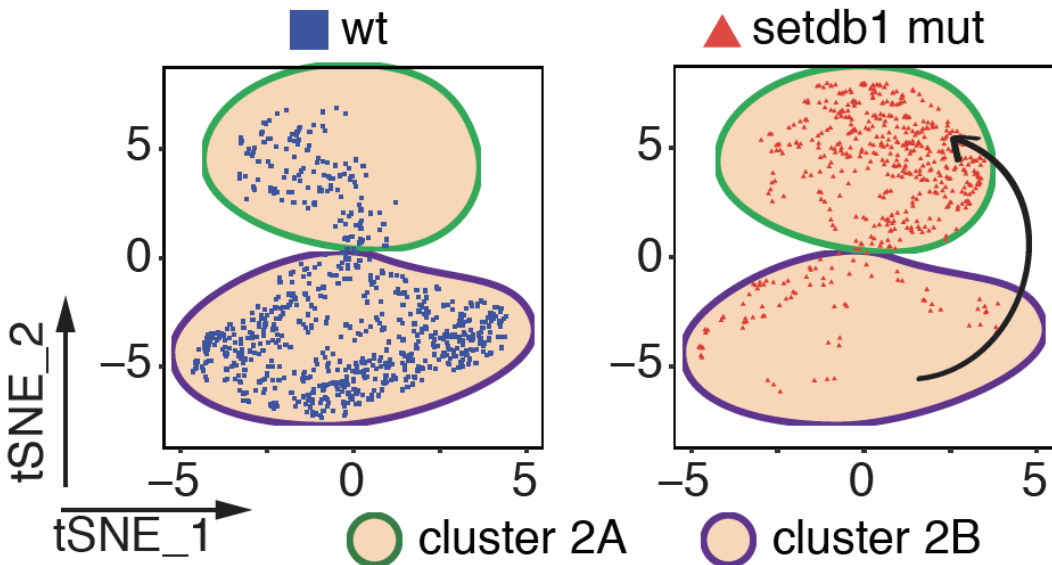


Figure 3: Setdb1 mutant hepatoblasts upregulate lineage non-specific genes.

(A) tSNE plots of single cell RNA-Seq data showing wt (cells from n= 7 embryos; left, blue squares) and Setdb1 mutant (cells from n=3 embryos; right, red triangles) cells in the three identified clusters. (B) tSNE plots of single cell RNA-Seq data showing wt (top, blue squared) and Setdb1 mutant (bottom, red triangles) Albumin positive cells from Cluster 2. The black arrow indicates transition of Setdb1 mutant cells (red triangles) to a different cluster than the wt cells (blue squares).

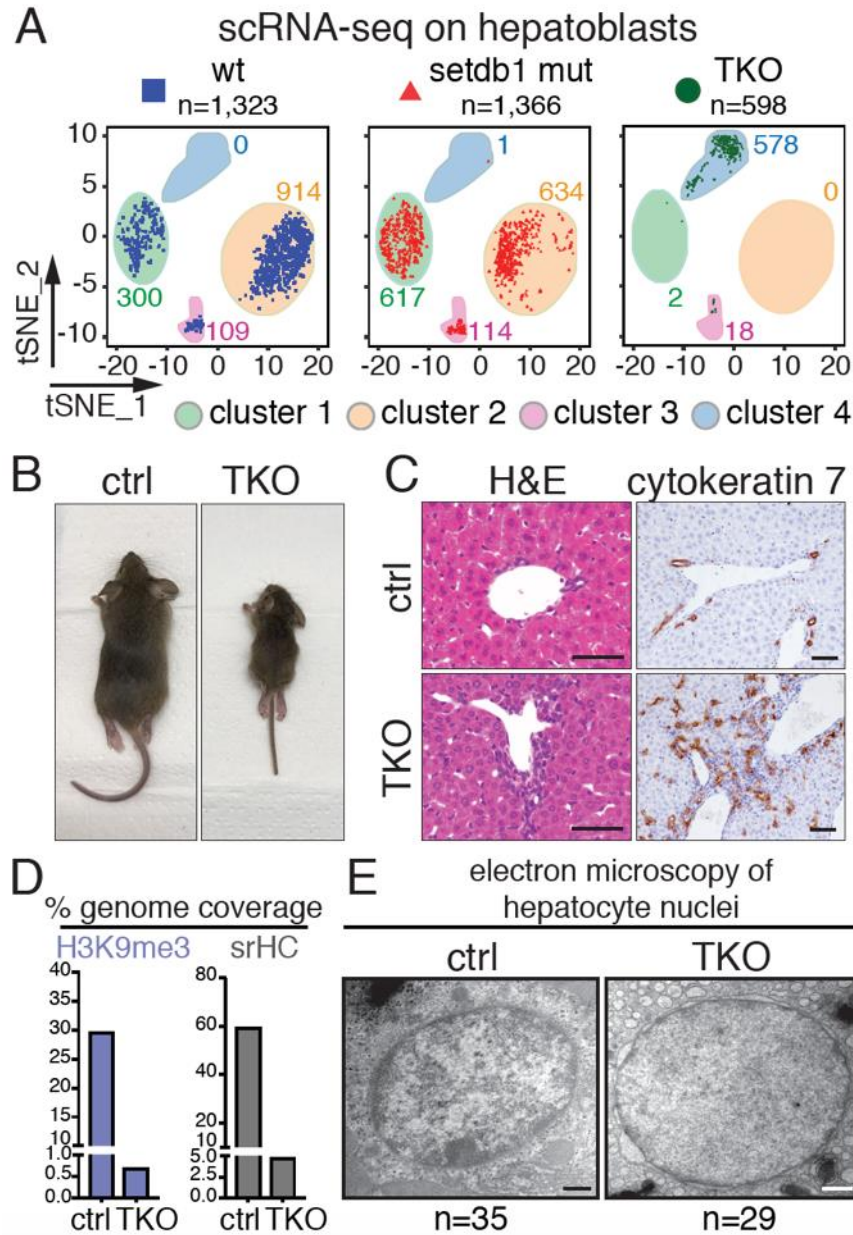


Figure 4: TKO mutant cells lose hepatic identity and show developmental phenotypes associated with decreased H3K9me3 and srHC levels.

(A) tSNE plots of e11.5 single cell RNA-Seq data showing wt, *Setdb1* mutant (same as Fig.3), and TKO (cells from $n=7$ embryos; right, dark green circles) cells in the four identified clusters. (B) Representative morphological phenotype of 1-month old ctrl ($n=3$) and *FoxA3*-cre; *Setdb1* fl/fl; *Suv39h1* fl/fl; *Suv39h2* KO/KO triple knockout (TKO) mutants ($n=5$). (C) H&E staining and cytokeratin 7 IHC in 1-month old ctrl and TKO livers. Scale bar: 50 μ m. (D) Percentage of genome covered by H3K9me3 and sr-HC domains in ctrl and TKO livers. (E) Representative electron microscopy images for ctrl and TKO one-month old hepatocytes. Scale bar: 600 nm. The number of cells recorded in the two groups is indicated at the bottom.

3.6. Supplementary Figures

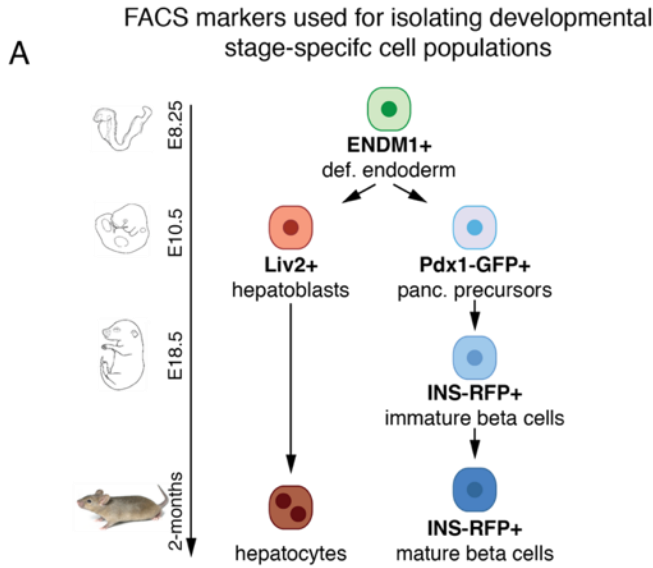
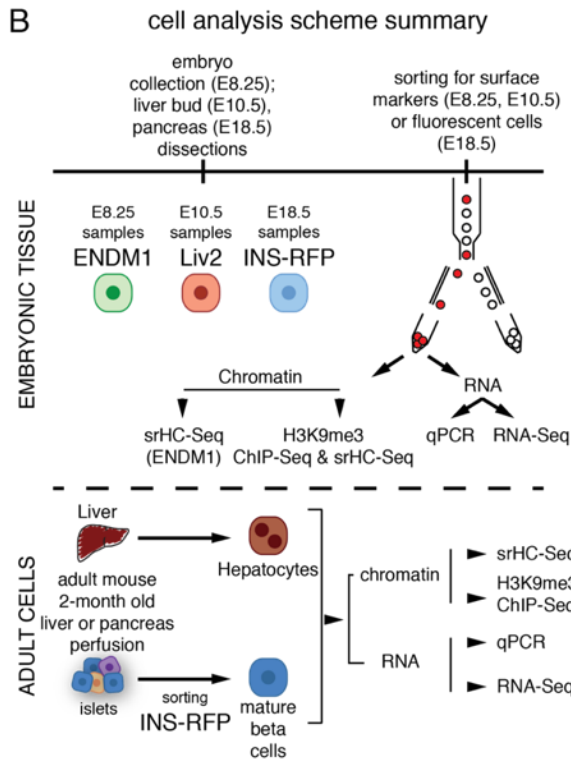


Fig. S1: Strategy for isolation of endoderm-derived cell populations.

(A) Schematic of the cell types, embryonic developmental stages and cellular surface markers or murine strains considered in this study. **(B)** Strategy for cell population isolation and downstream analysis.



cell sorting scheme and examples

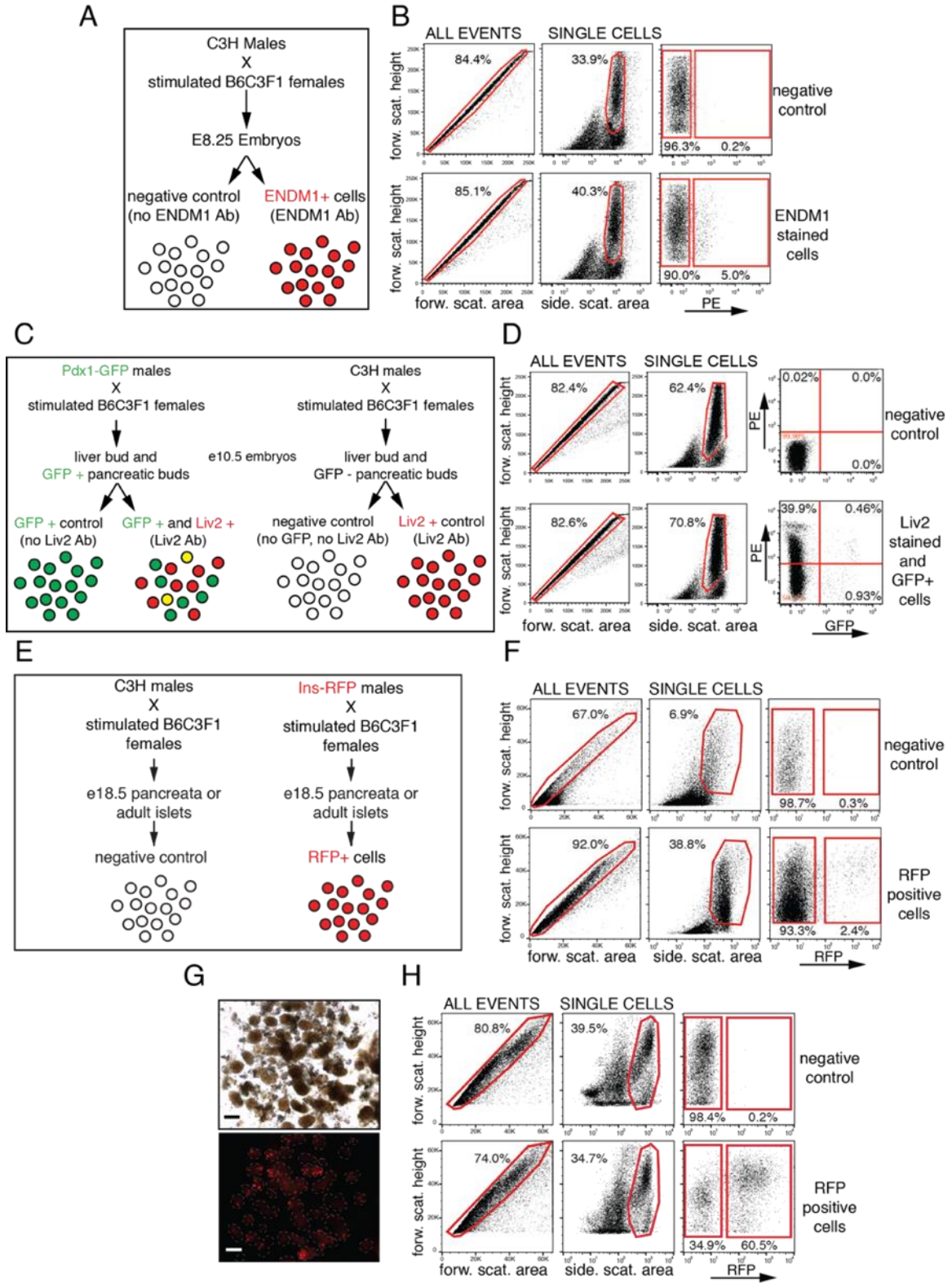


Fig. S2: Sorting strategy of embryonic cells

(A) Schematic representing the sorting strategy of definitive endodermal ENDM1+ cells from e8.25 embryos. **(B)** Representative sorting pattern of ENDM1+ cells. **(C)** Schematic representing the sorting strategy for Liv2+ hepatoblasts from e10.5 embryos. For this experiment Pdx1-GFP transgenic mouse strain have been used to separate pancreatic and hepatic cells. Single GFP+ and Liv2+ populations have been used to properly set sorting gates **(D)** Representative sorting pattern of Liv2+ cells. Note that Pdx1-GFP+/Liv2+ double positive population can be detected. For the purpose of this paper only Liv2+ and Pdx1-GFP+ single positive cells only have been considered. **(E)** Schematic representing the sorting strategy of e18.5 insulin-RFP+ cells. **(F)** Representative sorting pattern of e18.5 insulin-RFP+ cells. **(G)** Bright field (top) and fluorescent image (bottom) of isolated 2-month old islets from Insulin-RFP+ adult mice. **(H)** Representative sorting pattern of 2-month insulin-RFP+ cells.

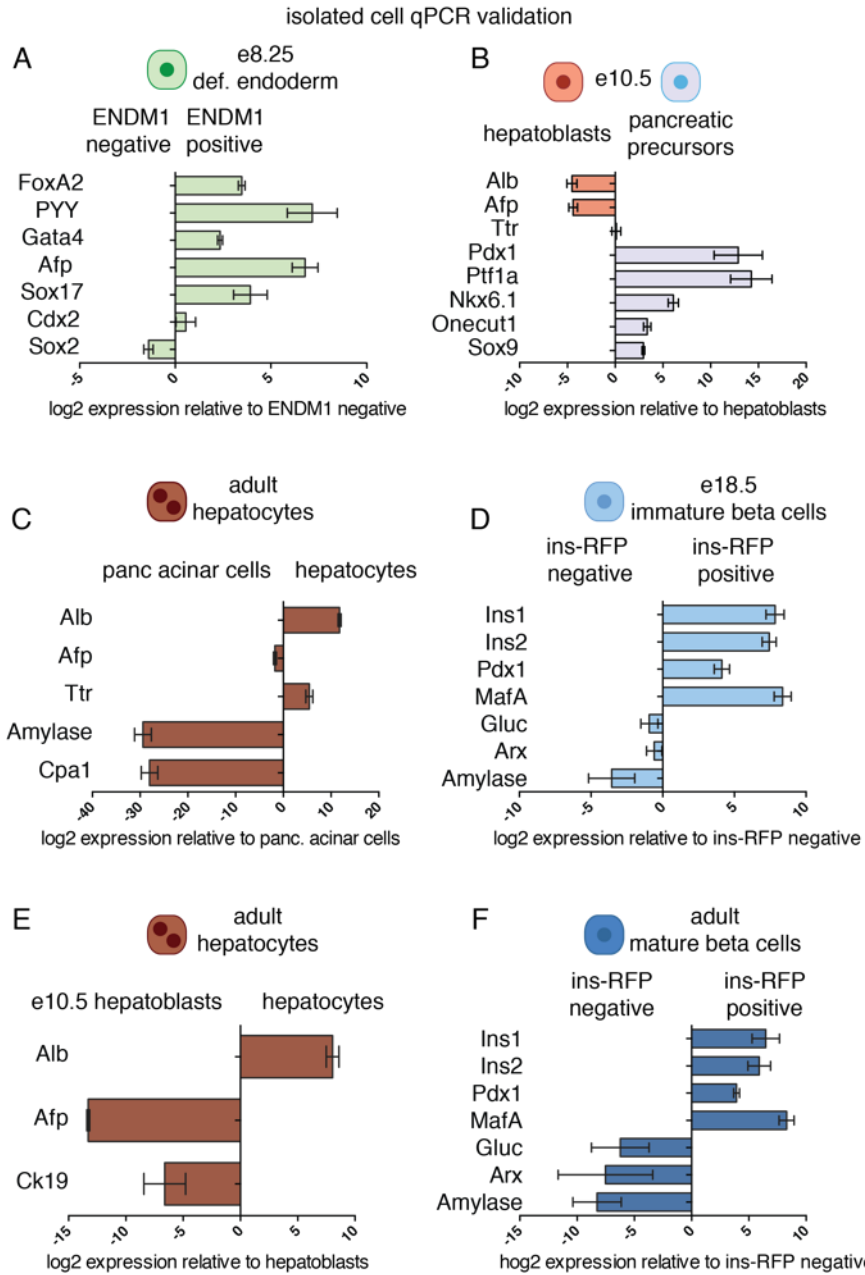


Fig. S3: Validation of sorted cell identities.

Quantitative PCR (qPCR) of definitive (A) endodermal ENDM1+ cells, (B) Liv2+ hepatoblasts and Pdx1-GFP+ pancreatic precursor cells, and (C and E) hepatocytes, (D) e18.5 immature and (F) 2-month old mature insulin-RFP+ cells, confirming the identity of the isolated cells. Ct values have been normalized to GAPDH and indicated populations of isolated cells have been used as reference. Expression is shown as Log₂ scale relative to reference population. At least three independent biological replicates have been considered for each quantification. Taqman probes used for this experiment are indicated in Table S1.

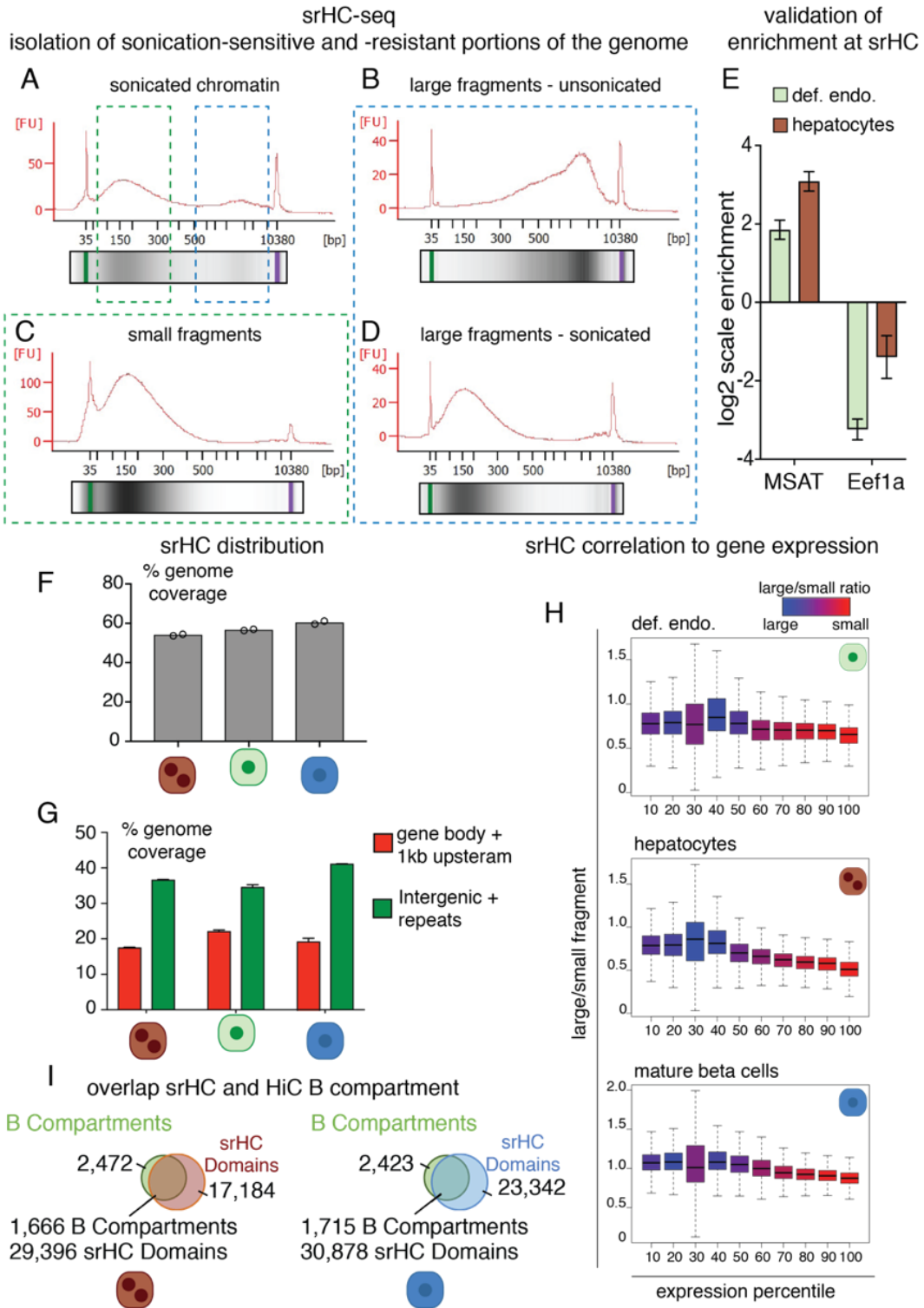


Fig. S4: srHC-seq strategy and genomic features.

Sonicated chromatin (**A**) has been separated in two fractions: sonication-resistant/large fragments (**B**) and sonication-sensitive/small fragments (**C**). Large fragments have been further sonicated to a size usable for sequencing (**D**). (**E**) Validation of srHC enrichment at MSAT and Eef1a in definitive endoderm and hepatocytes. (**F**) Percentage of genome and (**G**) genomic features covered by srHC domains in definitive endodermal cells, hepatocytes and mature beta cells. (**H**) Boxplot representation of srHC (large fragments) versus open chromatin (small fragments) ratio in relation to gene expression in definitive endodermal cells (top), adult hepatocytes (middle) and adult beta cells (bottom). Gene expression, indicated as percentile, couples with genes embedded in chromatin-sensitive portion of the genome, thus showing smaller large/small fragment ratio. (**I**) Venn diagrams showing overlap between srHC and Hi-C B compartment (A-B compartment data from series GSE93431) in adult hepatocytes and mature beta cells.

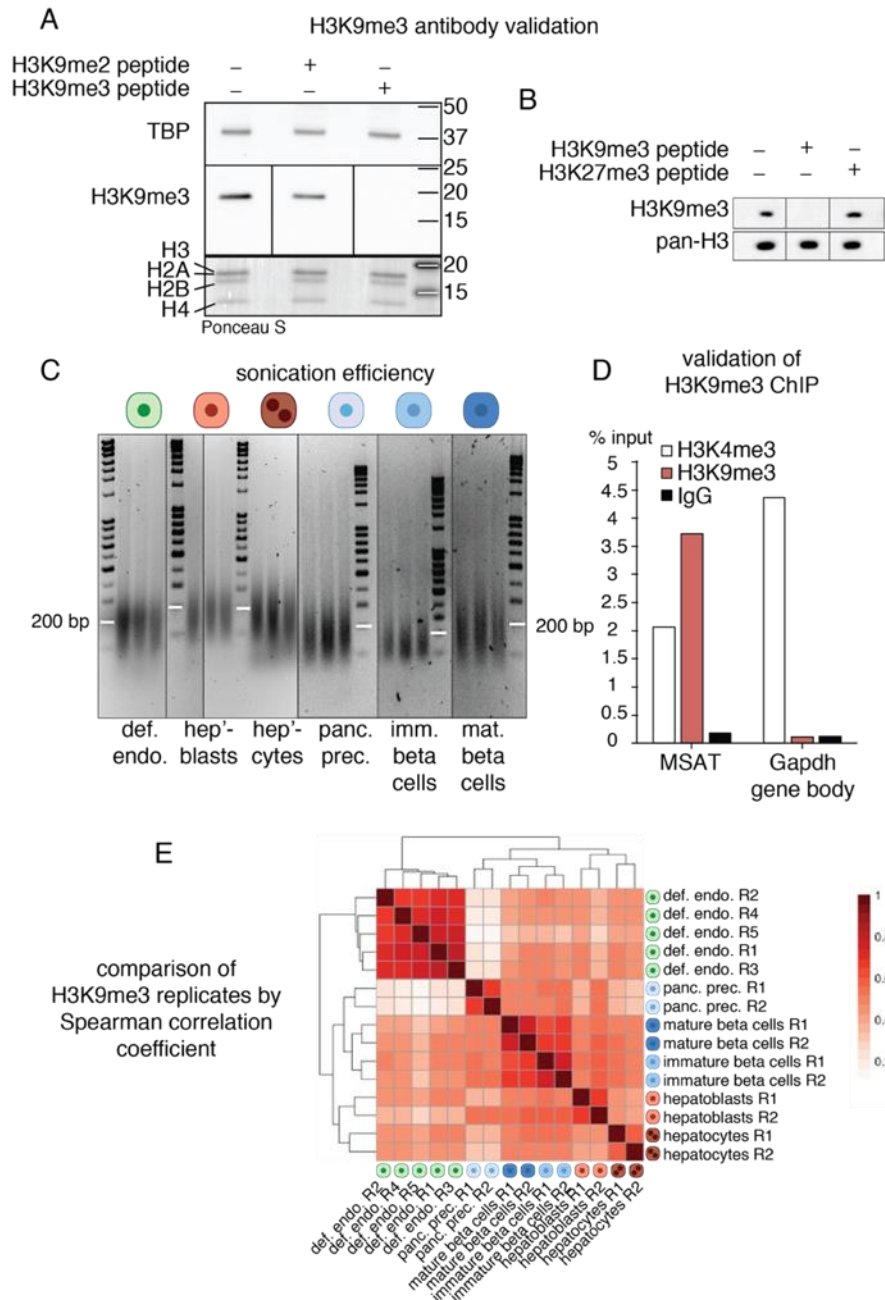


Fig. S5: H3K9me3 antibody validation and sample correlation.

Peptide competition assay on adult liver **(A)** and sorted definitive endoderm cells **(B)** protein extracts, testing H3K9me3 antibody specificity. TATA-Binding Protein (TBP) and pan-H3 antibodies were used as loading control in **(A)** and **(B)**, respectively. Ponceau staining of histone is indicated at the bottom of the panel **(A)**. **(C)** Representative sonication pattern of sorted cells. Cells have been prepared as indicated in the Material and Methods section. **(D)** Representative H3K9me3 and H3K4me3 ChIP-qPCR for sorted cells. The chart shows enrichment of H3K9me3 and H3K4me3 at major satellite repeats and Gapdh gene body. **(E)** Spearman correlation of H3K9me3 ChIP-Seq individual replicates considered in this study.

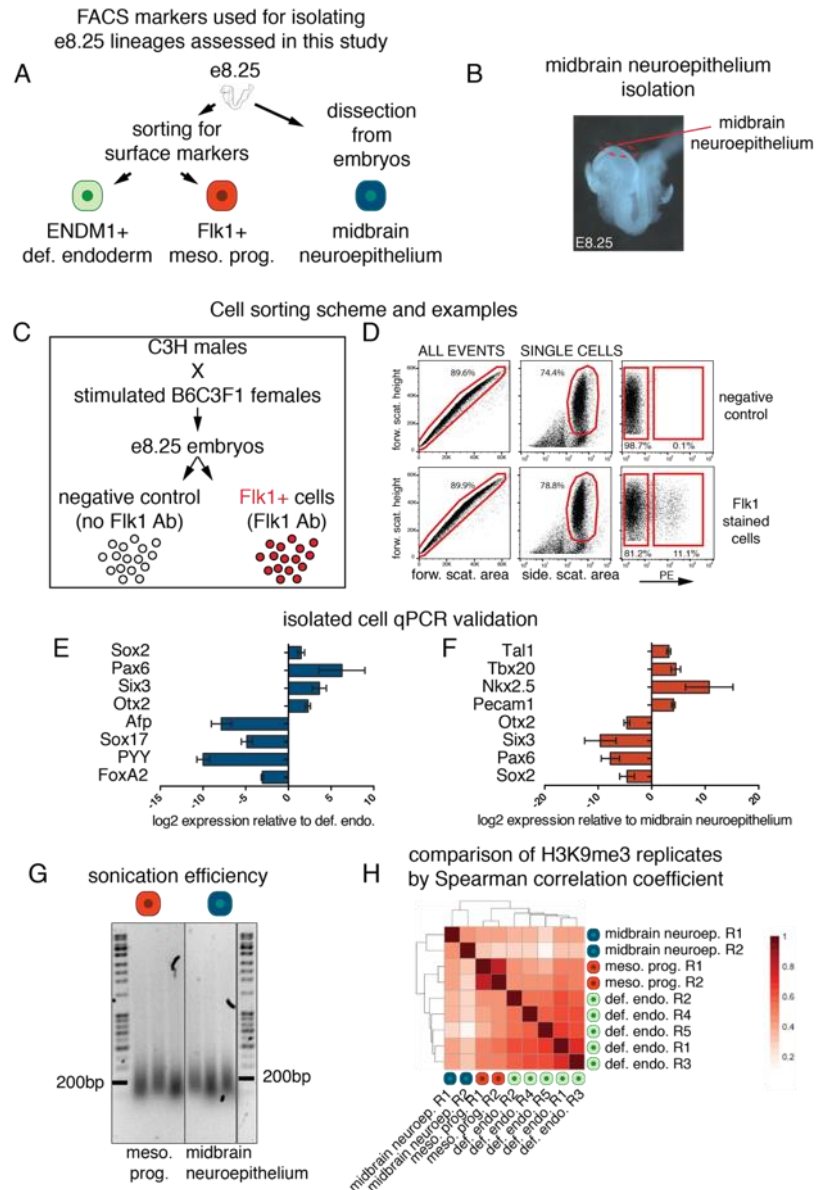


Fig. S6: E8.25 mesoderm progenitor and midbrain neuroepithelial cell identification. (A) Schematic of the cell types isolated at e8.25. (B) Midbrain neuroepithelial cells were isolated dissecting the neural fold from e8.25 embryos. (C) Schematic representing the sorting strategy of mesodermal progenitor Flk1+ cells from E8.25 embryos. (D) Representative sorting pattern of Flk1+ cells. (E and F) qPCR of mesoderm progenitor Flk1+ cells (E) and midbrain neuroepithelial cells (F) confirming the identity of the isolated cells. Ct values have been normalized to GAPDH and indicated populations of isolated cells have been used as reference. Expression is shown as Log2 scale relative to reference population. Three independent biological replicates have been considered for each quantification. (G) Representative sonication pattern of mesoderm progenitor Flk1+ and midbrain neuroepithelial cells. (H) Spearman correlation of individual replicates of samples isolated from e8.25 embryos.

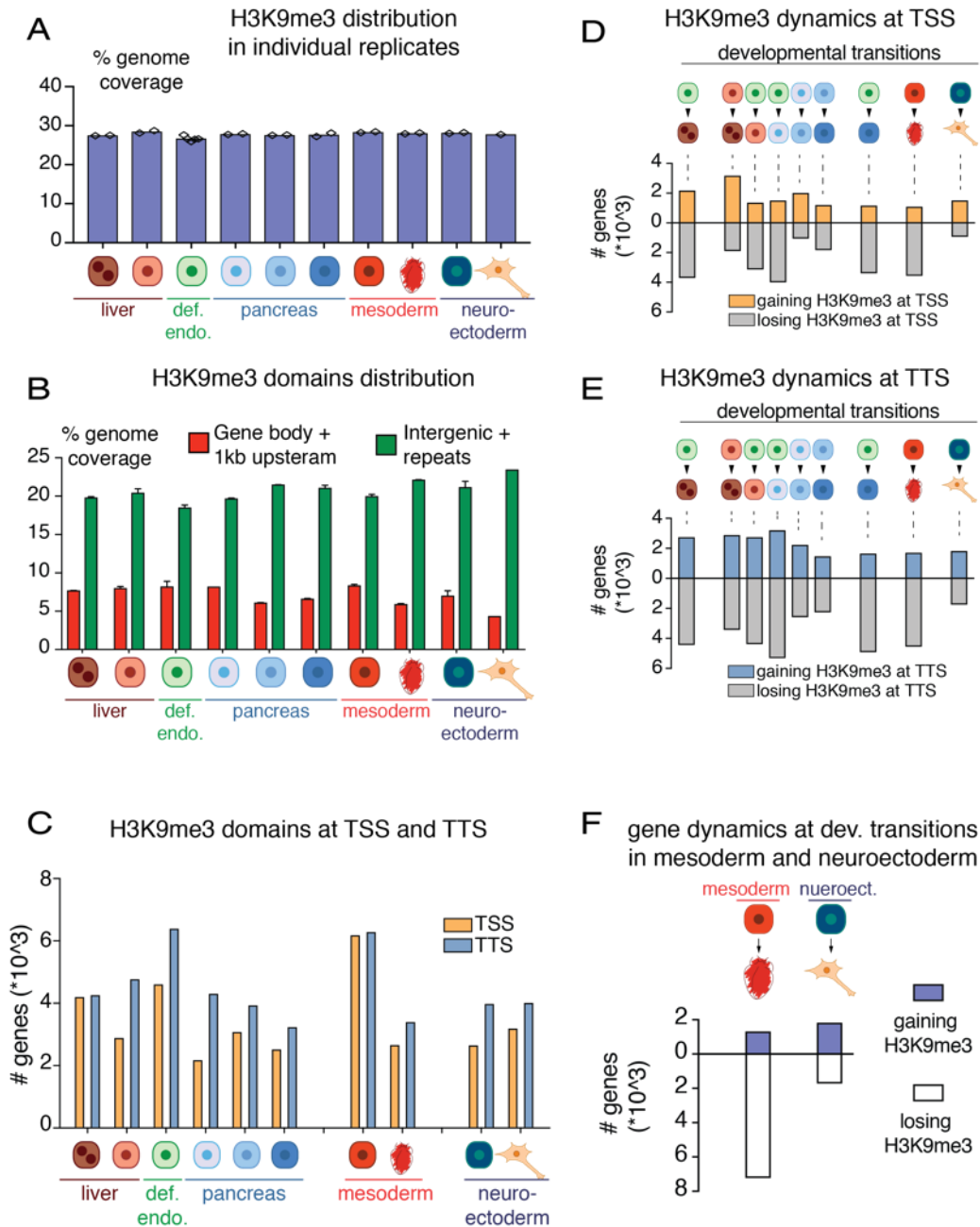


Fig. S7: H3K9me3 dynamics at genomic features.

(A) Percentage of genome and (B) genomic features covered by H3K9me3 domains in the cell types and embryonic developmental stages considered in this study. (C) Number of genes marked by H3K9me3 in the indicated cell types at the transcriptional start site (TSS, +/- 1kb) and termination site (TTS, +/- 1kb). (D and E) Number of genes gaining or losing H3K9me3 at TSS (D) or TTS (E) upon differentiation. Each transition is indicated above the corresponding bar indicating the gene number. (F) Number of genes gaining (purple) or losing (white) H3K9me3 in mesoderm and midbrain differentiation. Each transition is indicated above the corresponding bar indicating the gene number.

srHC, H3K9me3 and H3K27me3 overlap

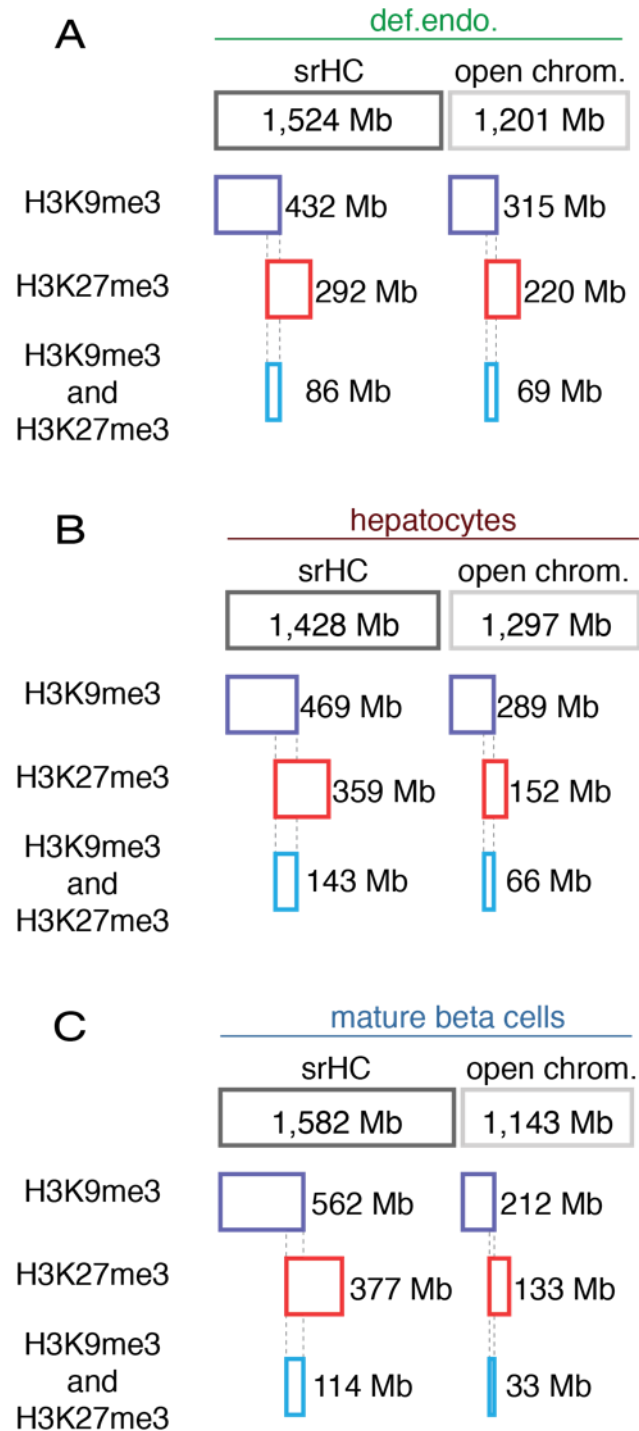


Fig. S8: Extent of srHC, H3K9me3, and H3K27me3 overlap. (A to C) Mb distribution of H3K9me3 and H3K27me3 in srHC and open chromatin in definitive endodermal cells (A), adult hepatocytes (B) and mature beta cells (C).

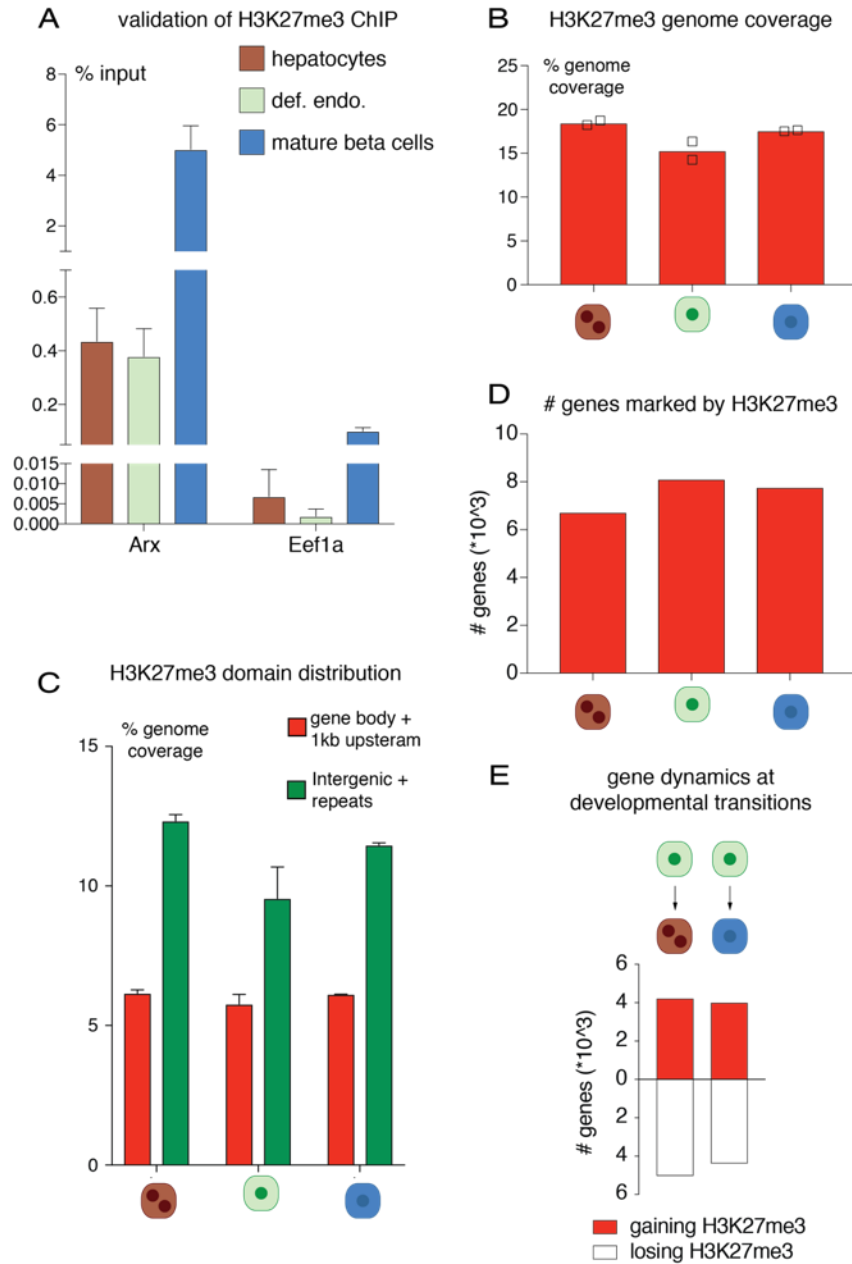
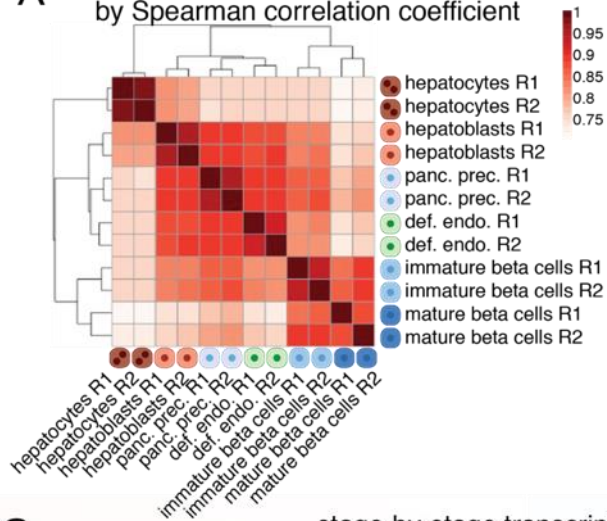


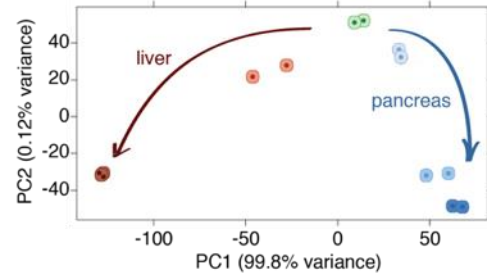
Fig. S9: H3K27me3 characterization during development.

(A) Representative H3K27me3 enrichment detected via ChIP-qPCR, in the indicated cell populations, at Arx (positive site) and Eef1a (negative site). (B and C) Percentage of genome (B) and genomic features (C) covered by H3K27me3 domains in the indicated cell populations. (D) Number of genes marked by H3K27me3 in each indicated stage. (E) Number of genes gaining (red) or losing (white) H3K27me3 upon step-wise transition in successive developmental stages. Each transition is indicated above the corresponding bar indicating the gene number.

A comparison of RNA-Seq Replicates by Spearman correlation coefficient



B PCA analysis



C

stage-by-stage transcriptome analysis

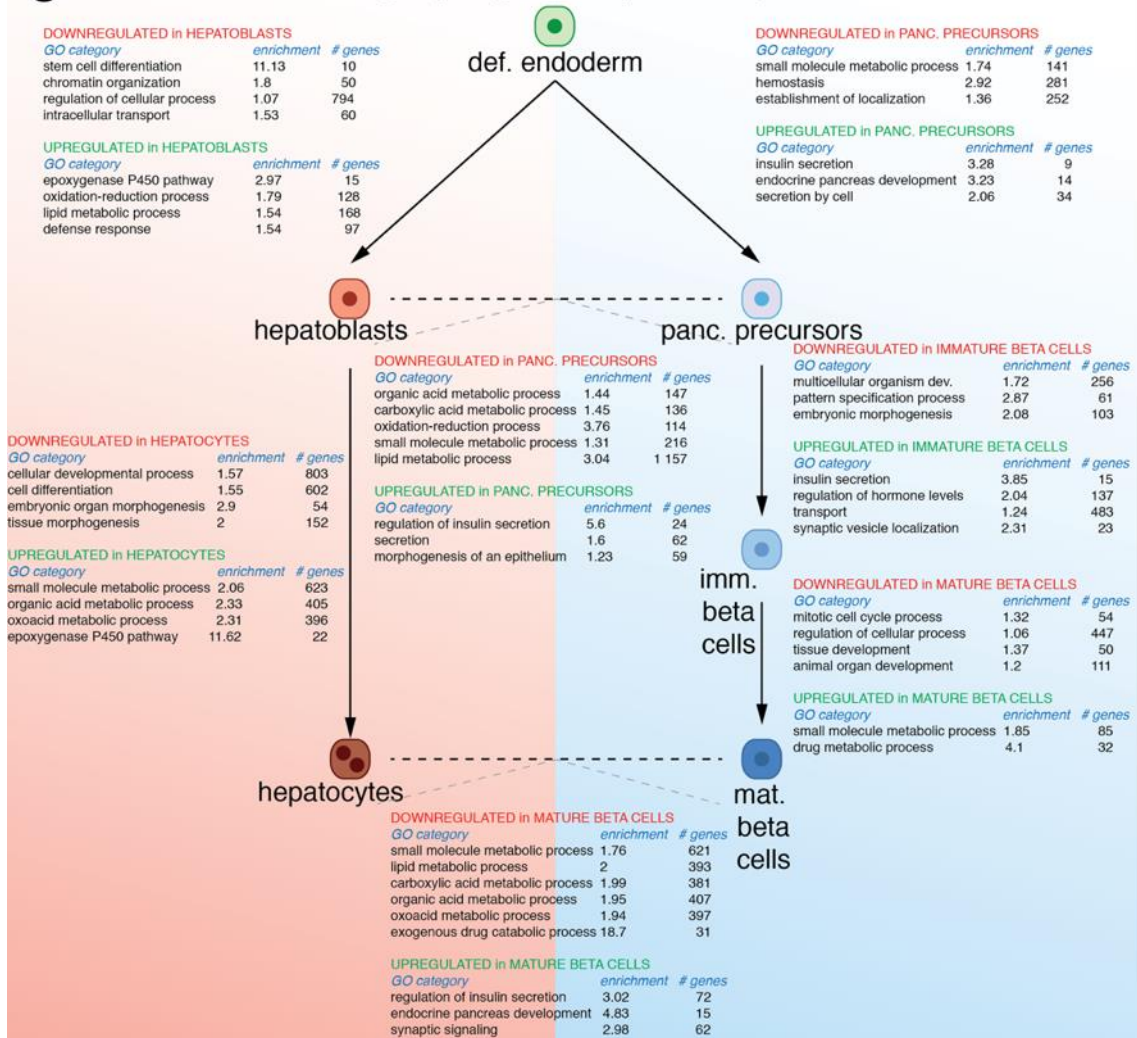


Fig. S10: Stage-by-stage transcriptome analysis.

(A) Spearman correlation of individual RNA-Seq replicates for endoderm differentiation along the hepatic and pancreatic samples. **(B)** PCA analysis of RNA-Seq data for the endoderm lineage samples considered in the study. **(C)** Representative GO categories and genome browser tracks for genes differentially expressed upon step-wise transition along the hepatic and pancreatic lineages. Comparison of e10.5 hepatoblasts vs pancreatic precursor cells and 2-month old hepatocytes vs mature beta cells have been included. For a comprehensive GO analysis, see Table S12.

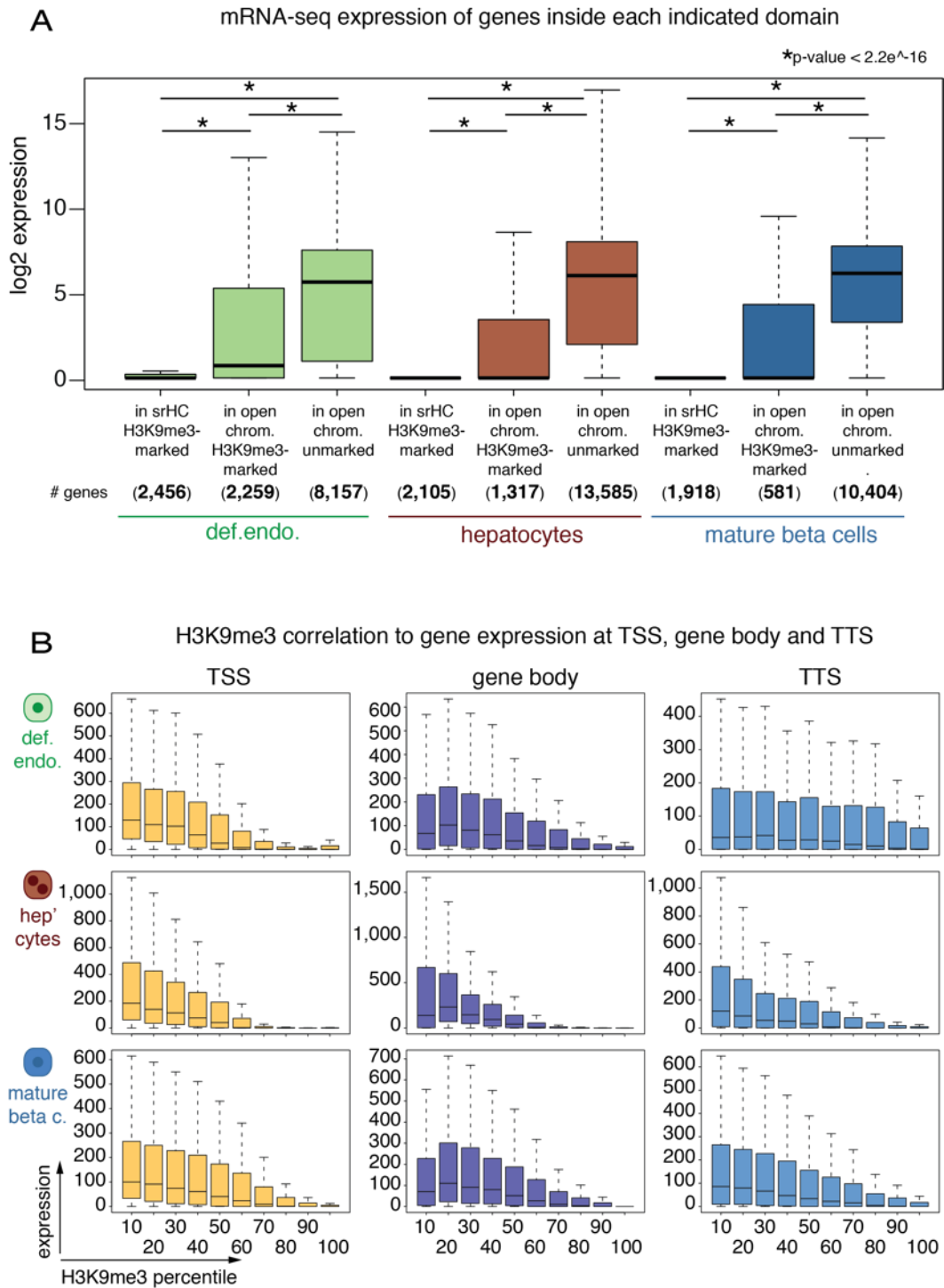


Fig. S11: srHC and H3K9me3 correlation to gene transcription.

(A) mRNA expression in definitive endoderm, adult hepatocytes and mature beta cells, for genes marked with H3K9me3 and present in srHC or open chromatin. Unmarked genes in open chromatin are indicated. (B) Correlation between H3K9me3 enrichment at TSS, gene body and TTS, and gene transcription in definitive endodermal cells, adult hepatocytes and mature beta cells.

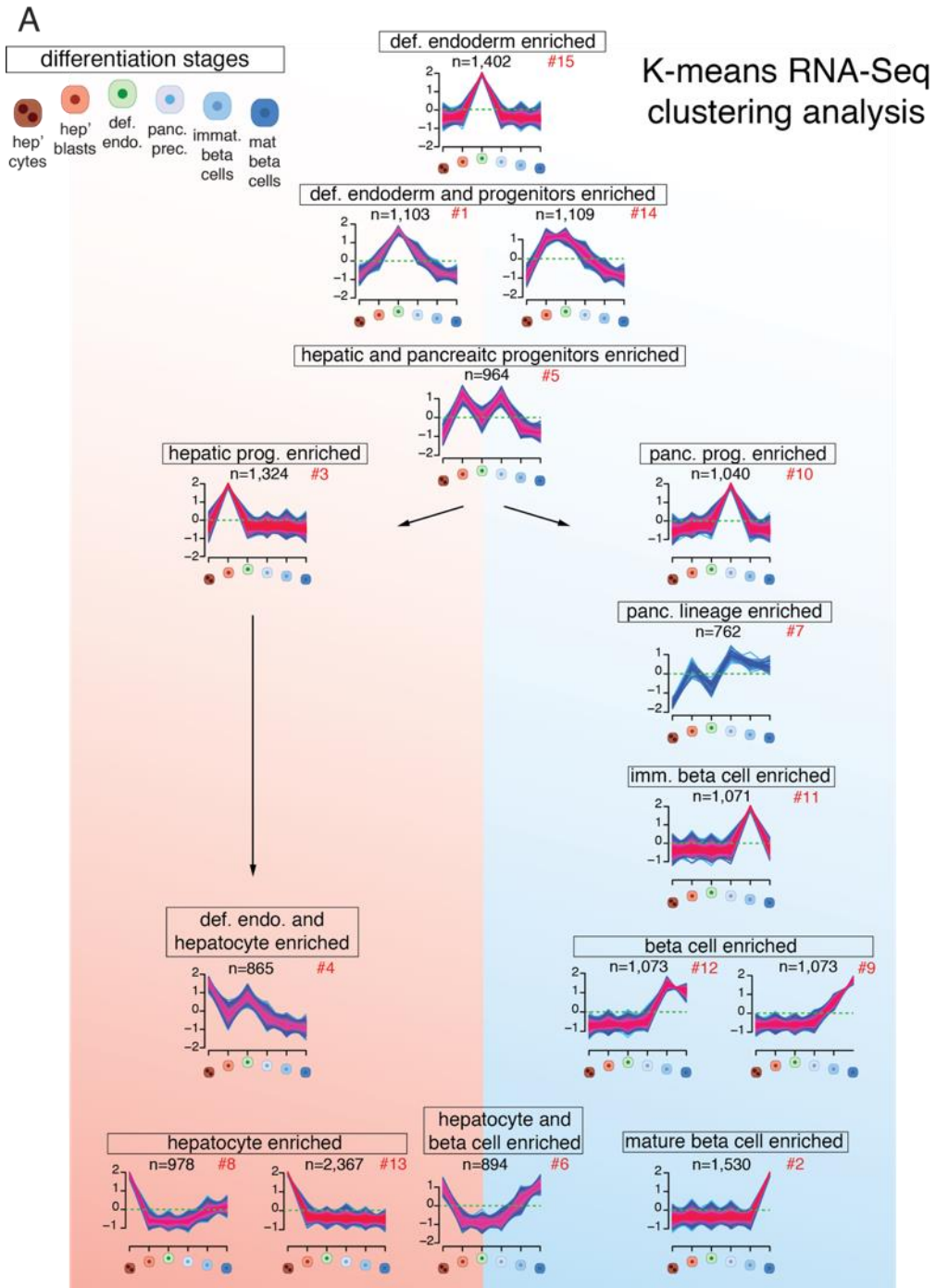


Fig. S12: RNA-seq cluster analysis.

(A) Cluster analysis of Z-score RNA-Seq values for endoderm differentiation along the hepatic and pancreatic samples. Each chart shows a profile of gene expression indicated as Z-score value. The numbers of cluster considered has been identified as described in Materials and methods. Clusters have been organized according to temporal expression of genes upon endoderm differentiation along the hepatic and pancreatic lineages. The red number beside each chart identifies the cluster number reported on Table S13.

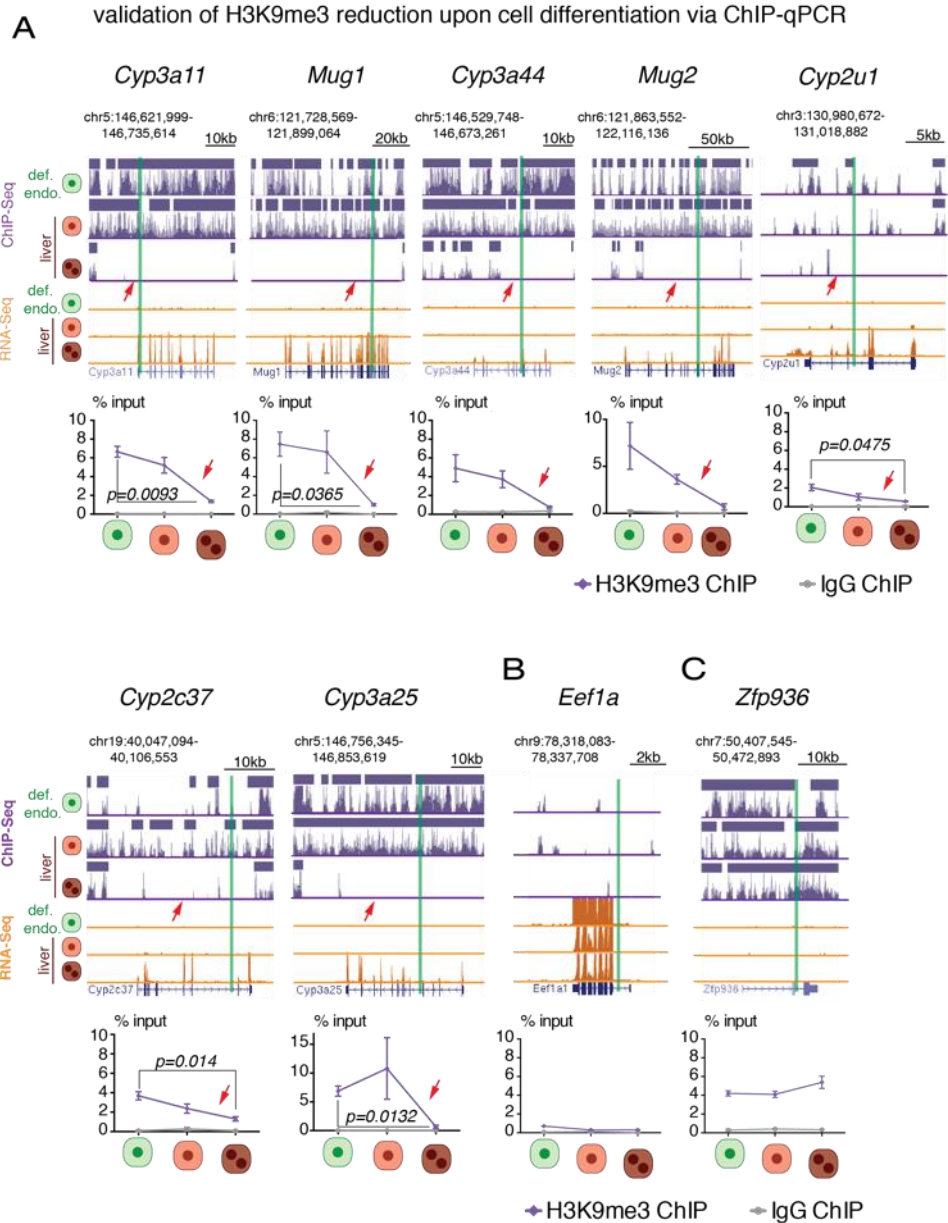


Fig. S13: ChIP-qPCR validation of H3K9m3 dynamics at hepatic genes.

(A to C) Representative UCSC genome browser tracks of input-divided H3K9me3 (purple) and RNA-Seq profiles (orange) upon definitive endoderm differentiation into adult hepatocytes. H3K9me3 patches are shown as purple bars above each profile. A ChIP-qPCR validation of candidate genes gaining expression upon H3K9me3 loss along the hepatic differentiation pathway is shown: H3K9me3 ChIP-qPCR (n=3) for (A) Cyp3a11, Mug1, Cyp3a44, Mug2 and Cyp2u1, Cyp2c37, Cyp3a25 showing progressive loss of H3K9me3 (red arrows) associated with gaining in gene expression. (B and C) show H3K9me3 enrichment at the constitutive active actin b gene and the permanently silenced zinc finger protein (Zfp) 936 gene. Each chart shows H3K9me3 and IgG enrichment at the indicated genic position (green bar). Results are mean. Error bars are SEM. Oligonucleotide sequences used for this experiment are indicated in Table S2.

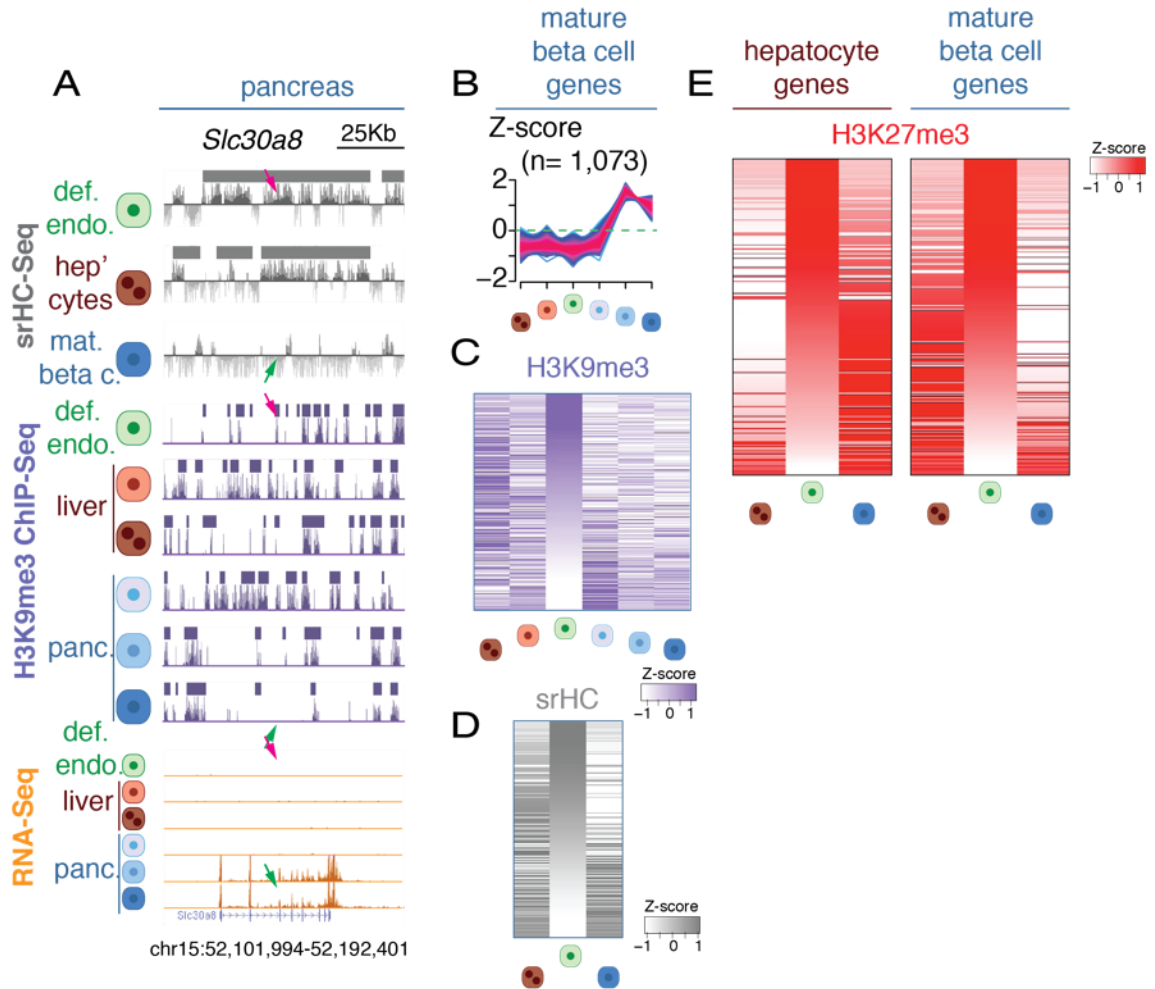


Fig. S14: srHC, H3K9me3 and H3K27me3 dynamics at lineage-specific genes.

(A) Representative UCSC genome browser tracks of srHC-Seq (grey), input-divided H3K9me3 (purple) and RNA-Seq profiles (orange) upon definitive endoderm differentiation into mature beta cells. SrHC and H3K9me3 patches are shown as grey and purple bars above each profile, respectively. *Slc30a8* on chromosome 15 (A) is shown. Magenta arrows indicate presence of srHC and H3K9me3 and absence of expression. Green arrows indicate absence of srHC, H3K9me3 and gene expression. (B) Z-score cluster representations for genes expressed in mature beta cells. (C and D) Heatmaps showing levels of srHC (C) and H3K9me3 (D) in the indicated stages. (E) Heatmaps showing levels of H3K27me3 in the hepatic and pancreatic specific genes identified with the K-means clustering. For each heatmaps definitive endodermal cells values have been ordered in a descendent manner.

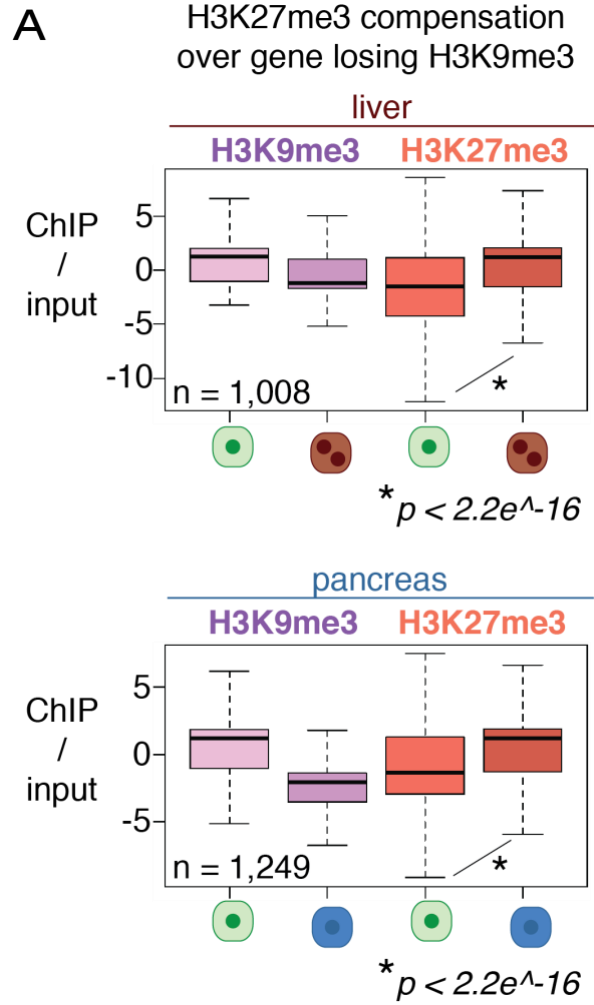


Fig. S15: H3K27me3 compensatory effects on genes losing H3K9me3.
(A) Boxplot showing H3K27me3 compensation on hepatic (top) and pancreatic (bottom) genes losing H3K9me3.

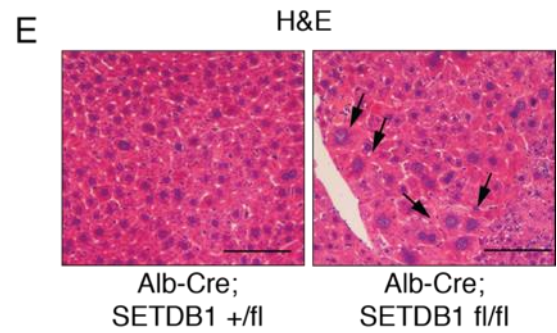
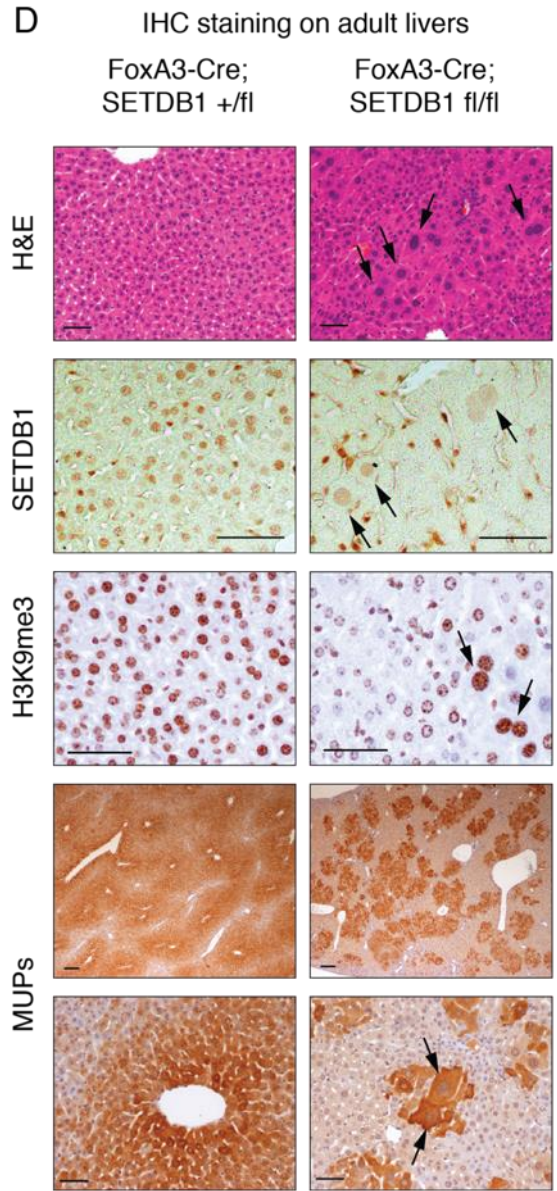
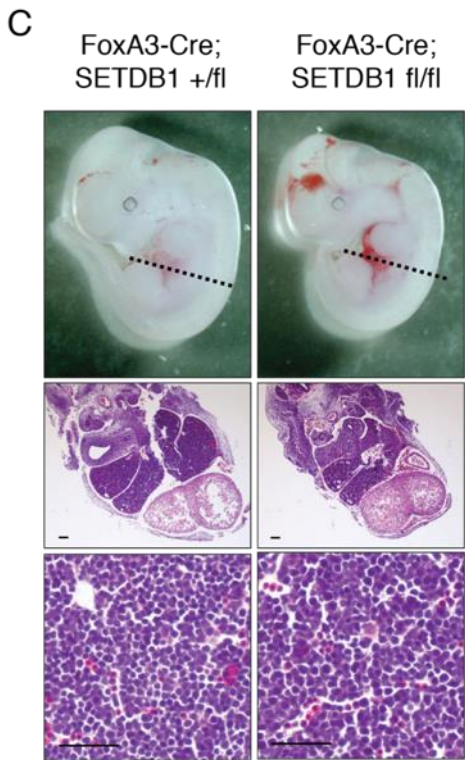
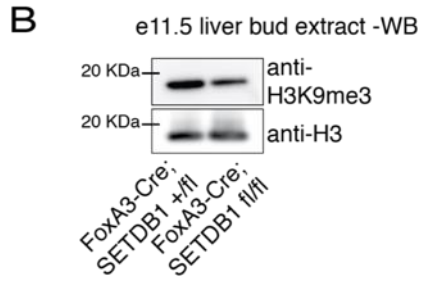
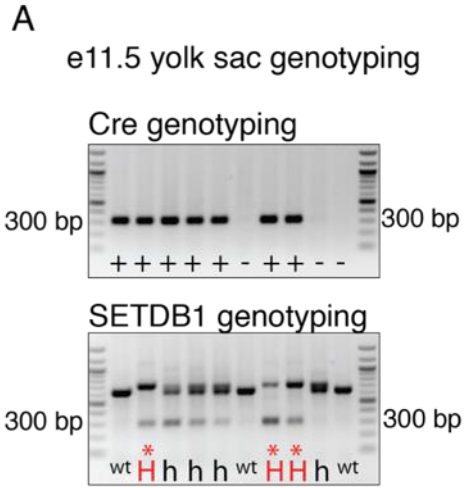


Fig. S16: E11.5 and adult Setdb1 mutants identification and characterization.

(A) E11.5 yolk sac genotyping for Cre (top) and SETDB1 (bottom). Note how in SETDB1 heterozygote and homozygote mutants a ~300bp band indicating rearrangement at the SETDB1 locus is detected. “+” and “-” in the top gel indicate samples positive or negative for Cre, respectively. In the bottom gel “wt” indicates Setdb1 wild type samples, “h” and “H” indicate samples where Setdb1 is floxed in one allele (heterozygous) or both (homozygous) alleles, respectively. Setdb1 homozygous floxed samples are also indicated by a red asterisk. **(B)** Western blot for H3K9me3 and panH3 in FoxA3-Cre; SETDB1 +/fl and FoxA3-Cre; SETDB1 fl/fl. **(C)** Top: representative images of FoxA3-Cre; SETDB1 +/fl and FoxA3-Cre; SETDB1 fl/fl e11.5 embryos. Middle: transverse sections stained with hematoxylin & eosin (H&E) of e11.5 FoxA3-Cre; SETDB1 +/fl and FoxA3-Cre; SETDB1 fl/fl embryos at the region indicated by the black, dashed line on top panels. Scale bar: 100 um. Bottom: H&E stained zoomed-in images of liver in FoxA3-Cre; SETDB1 +/fl and FoxA3-Cre; SETDB1 fl/fl e11.5 embryos. Scale bar: 50 um. **(D)** H&E staining and IHC for SETDB1, H3K9me3 and MUPs in 2-month old FoxA3-Cre; SETDB1 +/fl and FoxA3-Cre; SETDB1 fl/fl **(E)** H&E staining in 2-month old Alb-Cre; SETDB1 +/fl and Alb-Cre; SETDB1 fl/fl. In both **(D)** and **(E)** black arrows indicate overly big hepatocytes maintaining SETDB1 expression. Scale bar: 50 um.

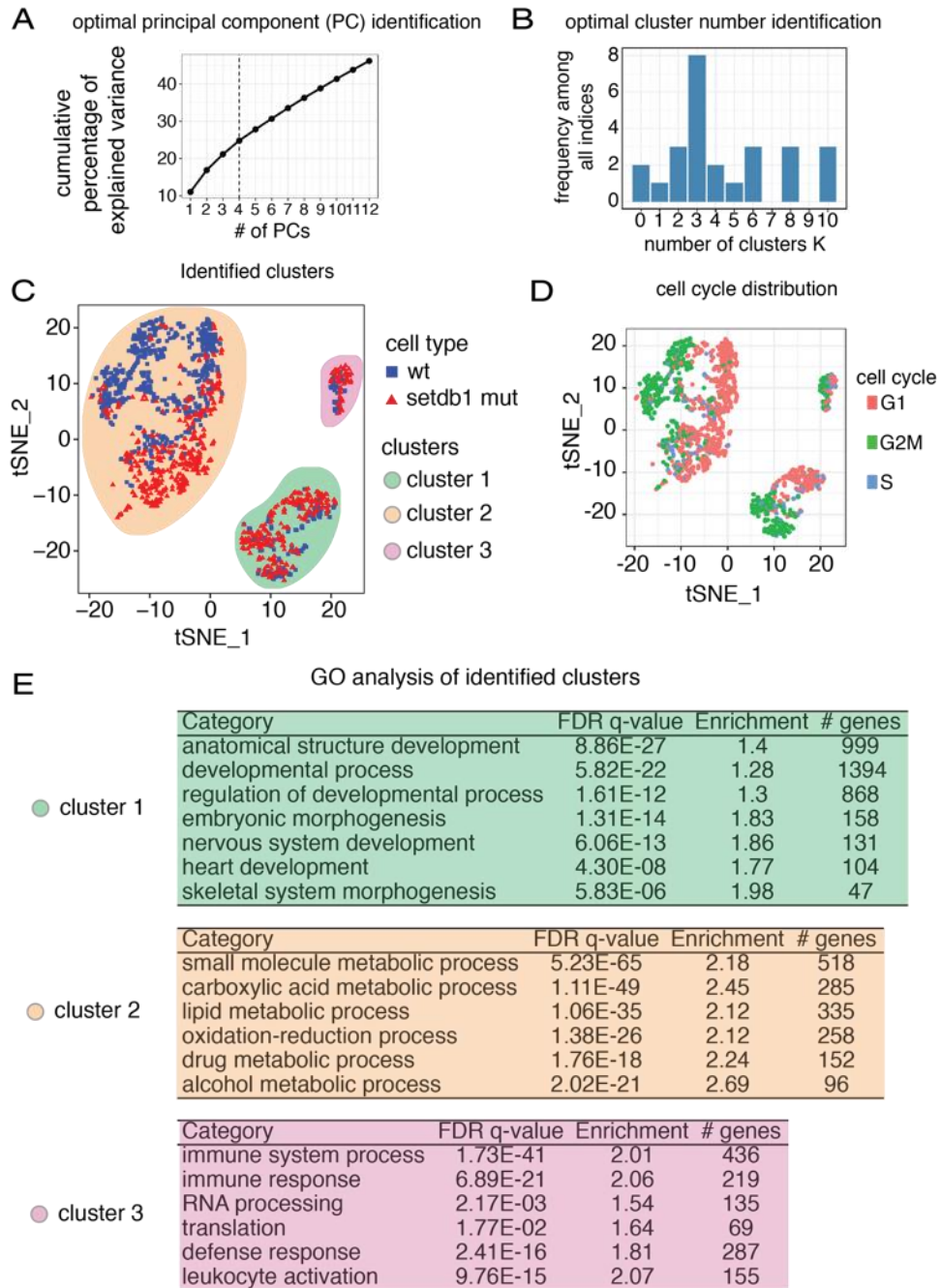


Fig. S17: Single cell RNA-Seq clustering and GO analysis of e11.5 wt and Setdb1 mutant hepatoblasts.

(A) Chart showing the optimal principal component (PC) for the data matrix of single cell RNA-Seq samples. (B) Histogram showing the optimal number of clusters for representing single cell RNA-Seq data. (C) t-distributed stochastic neighbor embedding (t-SNE) plot showing the three identified clusters. Wt and SETDB1 mutant cells are indicated as blue squares and red triangles, respectively. (D) t-SNE plot showing cell cycle phases in the sequenced cells. (E) Representative gene ontology categories for each of the three identified clusters.

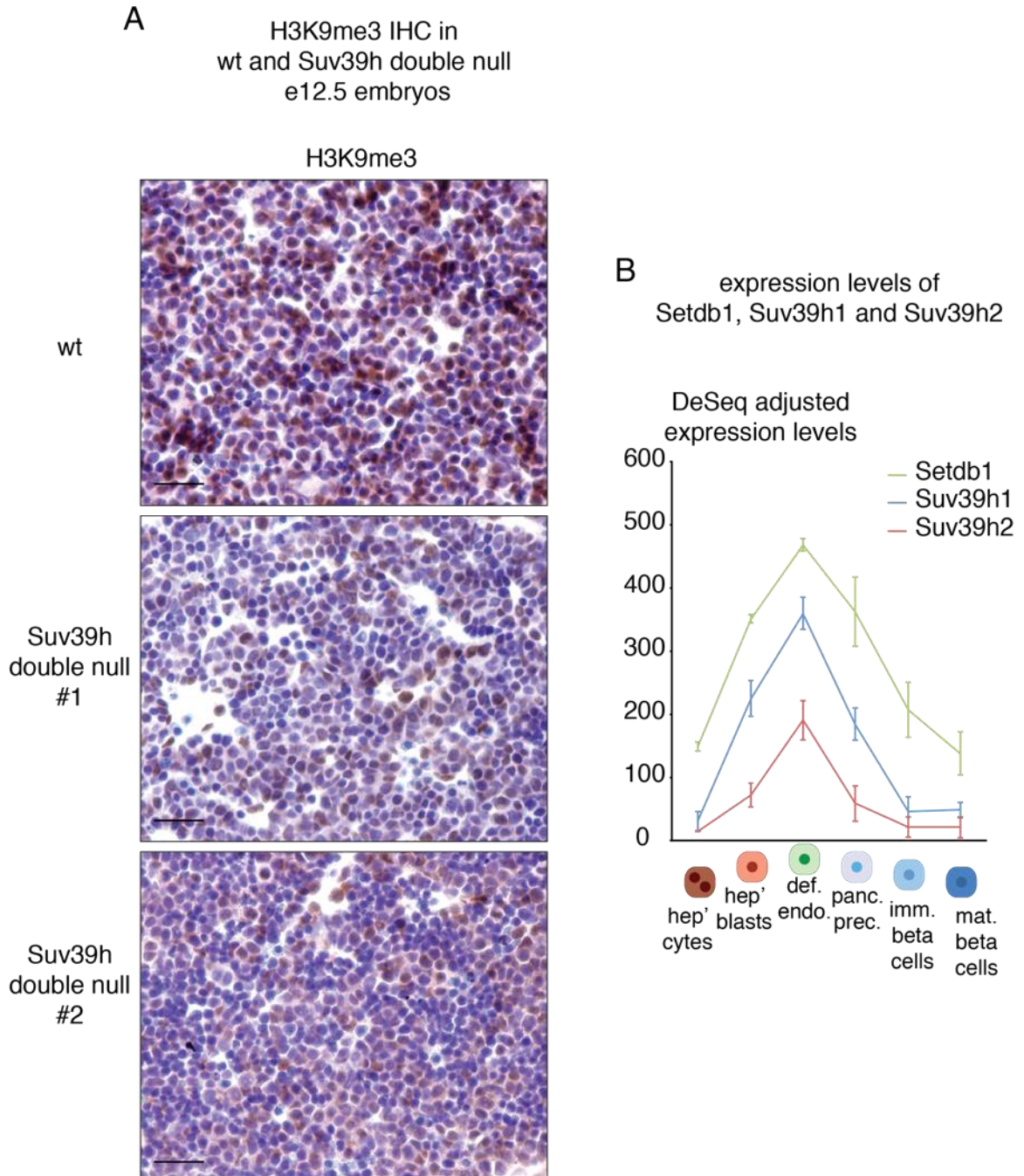


Fig. S18: Loss of Suv39h enzymes' effects on H3K9me3, and H3K9me3-related HMTase expression.

(A) IHC of e12.5 wt and two Suv39h double null livers showing a mild reduction in H3K9me3 levels. **(B)** Expression levels of Setdb1, Suv39h1 and Suv39h2 during definitive endodermal cell differentiation along the hepatic and pancreatic lineages.

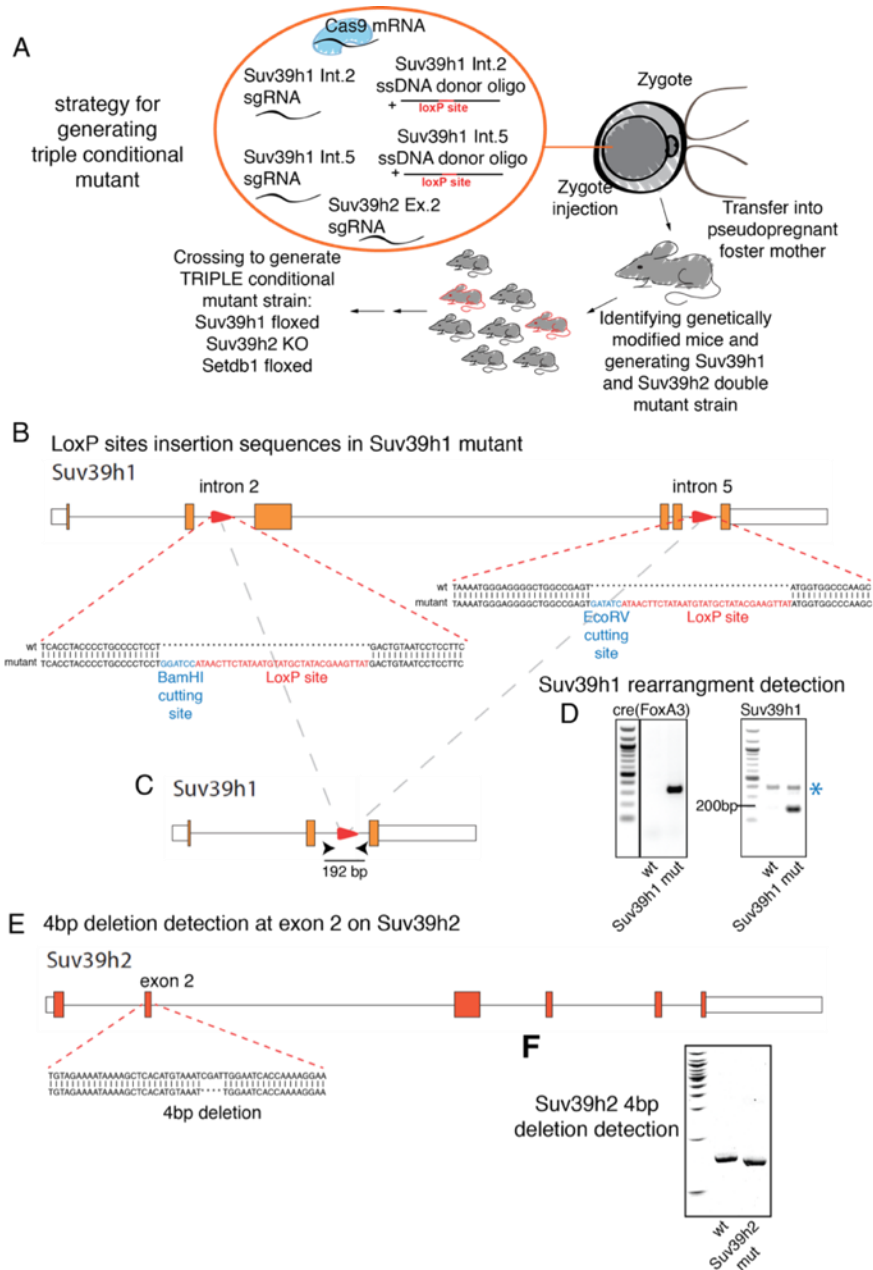


Fig S19: Strategy for the generation of Setdb1, Suv39h1, and Suv39h2 triple conditional mutant strain.

(A) Strategy for generating Setdb1, Suv39h1 and Suv39h2 triple conditional mutant mice. **(B)** Schematic of Suv39h1 gene. LoxP sites are indicated as red triangles on Intron 2 and Intron 5. DNA sequencing identified proper insertion of loxP sites in mutant mice compared to control. **(C)** Schematic of cre-driven rearrangement on Suv39h1. **(D)** Genotyping of control and mutant DNA from 1-month old liver. Suv39h1 rearrangement can be identified as 192 bp band on agarose gel. Asterisk indicates unspecific band. **(E)** Schematic of Suv39h2 gene. DNA sequencing identified 4bp deletion at Exon2. **(F)** Genotyping of control and mutant DNA from 1-month old liver. 4bp deletion leads to a lower band in 18% acrylamide gel.

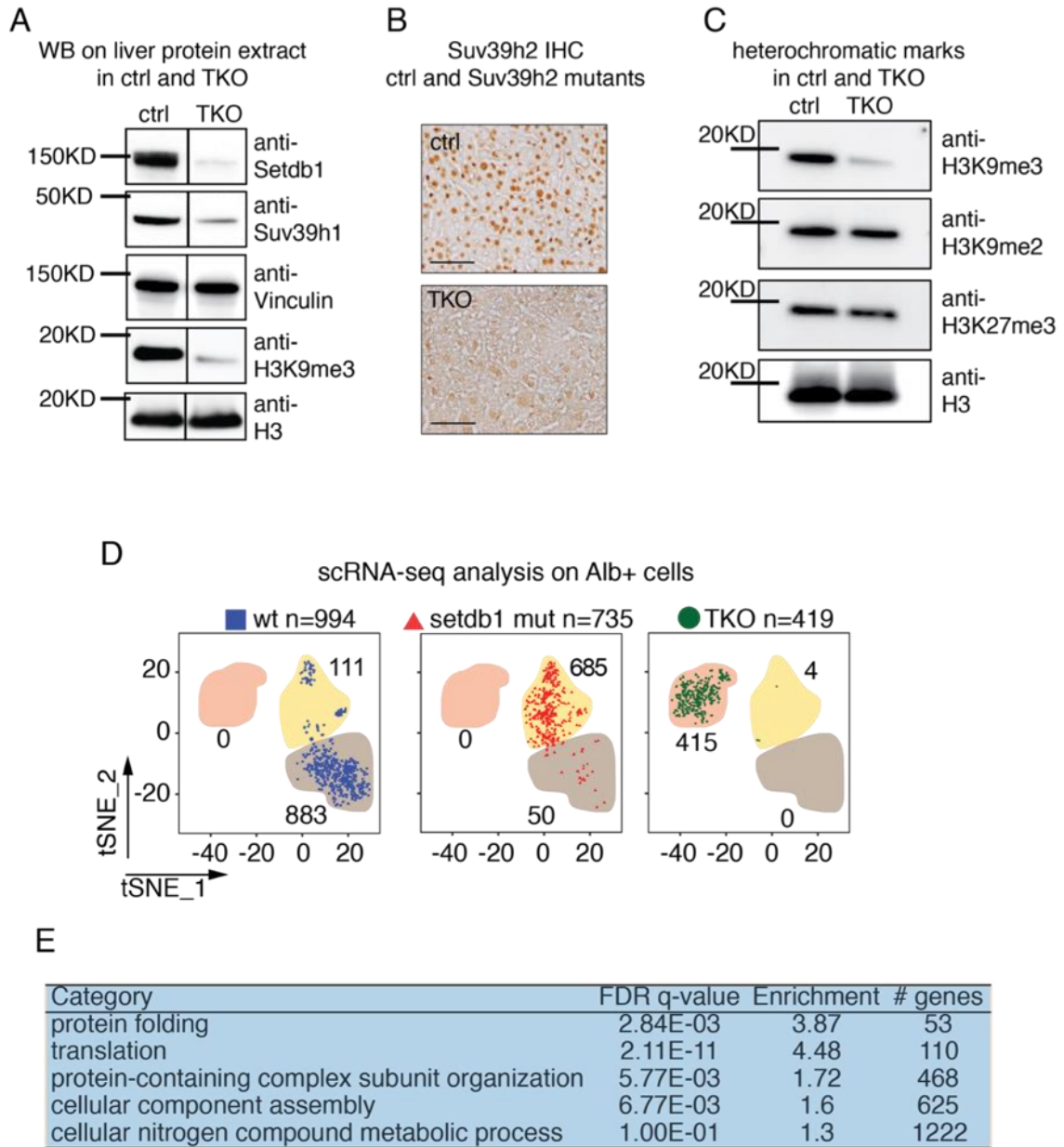


Fig. S20: TKO characterization and e11.5 scRNA-seq analysis.

(A) Western blot for Setdb1, Suv39h1, Vinculin, H3K9me3 and H3 in ctrl and Setdb1, Suv39h1 and Suv39h2 triple knock-out (TKO) liver protein extracts. **(B)** IHC for Suv39h2 in ctrl and Suv39h2 mutant livers. Scale bar: 50 um. **(C)** Western blot for H3K9me3, H3K9me2, H3K27me3 and panH3 in ctrl and TKO liver protein extracts. Please note that western blots in **(A)** and **(C)** have been performed using protein extracts from liver. Thus, in cells not expressing FoxA3 (e.g. biliary and endothelial cells), floxed alleles of *Setdb1* and *Suv39h1* are not recombined. This explains the residual H3K9me3 signal. **(D)** tSNE plots of single cell RNA-Seq data from e11.5 Liv2+ cells, showing wt (top, blue squared) Setdb1 mutant (bottom, red triangles) and TKO albumin positive cells distribution. **(E)** Representative gene ontology categories for TKO cells shown in Fig.3A.

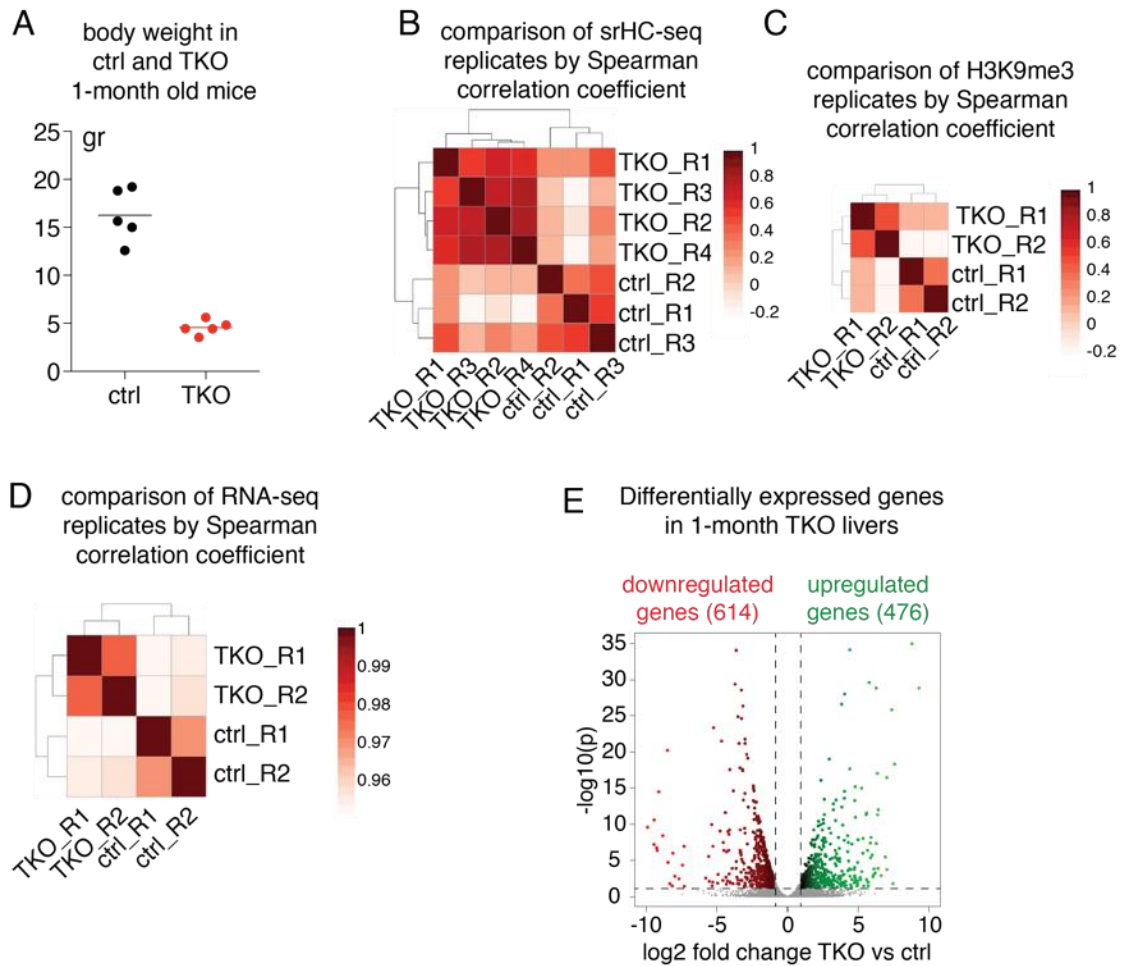


Fig. S21: One-month TKO characterization.

(A) Chart indicating body weight (gr) of ctrl (n=5) and TKO (n=5) mice. (B) Spearman correlation of srHC-seq individual replicates in ctrl and TKO livers. (C) Spearman correlation of H3K9me3 ChIP-seq individual replicates in ctrl and TKO livers. (D) Spearman correlation of RNA-seq individual replicates in ctrl and TKO livers. (E) Volcano plot showing downregulated (red) and upregulated (green) genes in TKO livers compared to ctrl.

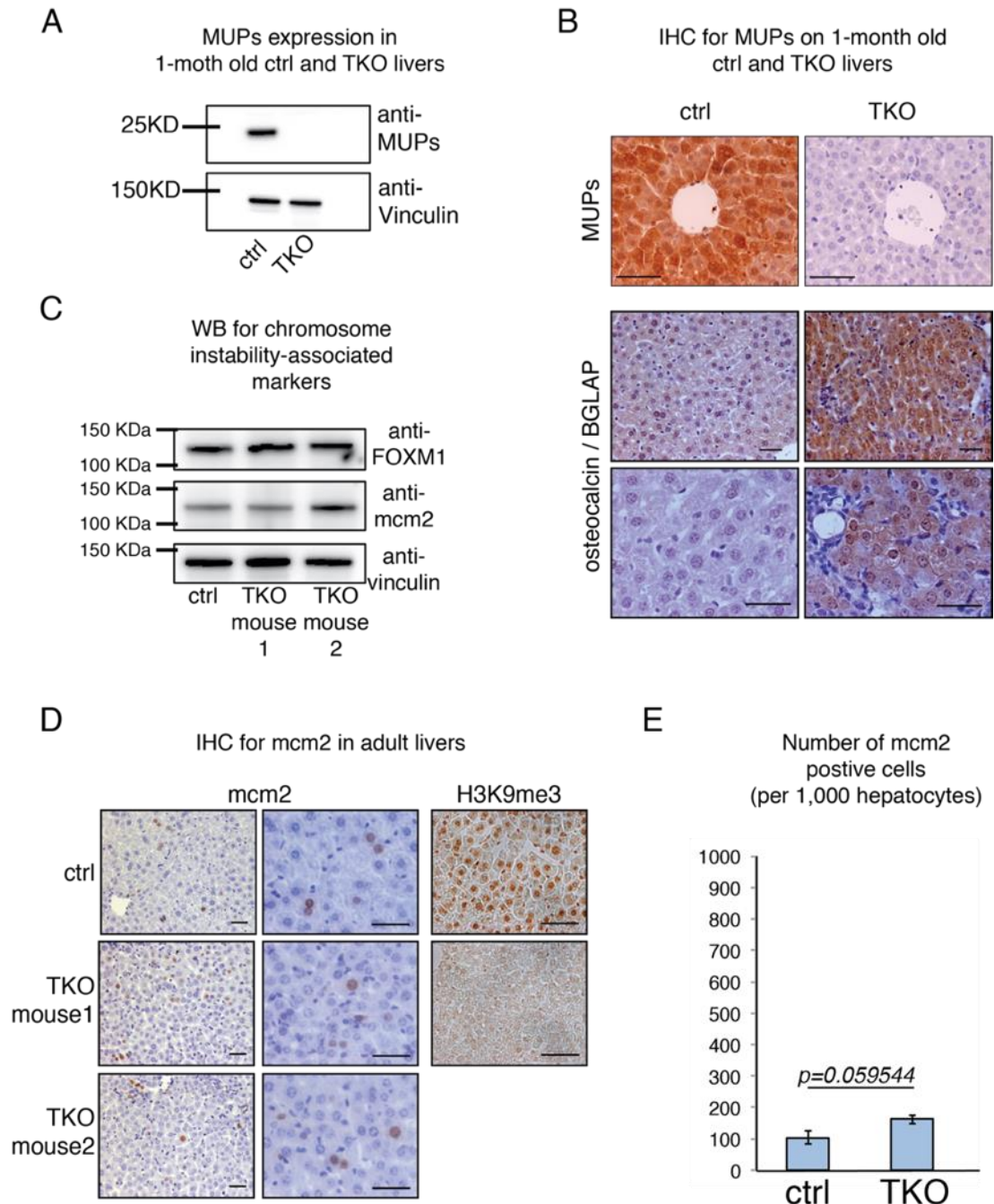


Fig. S22: TKO livers deregulate cell identity-related genes while chromosomal instability-associated markers' expression is unchanged.

(A) Western blot for MUPs and Vinculin in ctrl and TKO liver protein extracts. **(B)** IHC for MUPs and osteocalcin in ctrl and TKO mutant livers. Scale bar: 50 μ m. **(C)** Western blot for chromosome instability associated markers FOXM1 and mcm2 in ctrl and TKO livers. Vinculin was used as loading control. **(D)** IHC for mcm2 and H3K9me3 in ctrl and TKO mutant livers. Scale bar: 50 μ m. **(E)** Quantification of mcm2 positive cells in ctrl and TKO livers per 1,000 hepatocytes.

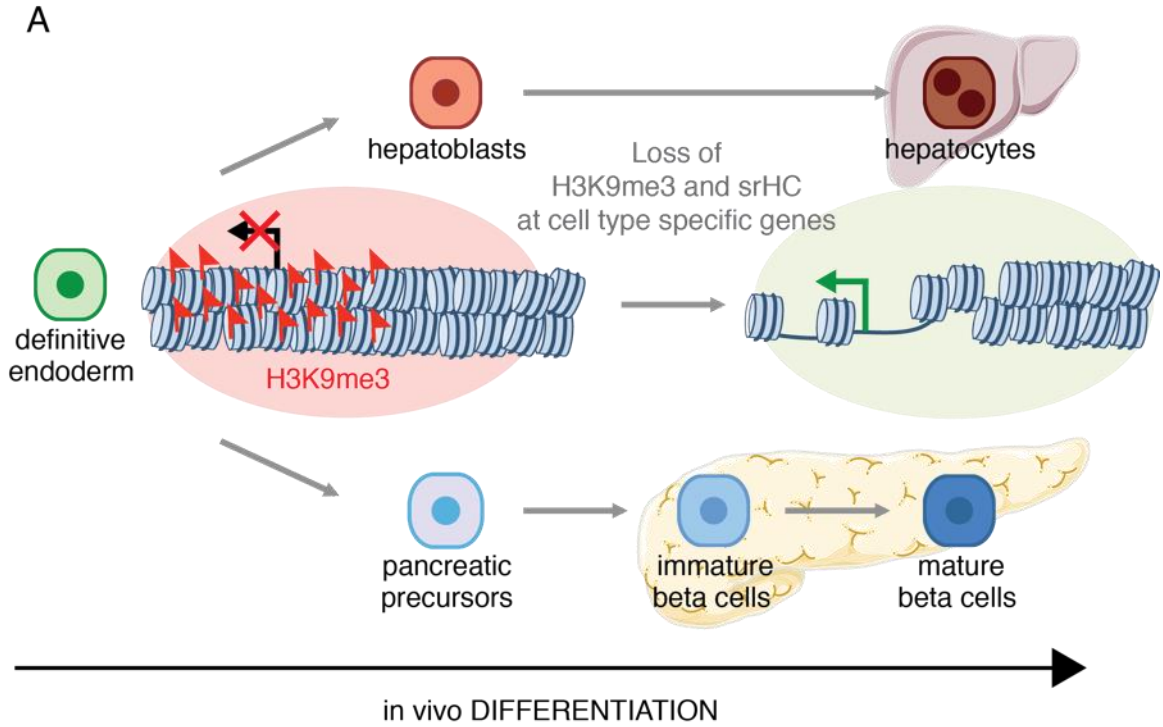


Fig. S23: Summary Model.

Lineage-specific genes expressed in mature cells are marked by H3K9me3, buried into srHC domains, and thereby repressed in early, uncommitted cells. During cell differentiation, many cell-appropriate H3K9me3 heterochromatic genes are opened up and expressed, while cell-inappropriate genes remain heterochromatic and silent.

3.7. Methods

Mouse Strains

Three weeks old F1 female progeny of C57BL/6 and CH3 crosses (B6C3F1) were regularly purchased from The Jackson Laboratory. Pdx1-eGFP, Ins-RFP, FoxA3^{Cre} (28) were used as described. SETDB1^{fl/fl} murine strain was a kind gift from Dr. Shinkai (30). Mice deleted for SETDB1 in foregut endoderm were generated by mating FoxA3^{Cre}; SETDB1^{fl/+} mice to one another. Suv39h1^{fl/fl} and Suv39h2^{KO/KO} mice were generated as described below.

Dissections and cell preparation for sorting

B6C3F1 females were superovulated by intraperitoneal injection (5 IU pregnant mare serum – PMSG, Prospec, cat# hor-272) followed, 48 h later, by a second injection (5 IU human chorionic gonadotropin - HCG, Peprotech cat#: 100-39-10UG), and then bred to C3H or Pdx1-eGFP or Ins-RFP males. To generate single cells for FACS, embryos at different developmental stages (i.e. e8.25, e10.5, e11.5, e18.5) were dissected out of decidua and collected in 1.5 ml tubes containing 1X PBS on ice. For e10.5 and e11.5 embryos, most of the dorsal part (somites) was removed. E18.5 pancreata, clearly visible and easy to distinguish, were dissociated from the rest of the organs. Collected samples were dissociated with 0.05% Trypsin-EDTA (Life Technologies, #25300-054) for 5 min at 37°C while gently shaking. The reaction was stopped by adding 0.4X volumes of FBS (Hyclone, SH30071). Explants were homogenized by gentle pipetting. Cells were spun down at 1500 rpm for 5min at 4°C, and the pellet was resuspended in 320 ul of IMDM (Life Technologies, #12440-053) (supplemented with L-glut – Life Technologies #25030-081, 0.1%BSA - Sigma A9418-50G, 2.5% FBS - Hyclone, SH30071, 1X pen/strep – Life

Technologies, #15140-122). Twenty (20) ul were transferred into a new tube and 280 ul IMDM were added to it (negative control tube - i.e. samples non incubated with the primary antibody). E8.25 definitive endodermal cells were purified using an antibody detecting ENDM1 (dilution: 1:50) surface antigen(31–33). Mesoderm progenitor cells were purified using anti-Flk1 antibody (dilution1:50). E10.5 hepatoblasts were sorted from Pdx1-GFP transgenic embryos, staining the cells for the surface marker Liv2 (MBL, cat# D118-3, dilution: 1:100). At the same stage pancreatic precursor Pdx1-GFP positive cells were collected. A subpopulation of Liv2+/Pdx1-GFP+ cells was detected at E10.5, as previously reported (33). These cells were not considered for further analysis. E11.5 hepatoblasts were sorted from wt embryos using the same antibody and conditions as for e10.5 embryos. Cells were incubated for 30-40min with primary antibody. Tubes were flicked every 5-8min. One (1) ml IMDM (supplemented with L-glut – Life Technologies #25030-081, 0.1%BSA - Sigma A9418-50G, 2.5% FBS - Hyclone, SH30071, 1X pen/strep – Life Technologies, #15140-122) was added to both the sample and negative ctrl tube. Cells were spun down at 1500 rpm for 5 min at 4°C and the pellet was resuspended in 300 ul of IMDM (supplemented with L-glut – Life Technologies #25030-081, 0.1% BSA - Sigma A9418-50G, 2.5% FBS - Hyclone, SH30071, 1X pen/strep – Life Technologies, #15140-122). PE-conjugated goat anti rat IgG (Biolegend 405406) was added (dilution 1:300). Samples were mixed well and incubated for 30min on ice in the dark. Tubes were flicked every 5-8 min. One (1) ml IMDM (supplemented with L-glut – Life Technologies #25030-081, 0.1% BSA - Sigma A9418-50G, 2.5% FBS - Hyclone, SH30071, 1X pen/strep – Life Technologies, #15140-122) was added to both the sample and negative ctrl tube. Cells were spun down at 1500 rpm for 5 min at 4°C and the pellet was resuspended in 400-800 ul 1X PBS + 3% BSA (Sigma A9418-50G). Samples were filtered through a cell strainer (FALCON 5 ml polystyrene round-bottom tube with cell-strainer cap, #352235). Samples

were kept on ice until sorting. Mature beta cells were isolated from Ins-RFP transgenic embryos as described (34).

Isolation of adult hepatocytes and sample preparation for chromatin-related assay

Two (2) -month old mice were anesthetized using isoflurane (Butler Animal Health Supply, #029405) and monitored by paw pinch. A V-shaped incision was used to reveal the abdominal cavity, the intestines moved aside, and a 22 or 24 gauge catheter (Midwest Veterinary Supply, #381423) was used to cannulate the venae cavae while liver perfusion media pumps into the liver, and the portal vein is immediately severed. 37°C liver perfusion media (Invitrogen #17701-038) and then liver digest media (Invitrogen #17703-034) are perfused through the liver (45mL each per animal). Livers are dissociated with a cutting motion with cell scrapers in William's E Medium (Sigma W4128) and strained through a 100 um filter (BD 352360). Hepatocytes are loosely pelleted by centrifugation at 50 g for 5 min at 4°C. Isolated hepatocytes were resuspended in 1% formaldehyde (Fisher Chemical F79-500) William's E Medium (Sigma W4128) and rocked for 10 min at RT. Fixation was stopped by adding 2.5 M glycine (Fisher BioReagents BP381-1) followed by 5 min incubation at RT. Samples were spun down 50 g for 5 min at 4°C. Cells were resuspended in RSB medium (10 mM Tris pH7.4, 10mM NaCl, 3mM MgCl₂) freshly supplemented with 0.5% NP40, 0.1 mM PMSF and 1 Complete EDTA-free protease inhibitor cocktail tablet (Roche, #11873580001). Nuclei were isolated by douncing the samples for 25 strokes. Isolated nuclei were washed with Np40-free RSB medium and counted using a hemocytometer. Nuclei were aliquoted in 10⁶ nuclei aliquots and resuspended in 2 ml ice-cold sonication buffer (10 mM Tris-HCL pH8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) freshly supplemented with 0.1% Na-deoxycholate (Sigma Aldrich, cat #D6750), 0.5% N-lauroylsarcosine (TEKnova, cat #S3379), 1 mM DTT, 0.1

mM PMSF and 1 Complete EDTA-free protease inhibitor cocktail tablet (Roche, #11873580001). Thirty thousand (30,000) cell aliquots were snap-frozen and stored at -80 C. Sample sonication was performed using COVARIS (S22 Focused Ultrasonicator) and the following conditions: Incident peak power: 200; Duty Cycle: 10%; Burst/Cycle: 200; Time: 10-12min. Chromatin immunoprecipitation on adult hepatocytes was performed as described in (25).

Isolation of one-month old liver cells and sample preparation for chromatin-related assay

One (1) month old control and triple knockout (TKO) mutant liver were dissected and cut in pieces. A small piece was transferred to 1.5 ml tube. Samples were minced using dissecting scissors. One (1) ml ACK (Life Technologies, #A10492-01) solution was added and samples were incubated 5-7 min on ice. Samples were spun down at 1,500 rpm for 5 min at 4°C; pellet was resuspended in 500 ul 1XPBS and dounced with microcentrifuge tubes pellet pestle (DWK Life Sciences – Kimble - #7495150000) for 10-15 sec on ice. Samples were spun down at 1,500 rpm for 5 min at 4°C. Pellets were resuspended in RT 1 ml 1X PBS + 1% formaldehyde (Fisher Chemical F79-500), and incubated for 10 min at RT on rocking table. The reaction was stopped by adding 2M glycine (Fisher BioReagents BP381-1) and incubating the samples at RT for 5 min on a rocking table. Samples were spun down at 2,000 rpm for 10 min at 4°C. Samples were resuspended in 500 ul ice-cold washing buffer (1X PBS + 2% FBS - Hyclone, SH30071). Samples were spun down at 2,000 rpm for 5 min at 4°C, twice. Samples were resuspended in 1 ml nuclear lysis buffer (50mM Tris-Cl pH 8.0, 10mM EDTA, 1% SDS -Lonza, #51213, supplemented with protease inhibitors -1 tablet Complete Mini Protease inhibitor cocktail tablets (Roche, #04693159001) per 10 ml Buffer). Samples were incubated on ice for 15 minutes, flash

frozen and stored at -80°C until used. Samples were sonicated as indicated for adult hepatocytes.

Generation of Suv39h1, Suv39h2 and SETDB1 triple mutant

Suv39h1 and Suv39h2 loci were targeted using CRISPR-Cas9 technology via zygote injection as previously described (35). Specifically, since absence of either of the two HMTases can be compensated by the other (13), we decided to obtain a conditional mutant for Suv39h1 and a germ-line KO for Suv39h2. Sequences of the oligonucleotides used for inserting loxP sites in Suv39h1 intron 2 and 5, or ind/del on Suv39h2 exon2 are indicated in Table S5. Zygote injections were performed at the Transgenic and Chimeric Mouse Facility (Dept. Genetics) at University of Pennsylvania. Genotyping was performed using the primers indicated on Table S5 as indicated in (36). For experiments where triple knockout (TKO) mutants were considered, controls were FoxA3-Cre negative; Setdb1 floxed/floxed; Suv39h1 floxed/floxed; Suv39h2 del/del.

RT-qPCR sample preparation

Total RNA was isolated using RNeasy Micro Kit (Qiagen, ID:74004) following manufacturers' instruction. RNA was reverse transcribed using Biorad iScript cDNA synthesis kit (bio-Rad #1708891). cDNA was pre-amplified with Taqman probes diluted 100-fold. Preamplified cDNA was then subjected to qPCR using specific Taqman probes. Delta-delta Ct analysis was performed. Taqman probes and primers used in this study are listed in Table S1.

Sample preparation for chromatin-related assays

We provide here a detailed protocol describing how to prepare low cell number sample for chromatin-related experiments.

1. Notes

The following protocol is modified from O'Geen, et al. (37) and recommended for small cell number ChIP-seq and is based on successful 3×10^4 cells for ChIP-seq with H3K9me3 and H3K27me3 antibodies in mouse sorted cells isolated by FACS with cell surface markers at different developmental stages (i.e. e8.25, e10.5). Please note for optimal DNA recovery when purifying ChIP, use a Phenol Chloroform method. This protocol was chosen due to its high specificity and low background.

2. Buffers

Cell Lysis Buffer (make 10 ml fresh each time)

3 mM MgCl₂

10 mM NaCl

10 mM Tris, pH 7.4

0.1% Igepal (Sigma Aldrich, cat #I8896)

Add protease inhibitors (1 tablet Complete Mini Protease inhibitor cocktail tablets - Roche, #04693159001-, per 10 ml Buffer).

Nuclei Lysis Buffer (prepare it fresh each time)

50 mM Tris-Cl pH 8.0

10 mM EDTA

1% SDS (Lonza, #51213)

Before use, add protease inhibitors (1 tablet Complete Mini Protease inhibitor cocktail tablets - Roche, #04693159001- per 10 ml Buffer. Never store at 4C).

IP Dilution Buffer (RIPA)

50 mM Tris pH 7.4

150 mM NaCl

1% Igepal (Sigma Aldrich, cat #I8896)

0.25% Sodium Deoxycholate (Sigma Aldrich, cat #D6750)

1mM EDTA

Add protease inhibitors (1 tablet Complete Mini Protease inhibitor cocktail tablets - Roche, #04693159001- per 10 ml Buffer).

IP Wash Buffer 1

(RIPA without protease inhibitors; place it on ice after preparation)

50 mM Tris pH 7.4

150 mM NaCl

1% Igepal (Sigma Aldrich, cat #I8896)

0.25% Deoxycholic Acid (Sigma Aldrich, cat #D6750)

1mM EDTA

IP Wash Buffer 2

(place it on ice after preparation)

100 mM Tris-Cl pH 9.0

500 mM LiCl
1% Igepal (Sigma Aldrich, cat #I8896)
1% Deoxycholic acid (Sigma Aldrich, cat #D6750)

IP Wash Buffer 3 (IP Wash Buffer 2 plus NaCl; place it on ice after preparation)

100 mM Tris-Cl pH 9.0
500 mM LiCl
1% Igepal (Sigma Aldrich, cat #I8896)
1% Deoxycholic acid (Sigma Aldrich, cat #D6750)
150 mM NaCl

Elution Buffer

50 mM NaHCO₃
1% SDS (Lonza, #51213)

3. Cell crosslinking:

The following has been optimized for crosslinking sorted cells:

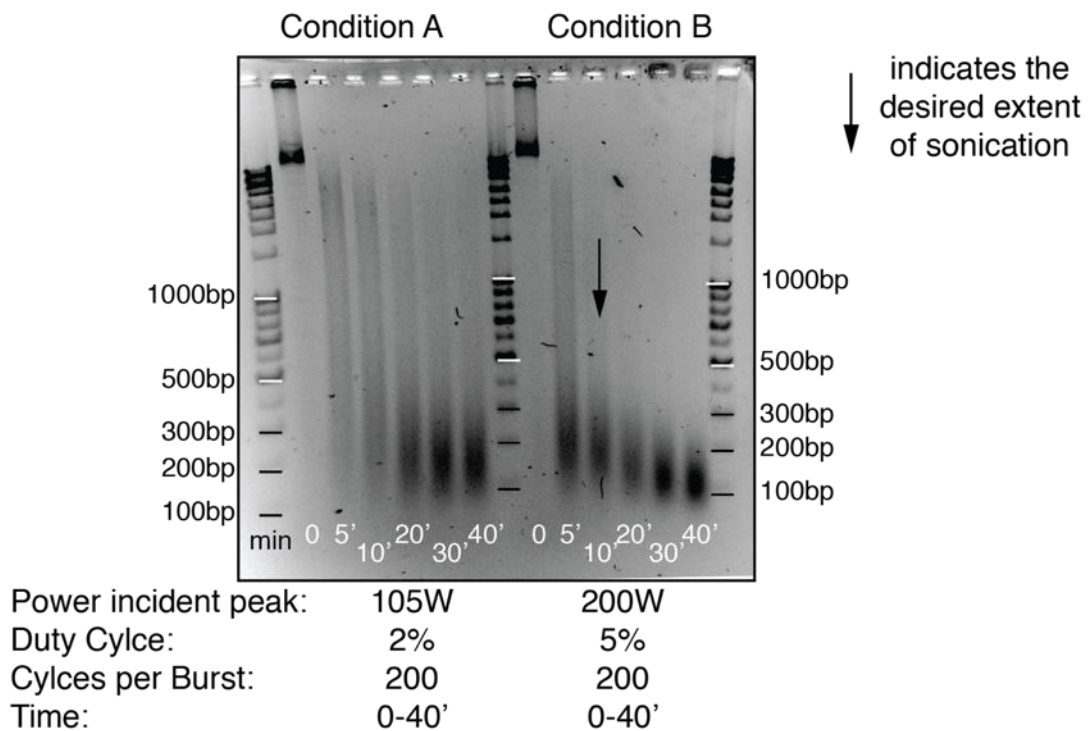
- a. Use FACs to sort $> 10^4$ cells into 1.5 ml tube, containing 50 ul 1X PBS supplemented with 1% BSA (Sigma A9418-50G) and 20% FBS (Hyclone, SH30071). Spin down the cells at 1700 rpm for 12-15 min at 4°C. Resuspend the pellet in 200-300ul PBS freshly prepared. Prepare 3 x 10^4 cells aliquots in 500ul PBS or alternatively, if you have a lot of cells, prepare 1-3 x 10^5 cells aliquots.
- b. Add 13.5 ul of 37% formaldehyde (Fisher Chemical F79-500) per 500 ul sample. Incubate exactly 10 min at room temp. on rotating platform.
- c. Stop crosslinking reaction by adding 36 ul of 2M glycine (Fisher BioReagents BP381-1) to the sample. Incubate 5 min at room temp. on rotating platform.
- d. Centrifuge at 2000 rpm for 15 min at 4°C
- e. Wash cells twice in cold PBS supplemented with 2% FBS (Hyclone, SH30071). For each wash, centrifuge at 2000 RPM for 5 min. Slowly aspirate supernatant with pipetman, leaving approximately 10 ul of solution behind. Be careful to not disturb the pellet.

4. Preparation of chromatin from crosslinked cells:

- a. If you have $> 10^5$ cells, resuspend cells in 300 ul freshly prepared cell lysis buffer and incubate on ice for 15 min. Centrifuge at 5000 rpm for 5 min at 4°C. Repeat centrifugation if cell lysate is not visible. Remove supernatant with pipetman, leaving pellet. Leave approximately 10 ul of supernatant. If starting with $< 10^5$ cells, proceed to step 2.
- b. Prepare aliquots of 3 x 10^4 cells in 130 ul Nuclei Lysis Buffer.
- c. Incubate on ice for 15 min, flash freeze (use dry ice or liquid nitrogen), then thaw to room temperature to aid in lysis. (You may keep chromatin at -80°C until a future day and then thaw to room temperature).

- d. Sonicate with predetermined conditions so that chromatin is between 200 and 500 bp in length. If using Covaris (S220 Focused Ultrasonicator) for chromatin sonication apply the following parameters (identified after time course experiment – see panel below):
 - i. Incident peak power: 200
 - ii. Duty Cycle: 5%
 - iii. Burst/Cycle: 200
 - iv. Time: 10-12 min
- e. Centrifuge sonicated chromatin for 10 min at 14K rpm at 4°C, collect supernatant and either use immediately in ChIP experiment or freeze at -80°C.

Sample sonication - time course



Panel A: COVARIS sonication time course on 30,000 cell aliquots using two different set of parameters. Condition B (12 min) is the one used in this paper.

5. Checking sonicated chromatin:

- a. Collect about 3×10^3 cells and resuspend them in elution buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 1% SDS - Lonza, #51213)
- b. Reverse crosslink at 65°C o/n on heat block
- c. Add 200 ul TE buffer (to dilute SDS - Lonza, #51213)
- d. Add 8 ul of 10 mg/ml RNase A (Roche, #10109169001) and incubate samples for 1.5-2 h at 37°C

- e. Add 4ul Proteinase K (20mg/ml stock) (Roche, #03115828001) and incubate 1.5-2 h at 55°C
- f. Precipitate DNA with standard phenol-chloroform protocol
- g. Transfer the aqueous phase to a new tube
- h. Add 16 ul of 5 M NaCl, 1.5 ul of 20 mg/ml glycogen (Roche, #10901393001) and 800 ul 100% EtOH
- i. Place the samples at -80°C for a couple of hours (you can do this step o/n)
- j. Pellet DNA spinning the samples at 15,000 rpm for 12 min at 4°C (pellet is faint!)
- k. Wash with 500 ul 80% EtOH, once
- l. Pellet DNA spinning the samples at 15,000 rpm for 12 min at 4°C
- m. Air dry the pellet and resuspend in 15 ul TE buffer (remove EtOH)
- n. Spin vacuum the samples and resuspend them in 3ul H₂O containing 10% glycerol
- o. Prepare 1% agarose gel (as thin as possible) without Ethidium Bromide
- p. Load the samples and marker
- q. Run the gel at 90V for 40-45 min
- r. Stain the gel (10 ug/ml final concentration) for 15 min RT while shaking
- s. Wash the gel 5 min RT while shaking in ddH₂O
- t. Image the gel.
- u. Note: the sonication can be checked with the Bioanalyzer (Agilent Technologies G2939AA) by resuspending the samples from step 14 in water (3 ul).

Chromatin Immunoprecipitation (ChIP) on low cell number samples:

We provide here a detailed protocol describing how to perform chromatin immunoprecipitation and how to obtain samples ready to be used for libraries preparation starting from low cell number samples.

6. Immunoprecipitation

- a. Transfer sonicated chromatin into 0.6 ml Axygen Maximum Recovery microtubes (MCT-060-L-C, #311-03-051). Dilute chromatin with Dilution Buffer to a final volume of 500 ul. For input, take 5 ul (1%) from the 500ul containing sheared chromatin and transfer to new tube containing 100 ul elution buffer.
- b. Add 1 ug of primary ChIP-grade antibody and incubate overnight on a rotating platform at 4°C.
- c. Add 15 ul Cell Signaling ChIP-grade Protein G magnetic Beads (Invitrogen #10004D) and incubate 2 h on rotating platform at 4°C.
- d. Transfer chromatin and beads to a pre-chilled 1.5 ml microtubes and allow beads to separate in magnetic rack (Thermo Fisher Scientific, #12321D) for > 30 se.
- e. Remove supernatant with pipetman. Using suction may cause you to lose beads and sample.

- f. Remove tube from rack, resuspend beads in IP Wash Buffer 1, and wash by pipetting up and down 12 times.
- g. Place tubes in magnetic rack (Thermo Fisher Scientific, #12321D) and allow beads to separate for > 30 sec.
- h. Repeat steps 5-7 once.
- i. Remove supernatant with pipetman.
- j. Remove tubes from rack, resuspend beads in IP Wash Buffer 2, and wash by pipetting up and down 12 times.
- k. Place tubes in magnetic rack (Thermo Fisher Scientific, #12321D) and allow beads to separate for > 30 sec.
- l. Repeat steps 9-11 2 times.
- m. Remove supernatant with pipetman.
- n. Remove tubes from rack, resuspend beads in IP Wash Buffer 3, and wash by pipetting up and down 12 times.
- o. Transfer beads and chromatin to a new 1.5 ml microtube.
- p. Place in magnetic rack (Thermo Fisher Scientific, #12321D) and allow beads to separate for > 30 seconds.
- q. Remove tubes from rack and resuspend in 100 ul Elution Buffer.
- r. Shake at RT at a setting of 1300 rpm for 1 h.
- s. Place tubes in magnetic rack (Thermo Fisher Scientific, #12321D) and allow beads to separate for > 30 seconds.
- t. Collect supernatant and add 12 ul 5M NaCl.
- u. Thaw the input and add 12ul 5M NaCl
- v. Incubate 4 hours to overnight at 65°C.
- w. Add 3 ul of 10 mg/ml RNase A (Roche, #10109169001) to each sample and incubate at 37°C for 30 min.

7. Phenol Chloroform sample purification

- a. Spin down Phase Lock Gel Light 1.5 ml tube (5 Prime, QuantaBio, #2302820) at 14000 rpm for 30 sec.
- b. Add 100 ul TE buffer to each samples (including input) and mix gently (total volume 200 ul)
- c. Mix 200 ul ChIP product and 200 ul Phenol/Chloroform/Isoamyl Alcohol by pipetting about 10 times and add to Phase Lock Gel Light 1.5 ml tube. Do not vortex.
- d. Centrifuge at room temperature for 5 minutes at 14000 rpm.
- e. Optional: Add an additional 200 ul of Phenol/Chloroform/Isoamyl Alcohol and mix by pipetting.
- f. Centrifuge at room temperature for 5 min at 14000 rpm.
- g. Collect aqueous solution on top and add to a new tube (should be about 200 ul). If the volume is less than 200 ul, add Buffer TE to reach 200 ul.
- h. Add 8 ul 5M NaCl, 10 ug glycogen, and 400 ul cold 100% ethanol, and mix.
- i. Precipitate DNA in ethanol at -80°C overnight at least for best recovery.

- j. Centrifuge at 4°C for 10 min at 11000 rpm.
- k. Aspirate supernatant carefully, without disturbing the very small white pellet. To be safe, you may leave 5 ul at the bottom.
- l. Wash with 500 ul 80% ethanol by pipetting five times.
- m. Centrifuge at 4°C for 10 min at 11000 rpm.
- n. Aspirate supernatant carefully, without disturbing the very small white pellet.
- o. Air dry 5 min, or until ethanol has evaporated, but pellet isn't completely dry. If pellet is too dry, it will be difficult to resuspend.
- p. Resuspend pellet in 30-40 ul Qiagen EB. It is now ready for qPCR. For library preparation resuspend the pellet in less volume (e.g. 10 ul).

srHC isolation and sample preparation for sequencing:

We provide a detailed protocol describing how to perform isolation of srHC and how to obtain samples ready to be used for libraries preparation starting from low cell number samples. Related publications at (38–40)

1. Reagents

Agencourt® AMPure® XP (Beckman Coulter, cat. no. A63881)

80% EtOH

Buffer EB

10 mM Tris-Cl, pH 8.5)

Glycine (2M solution) (Fisher BioReagents BP381-1)

nuclear lysis buffer

50mM Tris-Cl pH 8.0

10mM EDTA

1% SDS - Lonza, #51213, supplemented with protease inhibitors

TE buffer

10 mM Tris-Cl, pH 8.0

1 mM EDTA

RNase A (Roche, #10109169001)

Glycogen (Roche, #10901393001)

2. Equipment

COVARIS S220 (Focused Ultrasonicator S220)

microTUBE AFA Fiber Pre-Slit Snap-Cap (COVARIS, cat. #520045)

DynaMag™-2 Magnet (Thermo Fisher Scientific, #12321D)

Agilent 2100 Bioanalyzer system (Agilent Technologies, G2939AA)

Eppendorf Tubes

3. Cell preparation

- a. Prepare thirty thousand (30,000) cell aliquots in 500 ul PBS from sorted cells.

- b. Add 13.5 ul of 37% formaldehyde (Fisher Chemical F79-500) per 500 ul sample.
- c. Incubate samples 10 min at RT on rotating platform.
- d. Stop crosslinking reaction by adding 36 ul of 2M glycine (Fisher BioReagents BP381-1) to each sample.
- e. Incubate samples for 5 min incubation at RT on rotating platform.
- f. Centrifuge samples at 2000 RPM for 15 min at 4°C.
- g. Wash cells twice in cold PBS (use 10 ul of FBS - Hyclone, SH30071, in 500 ul of PBS). For each wash, centrifuge samples at 2000 rpm for 5 min at 4°C. Slowly aspirate supernatant with pipetman, leaving approximately 10 ul of solution behind.
- h. Resuspend the cells in 130 ul nuclear lysis buffer (50mM Tris-Cl pH 8.0, 10mM EDTA, 1% SDS - Lonza, #51213, supplemented with protease inhibitors - 1 tablet Complete Mini Protease inhibitor cocktail tablets - Roche, #04693159001- per 10 ml Buffer).

4. Chromatin Sonication

- a. Sonicate chromatin using COVARIS S220 following these parameters:
 - i. Peak Power 200
 - ii. Duty Cycle 5%
 - iii. Burst/Cycle 200
 - iv. Time 10'

5. DNA purification

- a. Adjust sample volume to 200 ul with nuclear lysis buffer.
- b. Reverse crosslinking by incubating samples at 65°C for 18h on heat block.
- c. Add 200 ul TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA), 8 ul of 10 mg/ml RNase A (Roche, #10109169001) to the samples.
- d. Incubate the samples for 2 h at 37°C.
- e. Add 4 ul, 20 mg/ml Proteinase K (Roche, #03115828001) to the samples.
- f. Incubate the samples for 2 h at 55°C.
- g. Purify DNA using conventional phenol:chloroform:isoamyl extraction. Briefly, add 400 ul phenol:chloroform:isoamyl mix to each sample. Mix and quickly vortex the samples.
- h. Spin down the samples at max speed for 5 min at RT.
- i. Transfer 350 ul of aqueous phase to a new eppendorf
- j. Add 16 ul 5M NaCl, 1.5 ul 20mg/ml glycogen (Roche, #10901393001) and 800 ul 100% EtOH to each sample.
- k. Mix the samples thoroughly by inverting the tube 10-15 times.
- l. Incubate the samples are incubated overnight at -80°C.
- m. Centrifuge the samples for 10 min at max speed at 4°C.
- n. Wash the samples with 500 ul 80% EtOH.
- o. Centrifuge the samples for 10 min at max speed at 4°C.

- p. Air-dry sample pellet at RT 5 min and resuspend it in 21 ul EB buffer.
- q. Save 1 ul purified DNA for checking sonication at the Bioanalyzer (Agilent Technologies, G2939AA) in parallel with further processed samples.

6. Sonication-resistant chromatin selection

- a. Add 0.7X AMPure XP beads (Beckman Coulter, A63881) to the each sample (14ul in 20ul initial solution). Please note that the amount of AMPure XP beads indicated has been optimized for 3×10^3 cell samples. Samples of different cell number might require adjustment in AMPure XP beads:volume ratio.
- b. Mix samples thoroughly by pipetting 10-12 times.
- c. Incubate samples at RT for 5 min.
- d. Transfer eppendorf tubes to a magnetic rack (Thermo Fisher Scientific, #12321D) and beads are allowed to separate for 5 min at RT.
- e. Transfer clear solution to a new eppendorf tube (this fraction is the small, sonication-sensitive fragments).
- f. Wash beads (to which large, sonication-resistant fragments are bound) with 200ul EtOH 80% while tubes are maintained on rack.
- g. Incubate samples for 30 sec with EtOH.
- h. Remove EtOH and wash once more each sample.
- i. Air-dry beads 5 min while tubes are kept on the rack.
- j. Remove tubes from the rack
- k. Elute the beads in 20 ul TE buffer, by pipetting thoroughly.
- l. Save 1 ul purified DNA before further sonication to check samples size at the Bioanalyzer (Agilent Technologies, G2939AA) in parallel with further processed samples.
- m. Incubate the samples are incubated 5 min at RT.
- n. Place the samples on the magnetic rack (Thermo Fisher Scientific, #12321D).
- o. Incubate the samples for 5 min.
- p. Transfer clear solution to a new tube (this fraction is the large, sonication-resistant portion of DNA).

7. Sonication of large DNA fragments

- a. Adjust sample volume to 130 ul adding TE buffer.
- b. Sonicate the samples using COVARIS S220 (Focused Ultrasonicator S220) following these parameters:
 - i. Peak Power 175
 - ii. Duty Cycle 10%
 - iii. Burst/Cycle 200
 - iv. Time 6'
- c. Adjust sample volume to 200 ul with TE buffer. DNA precipitation can be applied also to small, sonication-sensitive sample.

- d. Add 1 ul of 20 mg/ml glycogen (Roche, #10901393001), 1/10 volume of sodium acetate (3 M, pH 5.2) and 3.0X volume EtOH 100% to each sample
- e. Mix the samples thoroughly by inverting the tube 10-15 times.
- f. Incubate the samples 4 h at -80°C.
- g. Centrifuge the samples for 10 min at max speed at 4°C.
- h. Wash the samples with 500 ul 80% EtOH.
- i. Centrifuge the samples for 10 min at max speed at 4°C.
- j. Air-dry sample pellet at RT 5 min and resuspend it in 11 ul EB buffer.
- k. Save 1 ul of purified DNA for checking sonication at the Bioanalyzer (Agilent Technologies, G2939AA) in parallel with further processed samples

Sample preparation of polyA-selected RNA for sequencing:

Two replicates were used for each stage. Total RNA was isolated from 3×10^4 cell aliquots using RNeasy Micro Kit (Qiagen, ID:74004) following manufacturers' instruction. Total RNA was resuspended in 50 ul BTE solution (10 mM Bis-tris pH6.7, 1 mM EDTA). RNA secondary structures were removed by heating the samples at 65°C for 5 min, followed by incubation on ice. Oligo-dT25 dynabeads (Invitrogen #610-02) were used for selecting polyA mRNA. Beads were washed 3 times with 2X Oligo-dT binding buffer (OBB: 20 mM Tris pH7.5, 1 M LiCl, 2 mM EDTA), resuspended in 50 ul 2X OBB and added to RNA. Samples were incubated 5-10 min at RT while gently shaking. Samples were transferred to a (Thermo Fisher Scientific, #12321D) rack, and beads were allowed to separate for 3-5 min. Beads were washed 3 times with 200 ul Oligo-dT washing buffer (OWB: 10 mM Tris Ph7.5, 150 mM LiCl, 1 mM EDTA) and polyA mRNA was eluted into 10 ul BTE. Samples were incubated 2 min at 80°C and beads were allowed to separate on the magnetic rack (Thermo Fisher Scientific, #12321D). The solution was transferred to a fresh tube. Chemical RNA fragmentation was achieved using Mg^{2+} . On ice, for each 10 ul sample volume, 4 ul 5X 1st strand buffer from SuperScript III RT kit (Invitrogen #18080-044) and 2 ul 50 uM random hexamers (Roche) were added. Samples were incubated 16 min at 94°C in PCR machine and ramped down to 4°C. First strand synthesis was

performed using SuperScript III (Invitrogen #18080-044), following manufacturers' instructions and adding 1ug actinomycin D (Sigma #A1410) per reaction. Samples were incubated in PCR machine according to the following scheme: 10 min at 25°C, 20 min at 42°C, 30 min at 50°C, 10 min at 55°C. Samples were kept on ice. First strand cDNA was purified using conventional phenol:chloroform:isoamyl extraction. cDNA was precipitated using 3 M NaAc, 1 ul 20mg/ml glycogen (Roche, #10901393001) and 1 ml 100% EtOH. Samples were resuspended in milliQ water. Second strand synthesis incorporating dUTP was performed as following: for each sample, 2 ul First Strand bugger (from SuperScript III kit; Invitrogen #18080-044), 0.5 ul 0.1 M DTT, 5 ul 10X second strand buffer (dNTPs free; NEBnext #B6117S), 2 ul 10 mM dNTPs (with dUPT substituting in place of dTTP), 2 ul DNAPol I (20 U; NEB #M0209), 0.5 ul RNaseH (2.5 U: NEB #M0297) were added. Samples were incubated 2-3 h at 16°C. cDNA was purified using MiniElute kit (Qiagen, #28004) and samples were eluted in 10 ul RNase-free water.

Spike-in ChIP-Seq and srHC experiments:

H2A.v ChIP-seq data were aligned to the Drosophila genome assembly dm6/BDGP R6 using bwa version 0.7.10 with default parameters. Tags were extended to 200 bp and H2A.v peaks were called using SICER-rb (redundancy threshold 1, window size 200 bp, fragment size 200 bp, effective genome fraction 0.72, gap size 200 bp, E-value cutoff 0.1). The resulting peak regions were used to construct a bwa index and H3K9me3 and input data were aligned to this "H2A.v-ome" as well as to the standard mm9/NCBI v37 genome using bwa. Aligned tags were extended 200 bp to match the fragment size. Mouse-aligned read density was visualized similar to non-spiked-in H3K9me3, except that tracks are also scalar adjusted for the ratio of control:mutant spike-in tags as in (41). Because there is no logical pairing between mutant (n=2) and control (n=1, two technical repeats) samples, we

averaged control samples for each mutant. For srHC spike-in experiments DNA was purified from Ctrl (n=1, three technical repeats) and TKO (n=2) samples. 500 ng of purified DNA was considered to perform the chromatin fractions separation. Large and small fractions were purified and quantified Qubit 3.0 (Invitrogen) and 50 pg of Drosophila DNA was spiked-in in each sample, before dilution and libraries preparation.

Western blotting

Sample preparation, gel run and signal detection were performed as described in (25), with the following modifications: samples were denatured for 5 min at 95°C. Wet transfer was performed using in house-made transfer buffer (25 mM of Tris, 192 mM of glycine - Fisher BioReagents BP381-1, 20% methanol). Antibodies used in this study are listed on Table S3. For Thermo Scientific antibodies, the presented results have been achieved using the indicated lot numbers.

Immunohistochemistry

Embryos or tissues were fixed in 2% PFA (Electron Microscopy Sciences, #15714-S) o/n at 4°C. Samples were washed 3 times in 1X PBS and subjected to dehydration through sequential 15 min washes at increasing concentration of methanol in 1X PBS (25% - 50% - 75% -100%). Samples were washed twice with 100% EtOH, and stored at -20°C until further processed. Samples were equilibrated in 70% EtOH 1X PBS before paraffin embedding and sectioning. Sections were deparaffinized through two sequential 5 min washes in Xylenes (Fisher Chemical, X3P-1GAL), and reducing concentration of EtOH (100%, two washes; 95% EtOH, 75% EtOH, 50% EtOH, one wash). Samples were rinsed with water, immersed in sodium citrate buffer (10 mM Na-Citrate, pH6.0) supplemented with 0.05% Tween20 (Sigma Aldrich, cat #P1379) and microwaved for 2 min at Power Level 10, followed by 13 min microwaving incubation at Power Level 2. Samples were

cooled for 20 min at RT and washed 2-3 times in milliQ water for 5 min. Samples were washed 3 times in 1X PBS and endogenous peroxidases were quenched immersing the samples in 3% H₂O₂ solution (Sigma H1009-500ml) for 30 min at RT. Samples were washed 3 times in 1X PBS and blocking buffer (Thermo Fisher, Pierce Protein-Free T20 PBS blocking buffer, #37573) for 2-3h at RT. Primary antibodies were diluted in blocking buffer and samples were incubated o/n at 4°C. Samples were washed 3 times for 5 min with 1X PBS supplemented with 0.1% Tween20 (Sigma Aldrich, cat #P1379) (PBS-T). HRP-conjugated secondary antibodies were diluted in 1X PBS-T and samples were incubated 1h at 37°C. Samples were washed 3 times for 5 min with 1X PBS-T. DAB peroxidase staining (Vector, SK-4100) was performed following manufacturers' instructions. Reaction was stopped by immersing the samples in milliQ water. Samples were counterstained with Eosin (Millipore, #109844.1000) and dehydrated through sequential 2 min washes in 95% and 100% EtOH, followed by two 2 min washes in Xylene. Permount (Fisher, #SP15-500) was used as mounting solution. For Thermo Scientific Suv39h2 antibody, the presented results have been achieved using the indicated lot number. Antibodies used in this study are listed on Table S3.

Library preparation and next-generation sequencing:

Libraries were generated using ThruPLEX DNA-seq 48S Kit (Rubicon Genomics, #R400427). Libraries were quantified by qPCR using the NEBNext Library quantification kit for Illumina (NEB, #E7630L). Libraries were diluted to 8 nM and pooled for multiplexing. The diluted concentrations and were quantified a second time using NEBNext kit. Libraries were denatured in 0.2 M NaOH and loaded into NextSeq 75-cycles High Output v2 Kit (Illumina, FC-404-2005) cartridge at a concentration of 2.5-3 pM. Sequencing runs were performed in an Illumina NextSeq 500 machine.

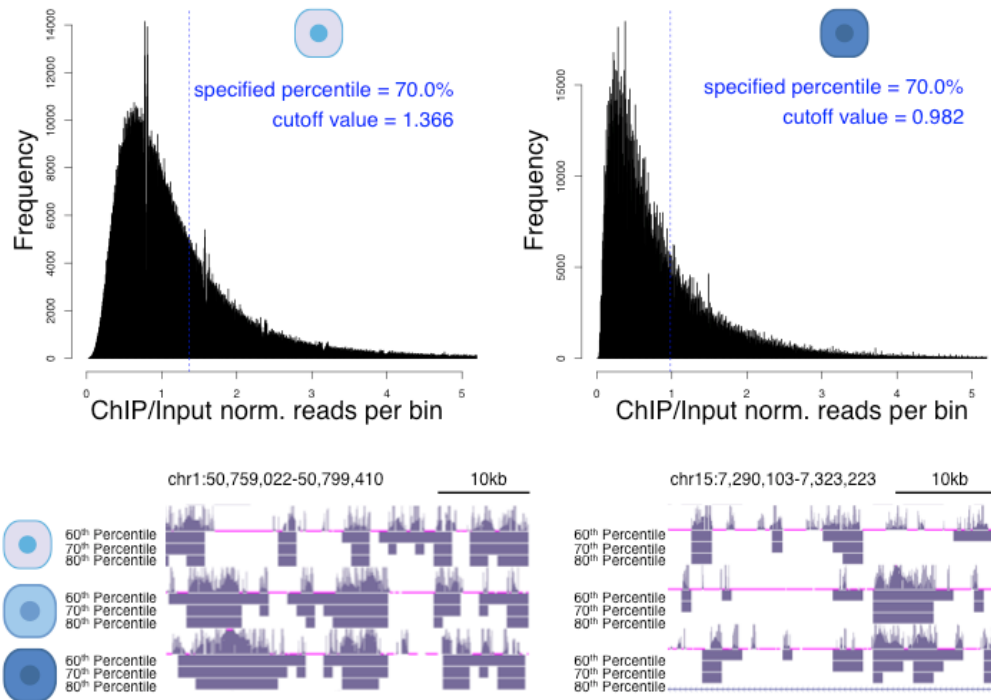
Bioinformatics

H3K9me3 ChIP-seq data for P0 heart and nucleus accumbens were downloaded from GSM2192520 and GSE24850, respectively. ChIP-seq data were aligned to mouse genome assembly mm9/NCBI v37 using bowtie2, version 2.1.0, using the --very-sensitive parameter setting. Read density was visualized using the UCSC Genome Browser to display bigWigs created with BEDTools genomeCoverageBed and UCSC Genome Browser's bedGraphToBigWig utility with a normalization step that adjusts each track to the library size (RPM) and which takes the fold-change of H3K9me3/Input. RNA-seq data were aligned using STAR version 2.5.2 and default parameters, then processed to get read counts per gene with HTSeq version 0.6.1. Resulting counts were analyzed for significant differential expression using DESeq (standard binomial test with an FDR control of 0.1). Spearman correlations were calculated as described in (25) using a 2 kb sliding window algorithm, with a sliding step of 1 kb for H3K9me3 and H3K27me3 and 10 kb sliding window algorithm, with a sliding step of 1 kb for srHC.

Calling enriched genomic domains

Enriched genomic domains (i.e. patches) were called as described in (25), with the following modifications. For H3K9me3 and H3K27me3 we used 2 kb sliding window algorithm, with a sliding step of 1 kb. For srHC we used 10 kb sliding window algorithm, with a sliding step of 1 kb. Fixed percentile enrichment thresholds were applied (to 30% for H3K9me3, top 20% for H3K27me3 and top 40% for srHC. A representative example of patch calling is indicated on panel B). For spike-in enriched genomic domains calling, the threshold value identified upon applied the fixed percentile in the control samples, was considered for setting the mutant threshold. RefSeq genes were considered decorated (i.e. patched) if at least 25% of the gene body was covered by called domain(s). For

H3K27me3 compensatory effect over H3K9me3, genes not changing expression across the indicated time-course (expression ratio of definitive endoderm to hepatocytes or mature beta cells < 1.1X) and going from being H3K9me3 patched (> 50% of the gene body overlapping the domain class) to unpatched (< 25% of the gene body overlapping the domain class) were considered. The area under the curve (AUCs) was computed for H3K27me3 and H3K9me3 over those genes. For each experiment, biological replicates have been pooled together. Downstream analysis has been performed on pooled replicates.

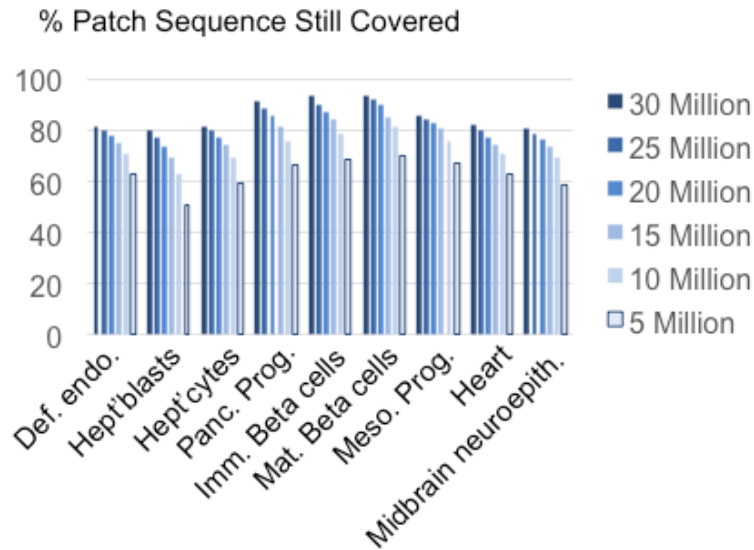


Panel B: Representative example of H3K9me3 patch calling for three indicated cell types. Top row shows histograms of Chip over input normalized reads per bin (2kb) and the cutoff threshold considered to call patches in the top 30% values. Bottom row shows two representative genome track examples showing three different thresholds considered to call patches.

Random down-sampling and data confirmation

Difference in the number of aligned reads might lead to artifact upon normalization. To assess the effect of genomic coverage on H3K9me3 patch definitions, we randomly down-

sampled the H3K9me3 and input tag counts (after pooling replicates) to one of six cutoffs (5M-10M-15M-20M-25M-30M). If the number of tags for a given sample was already under the cutoff, it was left as-is. Patches were then re-called using the pipeline used in the manuscript. The chart below shows the result of our down-sampling analysis.



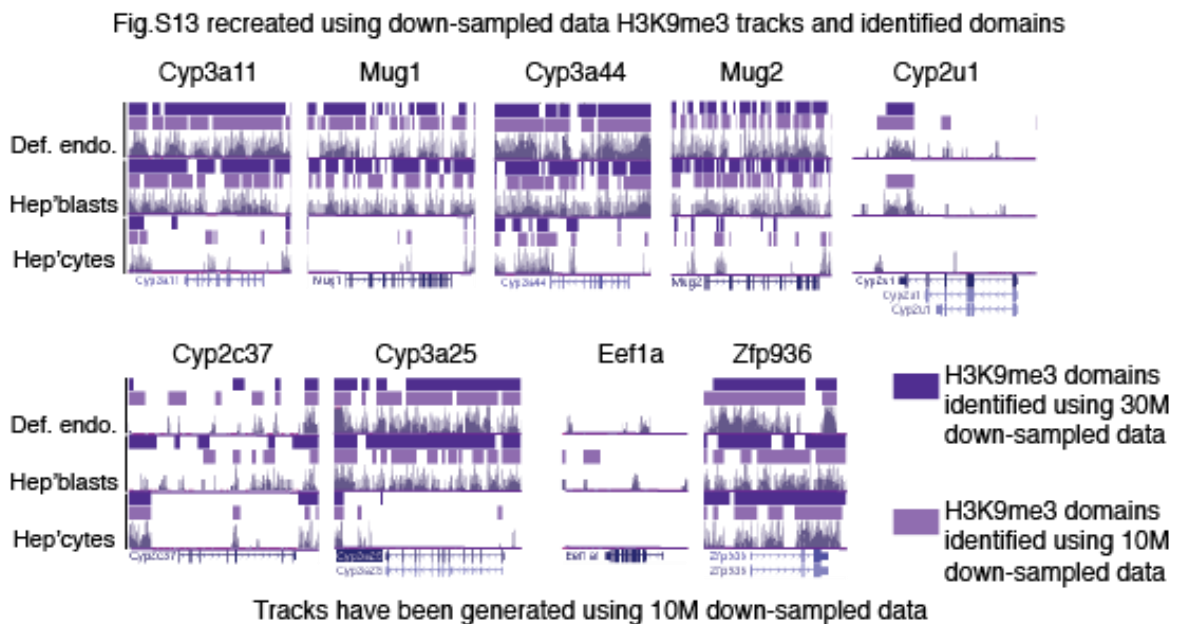
Panel C: Percentage of original H3K9me3 domain sequence covered upon down-sampling.

After patch calling, we counted the number of base-pair positions in the mouse genome covered by the original patches, then calculated the % bases still covered after down-sampling. We found that even at the limit of 5 million tags, the percentage of patch sequence called does not drop below half the original called, and even at the 10 million tag limit we are still calling between 70-80% of the original sequence for all samples but hepatoblasts. The table below shows the number of genes marked by H3K9me3 in the indicated stages after processing random down-sampled from initial reads to 10M and 30M. No matter how many reads were considered, the number of genes marked by H3K9me3 is highest in definitive endoderm and mesoderm progenitors and lower in the other developmental stages, confirming the data presented in this manuscript.

	down-sampling	
	10M	30M
def. endo.	10,648	9,161
hepatoblasts	8,138	6,407
hepatocytes	8,414	7,147
panc. prec.	6,589	5,538
imm. beta cells	5,555	4,963
mat. beta cells	5,138	4,502
meso. prog.	10,289	10,064
midbrain	5,770	5,205

Panel D: Number of genes marked by H3K9me3 upon down-sampling in the indicated samples

We also regenerated fig. S13 using randomly down-sampled data. Down-sampling of the data at 30M and 10M recapitulate accurately the figure originally presented in the manuscript.



Panel E: fig. S13 recreated using randomly down-sampled data. Tracks have been generated using 10M down-sampled data; H3K9me3 domains have been identified using 10M and 30M down-sampled data.

Alluvial Plot generation

Patches of preferential heterochromatin were divided into 1 kb windows, and all windows in at least one patch in any of definitive endoderm, hepatocytes, or mature beta cells were

scored for patch membership in each stage. A histogram of all windows was made and a table of window frequencies (e.g., patched in Def. Endo., unpatched in hepatocytes, patched in mature beta cells maps to N windows) was passed as input to the alluvial library in R. Alluvia passing through a patched state in definitive endoderm were colored dark gray, and those passing through an unpatched state in definitive endoderm were colored light gray.

Assessing overlap between srHC and Hi-C data

A-B compartment data from series GSE93431 were downloaded from NCBI GEO. Briefly, dominant eigenvector assignments from the untreated 20 kb cis contact map were downloaded and positive scoring, contiguous regions were treated as A compartments and negative scoring, contiguous regions were treated as B compartments. To assess the significance of overlap between A or B compartments and large srHC-seq domains, a circular shuffling approach was employed. First, the number of srHC-seq domains with at least 50% overlap by nucleotide with one or more compartments of a specified type was calculated. Next, a background overlap distribution was estimated by circular shuffling the SRHC-domains 1,000 times, in each iteration testing whether the overlap at 50% was greater than or equal to the observed overlap. The p-value was estimated at the number of iterations over 1,000 in which the shuffled overlap met or exceeded the observed overlap.

RNA-Seq data clustering analysis

From the DESeq value adjusted table, non-expressed genes and genes expressed in 1 single developmental stage were removed (total discarded gene = 6,665). Unique values were considered by removing duplicates. The final total number of genes considered for the clustering analysis was 17,570. Automated fuzzy c-means clustering was obtained

using the following software (<http://computproteomics.bmb.sdu.dk>) (42). Gene ontology analysis of individual clusters (Table S14) was performed using David (<https://david.ncifcrf.gov>) (43, 44) or GOrilla (<http://cbl-gorilla.cs.technion.ac.il>) (45, 46).

Single cell RNA-Seq sample preparation and data analysis

E11.5 Liv2+ sorted cells from wt (n=7) and FoxA3-Cre;SETDB1^{fl/fl} (n=3) embryos were fixed with methanol and stored at -80°C (47). For rehydration, cells were moved from -80°C to 4°C and kept on ice throughout the procedure. Cells were pelleted at 3000 x g, resuspended in PBS + 0.01% BSA (Sigma A9418-50G), centrifuged again, resuspended in PBS + 0.01% BSA (Sigma A9418-50G), passed through a 40 µm cell strainer, counted, and diluted for Drop-seq in PBS + 0.01% BSA (Sigma A9418-50G). To assess the single-cell resolution of the procedure we spiked in 10% of 293T cells in the mouse embryo single cell suspension. The final concentration of cells was 100 cells/ul and approximately 1.5 ml of cell suspension was loaded for each Drop-seq run. Barcoded beads were resuspended in lysis buffer (200 mM Tris-HCl pH 8.0_{RT}, 20 mM EDTA, 6% Ficoll PM-400 - GE Healthcare/Fisher Scientific -, 0.2% Sarkosyl - Sigma-Aldrich -, and 50 mM DTT - Fermentas; freshly made on the day of run) at a concentration of 120 beads/ml. The flow rate for cells and beads were set to 4,000 ml/h, while droplet generation oil (Bio-rad) was run at 15,000 ml/h. The droplets were then generated and collected in a 50 ml falcon tube for a run time of 15 min. Droplet breakage with Perfluoro-1-octanol (Sigma-Aldrich), reverse transcription, exonuclease I treatment, and amplification of cDNA were performed, following the Drop-seq protocol of the McCarroll lab (<http://mccarrollab.com/dropseq/>) with minor modifications. After two rounds of purification with 0.6x SPRISelect beads, we tagged 600 pg cDNA using the Nextera XT DNA sample preparation kit (Illumina, cat# FC-131-1096). Following cDNA tagmentation, we further amplified the libraries with 12

PCR cycles using custom P5-TSO hybrid and custom Nextera-compatible RB70X primers with different indexes. After quality control analysis by Qubit 3.0 (Invitrogen) and a Bioanalyzer (Agilent), libraries were sequenced on an Illumina Nextseq500 instrument using the 75-cycles High Output v2 Kit (Illumina, FC-404-2005). We loaded the library at 1.8 pM and used custom read 1 and index 2 primers. The sequencing configuration was 20 bp (Read1), 8 bp (Index1), 8 bp (Index2) and 56 bp (Read2). Primers used for sc-RNA-seq libraries preparation are listed in Table S4.

Upon sequencing the UMI (Unique Molecular Identifiers) count matrices of cells were extracted following the Drop-seq Core Computational Protocol from Steve McCarroll's lab at Harvard Medical School (See the details <http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf>). As comparing wt cells and setdb1 mutant cells, UMIs of 14,154 genes across all cells were obtained. We disregarded the cells with the total UMIs of those 14,154 genes less than 2,000. We also excluded out the outliers of cells with high proportions of UMIs of genes in the mitochondrial genome using MAD-based definition of outliers. After filtering, our data set included 1,919 cells including 952 wild type cells and 967 mutant cells. Data were normalized to the size of the library of each cell and then the normalized data were obtained by taking their logarithm with the base of 2, which were the input of the following analysis. We calculated the PCA of the data matrix and determined that 4 was the optimal number of optimal principle components since it was the elbow point of the cumulative of the percentage of the explained variance (Figure S17A). Next, we computed the two components of t-SNE using the first four principle components of PCA of the normalized data. The optimal number of clusters of cells is 3, which was determined from the different results obtained by varying all combinations of number of clusters and clustering methods under Euclidean distance (Figure S17B). There are 1,260 cells in the cluster 1 (the light

green shadow), 500 cells in cluster 2 (in the light orange shadow), and 159 cells in cluster 3 (the light magenta shadow) (Figure S17C). Then we identified the cell cycle stages of cells using the expression levels of cell cycle genes implemented by R package named “scater”. Fig. S17D shows that the cells in the same cell cycle states were distributed in the different clusters and it means that the cell cycle has no effects on the clustering. Based on the clustering information, we performed the differential expression analysis using the non-parametric Kruskal-Wallis test. A significant pvalue means that the gene expression levels in at least one cluster dominate the ones in one other cluster stochastically. The DE genes were selected with the threshold of the pvalue 0.1. Next, we calculated the fold changes of the DE genes between the wild type cells and the mutant cells. The up-regulated DE genes were chosen with the fold change greater than 1 and the down-regulated ED genes were determined by the fold changes less than 1. To show the expression patterns of DE genes in the three clusters, we use blue squares to denote the wild type cells and red triangles to denote the mutant cells. The marker genes in each cluster are the ones only expressing significantly in the corresponding clusters. Based on the marker gene list GO analysis was performed for each cluster and the representative GO analysis results are shown in Figure S17E. All cells in the cluster 2 with the expression of Alb greater than 0 (n= 1,257 single cells) were selected as Alb+ cells in the relevant analysis. tSNE plots were generated based on the first two principle components for the Alb+ cells (Figure 3B) and p-Value was calculated based on the non-parametric Kruskal-Wallis test comparing the gene expression in the wt Alb+ cells (n= 732) and the one in the mut Alb+ cells (n= 525). To compare wt cells, setdb1 mutant cells, and TKO cells with the total UMIs less than 1500 were disregarded and 1,323 wt cells, 1,366 setdb1 mutant cells, and 598 TKO cells were left to do the following analysis. Then 4 clusters were identified based on the first two components of tSNE, which were shown as four shadows with

different colors (Figure 4A). Then all cells with expression levels of Alb greater than 0 were chosen as Alb+ cells in the following analysis. Three clusters represented by three different shadows with different colors were classified based from the second component and the third component of tSNE (Figure 20D). Differentially expressed genes were detected by selecting expression with $p\text{val} < 0.01$. Upregulated genes between samples were identified considering $\log_2(\text{fold change}) > 0.5$, while downregulated genes were identified by selecting $\log_2(\text{fold change}) < -0.5$.

Statistics

A two-tailed t-test has been used for fig. S13. The Wilcoxon rank sum test was used for figure S11A and S15A. Spearman correlation coefficient was computed between all replicates and shown in figure S5E, S6H, S10A, S21B to D. Unless otherwise indicated, error bars indicate standard error mean (SEM).

3.8. Tables

Table S1. ThermoFisher Taqman probes and primers used in this study

Probe	Source	Identifier	Catalog #
Taqman mouse FoxA2 probe and primers	Thermo	Mm01976556_s1	4331182
Taqman mouse PYY probe and primers	Thermo	Mm00520716_g1	4331182
Taqman mouse Gata4 probe and primers	Thermo	Mm00484689_m1	4331182
Taqman mouse Afp probe and primers	Thermo	Mm00431715_m1	4331182
Taqman mouse Sox17 probe and primers	Thermo	Mm00488363_m1	4331182
Taqman mouse Cdx2 probe and primers	Thermo	Mm01212280_m1	4331182
Taqman mouse Sox2 probe and primers	Thermo	Mm03053810_s1	4331182
Taqman mouse Alb probe and primers	Thermo	Mm00802090_m1	4331182
Taqman mouse Ttr probe and primers	Thermo	Mm00443267_m1	4331182
Taqman mouse Pdx1 probe and primers	Thermo	Mm00435565_m1	4331182
Taqman mouse Ptf1a probe and primers	Thermo	Mm00479622_m1	4331182
Taqman mouse Nkx6.1 probe and primers	Thermo	Mm00454961_m1	4331182
Taqman mouse Onecut1 probe and primers	Thermo	Mm00839394_m1	4331182
Taqman mouse Sox9 probe and primers	Thermo	Mm00448840_m1	4331182
Taqman mouse Amy1 probe and primers	Thermo	Mm00651524_m1	4331182
Taqman mouse Cpa1 probe and primers	Thermo	Mm00465942_m1	4331182
Taqman mouse Ins1 probe and primers	Thermo	Mm01950294_s1	4331182
Taqman mouse Ins2 probe and primers	Thermo	Mm00731595_gH	4331182
Taqman mouse MafA probe and primers	Thermo	Mm00845206_s1	4331182
Taqman mouse Gcg probe and primers	Thermo	Mm00801714_m1	4331182
Taqman mouse Arx probe and primers	Thermo	Mm00545903_m1	4331182
Taqman mouse Kr19 probe and primers	Thermo	Mm00492980_m1	4331182
Taqman mouse Pax6 probe and primers	Thermo	Mm00443081_m1	4331182

Taqman mouse Six3 probe and primers	Thermo	Mm01237639_m1	4331182
Taqman mouse Otx2 probe and primers	Thermo	Mm00446859_m1	4331182
Taqman mouse Tal1 probe and primers	Thermo	Mm01187033_m1	4331182
Taqman mouse Tbx20 probe and primers	Thermo	Mm00451515_m1	4331182
Taqman mouse Nkx2.5 probe and primers	Thermo	Mm01309813_s1	4331182
Taqman mouse Pecam1 probe and primers	Thermo	Mm01242576_m1	4331182

Table S2. Primers used in ChIP-qPCR experiments

Name	Forward sequence (5'->3')	Reverse sequence (5" ->3')	Location
<i>Arx</i>	CTTGTTACCGCTTGTCTGA G	TGCGATCTTTGTTCCCTTCC	gene body
<i>Cyp2c37</i>	CTCGAGATGCAGGTAGTGA GAA	TGGGACCAATACTGCTGATT C	gene body
<i>Cyp2u1</i>	GTCACAGCCAATGACCCTTT	GCTCAACTCTCCCAGCCTTA	gene body
<i>Cyp3a11</i>	CAATAGTCAGCCCGTTTACG	CCATTCACCACATCCTTGG	gene body
<i>Cyp3a25</i>	TGTGTCTCACAATACAATGA CAC	CTGAGTCAAGGGTGTAAATCC	gene body
<i>Cyp3a44</i>	CTGCAGCATACAAAGCCA	CTGTCCTTTTCTCTGCTTTG	gene body
<i>Eef1a</i>	GCTATGCAAGCAACATTTCC TAT	CCTCCACTTGGTAAGCCTGA	gene body
<i>Gapdh</i>	TATAGGGCCTGGGTCAGTG C	AGCTGAGTCATGGTGGTTCT G	gene body
<i>MSAT</i>	GACGACTTGAAAAATGACGA AATC	CATATTCCAGGTCCTTCAGT GTGC	repeats
<i>Mug1</i>	ATAAGCCAGAGTATGTGAGG G	TGGAATACAAGGATTTGGGT G	gene body
<i>Mug2</i>	GATAGTTAAAATTATTCAGG ATCACC	CTTCTATAGTTTACTGCATGT CGC	gene body
<i>Zfp936</i>	CAACATAGTAAAGTGCTTGC ATGT	CGAAGATAGCTGGGTTTGTA GG	gene body

Table S3. List of antibodies used in this study

Antibody name	Source	Identifier	Application
anti-histone H3 (tri methyl K9) validated fig.S5A,B; S20A,C	Abcam	ab8898	ChIP, WB, IHC
anti-histone H3 (di methyl K9)	Abcam	ab1220	WB
anti-histone H3K27me3	Active Motif	#39155	WB, IHC
anti-histone H3K27me3	Millipore	07-449	ChIP
anti-histone H3 (tri methyl K4)	Abcam	ab8580	Chip
anti-histone H3	Abcam	ab1791	WB
anti-TBP	Abcam	ab51841	WB
anti-MUPs validated fig.S22A,B	Santa Cruz Biotechnology	sc-66976	IHC
anti-Cytokeratin 7 validated Fig.4C	Abcam	ab181598	IHC
anti-FLK1	BD Pharmingen	#550549	cell sorting
anti-ENDM1	In-house (Gadue et al., 2009)		cell sorting
anti-LIV2	MBL	D118-3	cell sorting
anti-Osteocalcin validated fig.S22B	Abcam	Ab93876	IHC
anti-FOXM1	Santa Cruz Biotechnology	sc-376471	WB
anti-mcm2	Cell Signaling	4007S	WB, IHC
anti-Setdb1 validated fig.S20A	Thermo Scientific	MA5-15722 Lot #RI2264653	WB, IHC
anti-Suv39h1 validated fig.S20A	Thermo Scientific	MA1-25505 Lot #SL2492781	WB
anti-Suv39h2 validated fig.S20B	Thermo Scientific	PA5-11366 Lot #SI2420746A	IHC
anti-Vinculin	Sigma	v9131	WB
goat anti-rabbit IgG-HRP	Santa Cruz Biotechnology	sc-2004	WB, IHC
goat anti-mouse IgG-HRP	Santa Cruz Biotechnology	sc-2005	WB, IHC
PE-conjugated goat anti-rat IgG	Biolegend	#405406	cell sorting

Table S4. List of primers used for single cell RNA-Seq

Primer	Sequence (5'->3')
Template Switch Oligo	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG
TSO-PCR	AAGCAGTGGTATCAACGCAGAGT
P5-TSO- Hybrid(RB519)	AATGATACGGCGACCACCGAGATCTACACACGAGTGAGCCTGTCCG GGAAGCAGTGGTATCAACGCAGAGT*A*C
P5-TSO- Hybrid(RB520)	AATGATACGGCGACCACCGAGATCTACACCGTGTCTAGCCTGTCCG CGGAAGCAGTGGTATCAACGCAGAGT*A*C
Nextera RB701	CAAGCAGAAGACGGCATAACGAGAT CGATAACC GTCTCGTGGGCTCGG AGATG
Nextera RB702	CAAGCAGAAGACGGCATAACGAGAT CTAGCGTA GTCTCGTGGGCTCGG AGATG
Nextera RB703	CAAGCAGAAGACGGCATAACGAGAT AATGGCCT GTCTCGTGGGCTCGG AGATG
Nextera RB704	CAAGCAGAAGACGGCATAACGAGAT GTTACCTC GTCTCGTGGGCTCGG AGATG
Nextera RB705	CAAGCAGAAGACGGCATAACGAGAT ACACGTGT GTCTCGTGGGCTCGG AGATG
Nextera RB706	CAAGCAGAAGACGGCATAACGAGAT CGTCATGA GTCTCGTGGGCTCGG AGATG
Read1 primer	GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC
DropSeqIndex2	GTTGATACCACTGCTTCCGCGGACAGGC

Table S5. Oligonucleotide sequences for zygote injection to generate Suv39h1 and Suv39h2 mutant alleles and for genotyping

Oligonucleotide name	Sequence (5'->3')
Suv39h1 intron 2 gRNA	GAGGGGCTGGCCGAGTATGGTGG
Suv39h1 intron 5 gRNA	GAAGGAGGATTACAGTCAGGAGG
Suv39h2 exon 2 gRNA	AAGCTCACATGTAAATCGATTGG
Suv39h1 intron 2 loxP site	GATATCATAACTTCGTATAATGTATGCTATACGAAGTTAT
Suv39h1 intron 5 loxP site	GGATCCATAACTTCGTATAATGTATGCTATACGAAGTTAT
Suv39h1 intron 2 HDR oligo	GCAGGATAGTTAGAGGAGCCCACTGAAAGTAGCATGCTTAGT TTAAATGTAGCCTTCATTTTCTGTCTATAAAATGGGAGGGGCT GGCCGAGTGATATCATAACTTCGTATAATGTATGCTATACGAA GTTATATGGTGGCCCAAGCCTTTAATCTCAGCACTGGGGAGG CAGAGGCAGGTTGATTTCTGAGTTTGAGGC
Suv39h1 intron 5 HDR oligo	TGTGGGAGAGGATGTTTGGCTGAAGAGTTAAATATACTCACT TCCATCTTAAATCACCTACCCCTGCCCTCCTGGATCCATAA CTTCGTATAATGTATGCTATACGAAGTTATGACTGTAATCCTC CTTCACCTTCCTTGTCTGAGCCTCTGGATACCCTTGTTTTT CTAGCCTGACCTCTTCTTG
Suv39h1 intron 2 oligo for mRNA synthesis	GAGGGGCTGGCCGAGTATGGGTTTTAGAGCTAGAAATAGCA AGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCA CCGAGTCGGTGCTTTTTT
Suv39h1 intron 5 oligo for mRNA synthesis	GAAGGAGGATTACAGTCAGGGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTTT
Suv39h2 exon 2 oligo for mRNA synthesis	AAGCTCACATGTAAATCGATGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTTTT
Suv39h1 intron 2 T7-sgRNA_primer_F	GAAATTAATACGACTCACTATAGGGAGAGAGGGGCTGGCCG AGTATGGGTTTTAGA
Suv39h1 intron 5 T7-sgRNA_primer_F	GAAATTAATACGACTCACTATAGGGAGAGAAGGAGGATTACA GTCAGGGTTTTAGA
Suv39h2 exon 2 T7-sgRNA_primer_F	GAAATTAATACGACTCACTATAGGGAGAAAGCTCACATGTAAA TCGATGTTTTAGA
T7-sgRNA_primer_R	AAAAAAGCACCGACTCGGTG
Suv39h1 intron 2 genotyping primer F	GGAGCCCACTGAAAGTAGCA
Suv39h1 intron 2 genotyping primer R	ACTCCAGCCCCTCCTTTTT

Suv39h1 intron 5 genotyping primer F	TCACTTCCATCTTAAATCACCTACC
Suv39h1 intron 5 genotyping primer R	GGTCAGGCTAGAAAACACAAGG
Suv39h2 exon 2 genotyping primer F	GCCTTGCCTAGTTTCACTTGAT
Suv39h2 exon 2 genotyping primer R	GGACATTGCCTTACCTTTGC

Other Supplementary Materials for this manuscript include the following:

Table S6. mm9_reads_alignment_summary.xlsx

Table S7. srHC_Seq_genes_in_alluvia.xlsx

Table S8. Genes_decorated_with_H3K9me3.xlsx

Table S9. H3K9me3_at_TSS_and_TTS.xlsx

Table S10. H3K9me3_dynamics_at_stage_transitions.xlsx

Table S11. H3K9me3_Patch_Gene_coverage.xlsx

Table S12. RNA-seq_differentially_expressed_genes_GO_analysis.xlsx

Table S13. RNA_seq_clustering.xlsx

Table S14. Quantile_normalized_AUCs_Z_score_hepatic_and_pancreatic_genes.xlsx

Table S15. scRNA_seq_genes_and_differentially_expressed_genes_analysis.xlsx

Table S16. Differentially_expressed_genes_1mo_Ctrl_vs_TKO_livers.xlsx

3.9. References

1. D. Duboule, Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Camb. Engl. Suppl.*, 135–142 (1994).
2. R. A. Raff, *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Chicago, IL: University of Chicago Press).
3. N. Irie, S. Kuratani, The developmental hourglass model: a predictor of the basic body plan? *Dev. Camb. Engl.* 141, 4649–4655 (2014).
4. E. Meshorer, T. Misteli, Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.* 7, 540–546 (2006).

5. B. Wen, H. Wu, Y. Shinkai, R. A. Irizarry, A. P. Feinberg, Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* 41, 246–250 (2009).
6. K. Ahmed et al., Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One.* 5, e10531 (2010).
7. R. D. Hawkins et al., Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell.* 6, 479–491 (2010).
8. J. Zhu et al., Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell.* 152, 642–654 (2013).
9. F. Ugarte et al., Progressive Chromatin Condensation and H3K9 Methylation Regulate the Differentiation of Embryonic and Hematopoietic Stem Cells. *Stem Cell Rep.* 5, 728–740 (2015).
10. M. J. Vogel et al., Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* 16, 1493–1504 (2006).
11. T. Chen, S. Y. R. Dent, Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat. Rev. Genet.* 15, 93–106 (2014).
12. L. L. Wallrath, S. C. Elgin, Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev.* 9, 1263–1277 (1995).
13. N. Gilbert et al., Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell.* 118, 555–566 (2004).
14. E. Fussner et al., Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* 30, 1778–1789 (2011).
15. A. H. Peters et al., Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell.* 107, 323–337 (2001).
16. C. Beisel, R. Paro, Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12, 123–135 (2011).
17. J. S. Becker, D. Nicetto, K. S. Zaret, H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet. TIG.* 32, 29–41 (2016).
18. G. Almouzni, A. V. Probst, Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucl. Acids Res.* 2, 332–338 (2011).
19. P. Martínez, M. A. Blasco, Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nat. Rev. Cancer.* 11, 161–176 (2011).
20. S. J. Nielsen et al., Rb targets histone H3 methylation and HP1 to promoters. *Nature.* 412, 561–565 (2001).
21. A. H. F. M. Peters et al., Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat. Genet.* 30, 77–80 (2002).
22. J. E. Dodge, Y.-K. Kang, H. Beppu, H. Lei, E. Li, Histone H3-K9 methyltransferase ESET is essential for early development. *Mol. Cell. Biol.* 24, 2478–2486 (2004).
23. M. Tachibana, M. Nozaki, N. Takeda, Y. Shinkai, Functional dynamics of H3K9 methylation during meiotic prophase progression. *EMBO J.* 26, 3346–3359 (2007).
24. C. Wang et al., Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* 20, 620–631 (2018).

25. J. S. Becker et al., Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell.* 68, 1023-1037.e15 (2017).
26. A. Soufi, G. Donahue, K. S. Zaret, Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell.* 151, 994–1004 (2012).
27. S. Matoba et al., Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell.* 159, 884–895 (2014).
28. C. S. Lee, J. R. Friedman, J. T. Fulmer, K. H. Kaestner, The initiation of liver development is dependent on Foxa transcription factors. *Nature.* 435, 944–947 (2005).
29. A. Calmont et al., An FGF response pathway that mediates hepatic gene induction in embryonic endoderm cells. *Dev. Cell.* 11, 339–348 (2006).
30. T. Matsui et al., Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature.* 464, 927–931 (2010).
31. P. Gadue et al., Generation of monoclonal antibodies specific for cell surface molecules expressed on early mouse endoderm. *Stem Cells Dayt. Ohio.* 27, 2103–2113 (2009).
32. C.-R. Xu et al., Chromatin “prepattern” and histone modifiers in a fate choice for liver and pancreas. *Science.* 332, 963–966 (2011).
33. C.-R. Xu et al., Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. *EMBO J.* 33, 2157–2170 (2014).
34. D.-S. Li, Y.-H. Yuan, H.-J. Tu, Q.-L. Liang, L.-J. Dai, A protocol for islet isolation from mouse pancreas. *Nat. Protoc.* 4, 1649–1652 (2009).
35. J. Henao-Mejia et al., *Cold Spring Harb. Protoc.*, in press, doi:10.1101/pdb.prot090704.
36. X. Zhu et al., An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.* 4, 6420 (2014).
37. H. O'Geen, L. Echipare, P. J. Farnham, Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol. Biol. Clifton NJ.* 791, 265–286 (2011).
38. S. Rodrigue et al., Unlocking short read sequencing for metagenomics. *PLoS One.* 5, e11840 (2010).
39. N. J. Lennon et al., A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol.* 11, R15 (2010).
40. A. C. Clarke et al., From cheek swabs to consensus sequences: an A to Z protocol for high-throughput DNA sequencing of complete human mitochondrial genomes. *BMC Genomics.* 15, 68 (2014).
41. B. Egan et al., An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. *PLoS One.* 11, e0166438 (2016).
42. V. Schwämmle, O. N. Jensen, A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinforma. Oxf. Engl.* 26, 2841–2848 (2010).
43. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57 (2009).
44. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13 (2009).

45. E. Eden, D. Lipson, S. Yogev, Z. Yakhini, Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* 3, e39 (2007).
46. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 10, 48 (2009).
47. J. Alles et al., Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 15, 44 (2017).

Chapter 4: Perspectives and Future Directions

4.1. Sonication-resistant heterochromatin sequencing as a method to map compacted heterochromatin and open euchromatin

There are hundreds of cell types that cooperate to perform the necessary physiologies of complex multicellular organisms such as mammals and, remarkably, each of these cell types have the same complement of chromosomes with essentially the same sequence. Packaging DNA into heterochromatin, a compact and repressive structure, is key for locking down cellular identity in development and maintenance (Becker et al., 2016; Fadloun et al., 2013), reprogramming (Becker et al., 2017; Mansour et al., 2012; Soufi et al., 2012), and is perturbed in disease settings (Rao et al., 2017; Schuettengruber et al., 2017). Classically, heterochromatin has been defined by physical compaction as seen through a microscope (Brown, 1966; Heitz, 1928), but more recently heterochromatin has been defined by the histone marks H3K9me3 or H3K27me3. H3K9me3-marked chromatin is often defined as constitutive heterochromatin that represses pericentromeric repeat elements to maintain genome stability (Fukagawa et al., 2004; Lehnertz et al., 2003), but more recent studies have shown cell type-specific repression of genes as well (Becker et al., 2017; Hawkins et al., 2010; Nicetto et al., 2019; Zhu et al., 2013). H3K27me3-marked chromatin is classically defined as facultative heterochromatin, as it has lineage- and temporally-specific patterns and represses many lineage-determining genes (Margueron and Reinberg, 2011; Xu et al., 2014; Zhu et al., 2013). However, definitions of heterochromatin based on histone marks are limited, as these H3K9me3- or H3K27me3-marked chromatin can be physically heterochromatic or euchromatic with

compaction state predicting gene activity and accessibility to binding by transcription factors (Becker et al., 2016; Beisel and Paro, 2011; Blahnik et al., 2011; Breiling et al., 2001; Grindheim et al., 2019; Hawkins et al., 2010; Nicetto et al., 2019; Piacentini et al., 2003; Riddle et al., 2012; Trojer and Reinberg, 2007; Vakoc et al., 2005).

Heterochromatic and euchromatic regions can be mapped genome-wide with a new technique termed gradient-seq (Becker et al., 2017). Gradient-seq was inspired by an initial observation that input sequencing tracks from ChIP-sequencing experiments are depleted for signal in H3K9me3-marked heterochromatin domains, suggesting that heterochromatin is sonication-resistant and thus underrepresented in most sequencing protocols as they involve size selection or bias for small fragments (Becker et al., 2017; Soufi et al., 2012). Supporting this hypothesis, heterochromatin is resistant to mechanical shearing (Frenster et al., 1963) and promoters of highly active genes are sonication-sensitive (Auerbach et al., 2009). The gradient-based method was chosen based on previous studies using centrifugation to analyze chromatin structure (Ghirlando and Felsenfeld, 2008; Gilbert et al., 2004; Ishihara et al., 2010) and involves fractionation of crosslinked, sonicated heterochromatin on a sucrose gradient, a setup in which heterochromatin is fast migrating and euchromatin is slow migrating (Appendix B, Figure 1). Crosslinking with formaldehyde, which creates linkages between closely-spaced molecules, about 2 Å, aids in making heterochromatin more sonication-resistant, (Quievryn and Zhitkovich, 2000; Solomon and Varshavsky, 1985). While Gradient-seq has the advantage of being able to isolate RNA, DNA, and protein from heterochromatin and euchromatin, it is also laborious, requiring tens of millions of cells and characterization of many fractions of the sucrose gradient.

To map euchromatin and heterochromatin while circumventing high cell numbers, laborious sucrose gradients, and fraction analysis, Dario Nicetto and I developed a novel

technique termed srHC-seq. srHC-seq was inspired by the observation from gradient-seq data (before sequencing) that DNA from heterochromatin is enriched for high molecular weight fragments, while DNA from euchromatin is enriched for low molecular weight fragments (Appendix B, Figure 1). My version of the srHC-seq protocol involved isolation of 10 µg DNA from crosslinked, sonicated chromatin, followed by size fractionation with SPRI beads of large, medium, and small fragments of DNA fragments. In contrast, the Nicetto version (Chapter 3, (Nicetto et al., 2019)) is aimed at lower amounts in input cells (30,000 cells) and to accommodate this, does not involve isolation of the medium fragment size population. Given that the medium fragment size population overlaps with both small and large fragment size populations, we believe that the tripartite separation will create a clearer separation between heterochromatin and euchromatin when mapped. Significantly however, results from both protocols on adult mouse hepatocytes show highly similar results, despite being performed by two different people, on two different sonication instruments, with two related but distinct size selection protocols, and with two different library prep protocols (Appendix B, Figure 2).

Robust calling of promoter and gene srHC chromatin compaction state is also possible with srHC-seq. With increasingly lax enriched window cutoffs from 10-40%, an increasing portion of the genome is called as heterochromatic or euchromatic, from approximately 500 MB of each at 10% to approximately 1100 MB at 40% (Appendix B, Figure 3AB). Despite the increasing fraction of the genome which is called as heterochromatic or euchromatic with 10% to 40% enriched window cutoffs, there is relatively mild change in the number of genes or promoters called as heterochromatic, euchromatic, or intermediate and, reflecting the robustness of this technique, euchromatic genes have high expression and heterochromatic genes have relatively low expression (Appendix B, Figure 3CD). Contrary to the term “Polycomb-marked heterochromatin”, we

find that there are genes and promoters called both as euchromatic and H3K27me3+ (Appendix B, Figure 3CD) and are expressed, albeit at lower levels than genes that are called as euchromatic and H3K27me3- at their gene bodies or promoters. We also note that promoter srHC chromatin state is a better predictor of expression than coding sequence srHC chromatin state, likely due to the large size of genes and the extreme openness of the promoters of expressed genes. Curiously, genes called as having heterochromatic gene bodies or promoters at P14 have higher levels of expression than genes called as having heterochromatic gene bodies or promoters at M2, suggesting cellular heterogeneity or incomplete “lockdown” of the genome occurs at the relatively immature P14 timepoint.

We find that srHC-seq data agrees strongly with observations that the higher-order Type A and B chromatin compartments associated with active and inactive chromatin (Boettiger et al., 2016; Lieberman-Aiden et al., 2009; Lin et al., 2012; Rao et al., 2014; Vieux-Rochas et al., 2015; Wang et al., 2016) (Appendix B, Figure 4). Interestingly, we find that a quarter of Type A compartments were called as heterochromatic and 70% of those regions are marked by H3K27me3 domains, suggesting that A heterochromatin compartments are particularly dependent on Polycomb-related compaction. Future work could investigate whether A compartment H3K27me3 domains represent the Polycomb-repressed aggregates termed “Polycomb Bodies” that localize to the nuclear interior and are dependent on nucleoplasmic LAMIN A/C to maintain transcriptional repression and aggregation (Marullo et al., 2016) and whether methods of PRC2/H3K27me3 repression sorts with A/B compartment.

4.2. PRC2/H3K27me3 function at euchromatic promoters

The factors that determine which H3K27me3-marked or PRC2-bound genes that respond to *Ezh1/2* loss with subsequent gene activation has been an ongoing question in the field. Several groups have noted that genes with promoter H3K4me2/3 (Bae et al., 2015; Ezhkova et al., 2011; Jadhav et al., 2016) and genes marked by tissue-specific H3K27me3 are also more likely to derepress in response to H3K27me3 loss (Jadhav et al., 2016). We also find low levels of pre-existing H3K4me3 at *Ezh1/2*-sensitive promoters and by plotting srHC-seq signal at the transcriptional start sites of genes with promoter H3K27me3 which do and do not upregulate in response to *Ezh1/2* loss, we additionally find that euchromatic promoter signal is a predictor of *Ezh1/2* sensitivity (Chapter 2, Figure 4).

The conflicting repression-associated and activation-associated features at the promoters of H3K27me3+ *Ezh1/2*-sensitive genes suggests that there is a fine balance being maintained at these promoters. On the pro-activation side, a transcription factor may be binding, hence the open promoter state and H3K4me3 presence. The open state may be preventing the spread of H3K27me3, since compact chromatin templates are better substrates for the PRC2 complex (Jiao and Liu, 2015; Yuan et al., 2012). H3K4me3 can inhibit PRC2 methyltransferase activity and may be preventing H3K27me3 spread (Schmitges et al., 2011). On the pro-repression side, PRC2 is recruited to unmethylated CpG islands and H3K27me3 at the promoter blocks acetylation of H3K27, which is involved in gene activation (Tie et al., 2016). Additionally, while PRC2 at promoters seemingly is not functioning through chromatin compaction, it may be able to methylate elongation factor A, which leads to reduced transcription (Ardehali et al., 2017). We additionally found that the open promoters become further euchromatic with wider open

regions in *Ezh1/2* cells, suggesting that SWI/SNF remodeling may be inhibited by PRC2 (Schuettengruber et al., 2017). Our finding that *Ezh1/2*-sensitive promoters are euchromatic underline the importance of compaction-independent repressive activities of PRC2.

4.3. PRC2/H3K27me3 involvement in postnatal hepatocyte identity

In postnatal liver maturation, we found that PRC2 is used extensively to repress multiple functional classes of genes with pre-existing euchromatic promoters, including late and early maturation genes. In the case of late maturation genes with promoter H3K27me3, the fact that these genes have euchromatic promoters and derepress, once PRC2 repression is relieved, suggests that Polycomb repression was the “last line of defense” holding these genes from being activated by transcription factors already present in the nuclear milieu. Many key liver transcription factors are expressed in the embryo and form an increasingly complex and self-reinforcing transcriptional network that persists in adults (Kyrnizi et al., 2006; Odom et al., 2006). This leaves open the question of what upstream cues signal through transcription factors to tip the balance between repression and activation in the context of maturation. Sexual maturation is one obvious potential candidate, with known growth hormone, estrogen, and androgen signaling differences in liver between males and females, and even sex-specific transcription factors (Ma et al., 2014; Torre et al., 2017), though we did not observe obvious differences. Weaning and the consequential switch from a high fat milk diet to a high carbohydrate diet with additional metabolites is another possible cue for maturation (Hashimoto and Ogawa, 2018). These same upstream hormone and metabolic cues may also cause the initial repression of early maturation genes, for which PRC2/H3K27me3 can then maintain repression. A

hypothesis-generating experiment to identify factors which affect PRC2-repressed maturation genes could be done by comparing the list of PRC2-maturation genes with the wealth of liver expression data in mutants that is publicly available.

As PRC2 represses maturation genes in a time-appropriate manner and loss of *Ezh1/2* lead to premature maturation at the transcriptional level, Polycomb repression may be impacting physiological maturation changes such as liver lobule zonation and hepatocyte polyploidization. While we did not see preferences for periportal or pericentral zoned genes to change expression as a group (Chapter 2, Supplemental Figure 2), this does not rule out the possibility that individual zoned genes are regulated by PRC2. This could be addressed in future studies using H3K27me3 ChIP-seq in periportal and pericentral hepatocytes to identify PRC2 targets and histology or single cell RNA-seq to assess gene expression in periportal and pericentral hepatocytes. Hepatocyte polyploidization is also possibly regulated by PRC2, as polyploidization is affected by expression of cell cycle factors (Chen et al., 2012a; Chipchase et al., 2003; Hsu et al., 2016; Li et al., 2013, 2013; Mayhew et al., 2005; Pandit et al., 2012), which we find downregulated prematurely in P14 *Ezh1/2* hepatocytes (Chapter 2, Figure 1), and P14 *Ezh1/2* hepatocytes stain for a proliferation marker less frequently than P14 *Wt* hepatocytes (data not shown). While pinpointing which of the many differentially expressed cell cycle or other genes are regulated by PRC2 may be complicated, overall polyploidization could be assessed with propidium iodide staining and cell sorting. Overall, finer characterization of discrete hepatocyte physiological changes that are regulated by PRC2 is an exciting new research area.

PcG proteins are noted for their roles in repressing transcription factors to influence cell fate (Bracken et al., 2006; Lewis, 1947; Slifer, 1942; Snitow et al., 2016; Xu et al., 2011), and consistent with this, we find that PRC2 represses the expression of alternative

lineage genes specifically with a euchromatic H3K27me3+/H3K4me3+ promoter state in postnatal hepatocytes. While euchromatic H3K27me3+/H3K4me3+ promoter state explains at the chromatin-level, why some alternative lineage genes are activated in response to *Ezh1/2* loss, it does not explain the physiological benefits. One hypothesis could be that *Ezh1/2*-sensitive alternative lineage genes are primed to respond to some unidentified stimulus to aid in liver function. Another hypothesis could be that these genes are simply “leftovers” from development that do not have a second redundant layer of repression. Future studies may keep in mind whether these derepressed alternative lineage genes are used in physiological responses such as hepatocyte to cholangiocyte transdifferentiation or response to acute liver injury (Yanger et al., 2013).

Future work might also investigate whether certain transcription factors or classes of transcription factors maintain the open state of sensitive promoters, what physiological functions these primed or restrained genes have, and what non-histone targets of *Ezh1/2* are involved in liver maturation. Work relative to the clinic could investigate whether human hepatocytes maturation genes are similarly regulated by PRC2/H3K27me3, whether artificially derived hepatocytes have defects in maturation gene repression or activation at genes that should or should not be repressed by PRC2, and whether manipulation of Polycomb proteins affects the maturation state of artificially derived hepatocytes.

4.4. Implications for artificially derived hepatocytes

Artificially derived hepatocytes are currently benchmarked by comparison to gene expression profiles of the starting cell type and mature hepatocytes, some metabolic tests, and transplantation assays. We propose that benchmarking could be improved by

inclusion of immature and mature transcriptional, H3K27me3, and chromatin compaction profiles. A comparison of artificially derived, immature, and mature hepatocyte transcriptional profiles could identify the extent of hepatocyte maturity and lineage fidelity by identifying three gene classes and to what extent they are expressed: 1) immature liver genes have been silenced, 2) mature liver genes have been upregulated, and 3) “hepatocyte-inappropriate”, or genes that should not be expressed in hepatocytes at all, are aberrantly expressed. Further benchmarking of correct H3K27me3 and chromatin compaction profiles would not only benchmark artificially derived hepatocytes, but also suggest why transcriptional regulation is not occurring as it does in mature hepatocytes. In the cases of immature liver genes and hepatocyte-inappropriate genes that are expressed in artificially derived hepatocytes, follow-up could determine whether these genes are failing to be repressed by PRC2 as assessed by H3K27me3 and euchromatic signal as assessed by srHC-seq at promoters, gene bodies, and possibly enhancers. Repression of immature liver genes and hepatocyte-inappropriate genes could be key for clinical efficacy because as seen in the *Ezh1/2* mouse model (Chapter 2), liver fibrosis may stem or fail to be resolved because of derepression of either or both of these gene sets. In the case of mature liver genes that fail to activate or fail to activate to the sufficient levels, a follow-up would be to ask whether these genes are inappropriately repressed by PRC2 with either heterochromatic or euchromatic H3K27me3. Mismatches in chromatin state between artificially derived hepatocytes and mature hepatocytes as defined by benchmarking could further be used to design and test methods to improve protocols.

As perturbation of PRC2 repression results in premature maturation of postnatal day 14 immature mouse hepatocytes and represses maturation genes in a time-appropriate manner, manipulation of PRC2 repression should be tested as a method to improve artificially derived hepatocytes. A first-order experiment would be to see if

transient knockdown or inhibition of *Ezh1/2* would cause a further maturation of artificially derived hepatocytes. Theoretically, transient relief of PRC2 repression could allow for transcriptional activation of inappropriately repressed liver factors and hepatocyte transcriptional network stabilization that would remain after PRC2 repression is restored. One drawback of removing all PRC2 repression is that there may be sustained expression of hepatocyte-inappropriate genes even after restoration of PRC2 activity. Further experiments would seek to determine which discrete genes or gene sets are preferentially affected by activation or inhibition of different Polycomb proteins, as it is likely that PRC2 variant complexes, PRC2 accessory proteins, and H3K27me3 histone demethylases differentially affect genes in heterochromatin or euchromatin, with or without nearby transcription, with or without transcription factor binding, and from methods that are H3K27me3-dependent or non-histone target-dependent. Inhibition of PRC2 variant complexes, PRC2.1 and PRC2.2 (Chittock et al., 2017), or Polycomb accessory proteins PHF1, JARID2, and MTF2, which affect PRC2 recruitment and enzymatic activity (Landeira et al., 2010; Li et al., 2010, 2017; Oksuz et al., 2018; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009; Son et al., 2013), may allow for derepression of liver genes. Overexpression of H3K27me3 histone demethylases JMJD3 and UTX (Hong et al., 2007) may be desirable if H3K27me3 is used to repress liver genes, while inhibition may be desired if they remove H3K27me3 from hepatocyte-inappropriate genes. Thus, perturbation of one or a combination of Polycomb factors might be used to cause repression or derepression of discrete gene sets to generate artificially derived hepatocytes more similar to native mature hepatocytes.

4.5. PRC2/H3K27me3 involvement in liver fibrosis

Liver cirrhosis, the end stage of liver fibrosis, is a complicated disease involving multiple cell types and signaling pathways with underlying mechanisms remaining incompletely understood. The most common causes in Western countries include alcoholism, hepatitis C virus infection, and nonalcoholic fatty liver disease (Di Bisceglie, 2000; Innes et al., 2013; Naveau et al., 2005). Hepatocytes, the main liver parenchymal cell type, are one of the key cell types injured by hepatotoxic agents and, when damaged, release reactive oxygen species and fibrogenic mediators. These paracrine factors then activate hepatic stellate cells to become the main recognized collagen-secreting cells in liver damage (Zhou et al., 2014). With chronic activation, the extracellular matrix builds up and compresses the sinusoids, causing portal hypertension. Portal hypertension from fibrosis causes systemwide problems with blood flow to the kidneys and kidney failure, and in the liver there is a reduction in the functional number of functional portal triads. Less liver function leads to buildup in ammonia, which is particularly toxic for the brain, decreased bilirubin conjugation leading to jaundice, hypoalbuminemia, and decreased clotting factor production which leads to coagulation deficiencies.

We observed that *Ezh1/2* hepatocytes exhibit increased apoptosis in animals as young as 1-month-old, which progresses to chronic liver damage and liver fibrosis by 2-months (Chapter 2, Figure 5), results which suggest that PRC2 is involved in repression of fibrosis-driving or fibrosis-response genes. To investigate this, we utilized two published datasets; a 232 gene signature predicting fibrosis before histopathological detection from a non-alcoholic steatohepatitis (NASH) mouse model (van Koppen et al., 2018; Teufel et al., 2016) and 121 genes upregulated in advanced human liver fibrosis as compared to early fibrosis from a mixed cohort of chronic liver disease patients with

Hepatitis C and/or fatty liver disease (Ramnath et al., 2018) (Chapter 2, Figure 6). Strikingly, a third to a half of genes from both datasets have, the predictive euchromatic H3K27me3+/H3K4me3+ promoter chromatin state that we identified as sensitizing genes to *Ezh1/2* loss and are normally downregulated in postnatal hepatic maturation. We found it curious that these fibrosis genes were not previously seen in the classes of genes aberrantly upregulated in P14 *Ezh1/2* hepatocytes, ie, prematurely upregulated late maturation genes at P14, early maturation gene that failed to be repressed at P14, and alternative lineage genes that are aberrantly upregulated at P14. Further analysis led to the observation that a striking 79 of 232 (34%) mouse nonalcoholic steatohepatitis-related fibrosis genes and 35 of 121 (29%) human advanced fibrosis genes are significantly downregulated in maturation and not significantly upregulated by P14 in *Ezh1/2* animals; thus PRC2 is used to repress these fibrosis genes at a later timepoint than P14. The temporal delay in both the liver damage phenotype and liver fibrosis gene upregulation data support a model where PRC2 is used to repress and prime fibrosis-*response* genes.

Of the liver fibrosis genes that are repressed by PRC2, *Tgfb1* and *Pdgfa* are of particular interest (while the *Pdgfa* promoter is not called as marked by H3K27me3 in *Wt* M2 hepatocytes, visual inspection of track views reveals significant upstream H3K27me3 signal in *Wt* M2 hepatocytes). TGF- β 1 is the among the most potent inducers of liver fibrogenesis (Kirmaz et al., 2004; Matsuoka and Tsukamoto, 1990; Nakatsukasa et al., 1990; Schuppan et al., 2003), PDGF-A is a strong HSC mitogen and activator, and levels of both increase with the severity of fibrosis and cirrhosis (Borkham-Kamphorst et al., 2004; Cao et al., 2010; Jeong et al., 2004; Thieringer et al., 2008). TGF- β 1 can be secreted by hepatocytes in response to liver damage and primarily targets hepatic stellate cells, where it stimulates expression of matrix-producing genes (Wells et al., 2004; Zhou et al., 2014), and can also can inhibit hepatocyte DNA synthesis and induce apoptosis

(Kirmaz et al., 2004). Mice overexpressing PDGF-A in hepatocytes develop spontaneous fibrosis (Thieringer et al., 2008) and dominant negative PDGF receptor treatment can inhibit hepatic stellate cell activation and attenuate fibrosis (Borkham-Kamphorst et al., 2004). The chronic liver damage experienced by livers with *Ezh1/2* hepatocytes may come from an inability to silence *Tgfb1* and *Pdgfa* transcription, and thus are unable to stop a feed forward liver damage response.

While further experiments are needed to confirm PRC2 repression of fibrosis genes in human liver samples, there are impactful clinical implications. Fibrosis was previously thought to be irreversible but there is growing evidence that treatment of the underlying cause of damage, such as hepatitis B, C, and autoimmune hepatitis, can lead to fibrotic reversal (Jung and Yim, 2017). In non-liver fibrosis models inhibition of an H3K27me3 demethylase with a small molecule inhibitor leads to accumulation of H3K27me3 and reduced expression at key fibrosis genes (Bergmann et al., 2018). If re-establishment of PRC2 repression at fibrosis genes in hepatocytes is part of the process of fibrotic reversal, then inhibition of H3K27me3 demethylases could be a therapeutic target.

4.5. References

- Ardehali, M.B., Anselmo, A., Cochrane, J.C., Kundu, S., Sadreyev, R.I., and Kingston, R.E. (2017). Polycomb Repressive Complex 2 Methylates Elongin A to Regulate Transcription. *Mol. Cell* 68, 872-884.e6.
- Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14926–14931.
- Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* 32, 29–41.
- Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* 68, 1023-1037.e15.

- Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* *12*, 123–135.
- Bergmann, C., Brandt, A., Merlevede, B., Hallenberger, L., Dees, C., Wohlfahrt, T., Pötter, S., Zhang, Y., Chen, C.-W., Mallano, T., et al. (2018). The histone demethylase Jumonji domain-containing protein 3 (JMJD3) regulates fibroblast activation in systemic sclerosis. *Ann. Rheum. Dis.* *77*, 150–158.
- Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O'Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., Costello, J.F., et al. (2011). Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. *PLoS ONE* *6*, e17121.
- Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Believeau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* *529*, 418–422.
- Borkham-Kamphorst, E., Herrmann, J., Stoll, D., Treptau, J., Gressner, A.M., and Weiskirchen, R. (2004). Dominant-negative soluble PDGF-beta receptor inhibits hepatic stellate cell activation and attenuates liver fibrosis. *Lab. Invest.* *84*, 766–777.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* *20*, 1123–1136.
- Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* *412*, 651–655.
- Brown, S.W. (1966). Heterochromatin. *Science* *151*, 417–425.
- Cao, S., Yaqoob, U., Das, A., Shergill, U., Jagavelu, K., Huebert, R.C., Routray, C., Abdelmoneim, S., Vasdev, M., Leof, E., et al. (2010). Neuropilin-1 promotes cirrhosis of the rodent and human liver by enhancing PDGF/TGF-beta signaling in hepatic stellate cells. *J. Clin. Invest.* *120*, 2379–2394.
- Chen, H.-Z., Ouseph, M.M., Li, J., Pécot, T., Chokshi, V., Kent, L., Bae, S., Byrne, M., Duran, C., Comstock, G., et al. (2012). Canonical and atypical E2Fs regulate the mammalian endocycle. *Nat. Cell Biol.* *14*, 1192–1202.
- Chipchase, M.D., O'Neill, M., and Melton, D.W. (2003). Characterization of premature liver polyploidy in DNA repair (Erc1)-deficient mice. *Hepatology* *38*, 958–966.
- Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. *Biochem. Soc. Trans.* *45*, 193–205.
- Di Bisceglie, A.M. (2000). Natural history of hepatitis C: its impact on clinical management. *Hepatology* *31*, 1014–1018.
- Fadloun, A., Eid, A., and Torres-Padilla, M.-E. (2013). Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions. *Curr. Top. Dev. Biol.* *104*, 1–45.
- Frenster, J.H., Allfrey, V.G., and Mirsky, A.E. (1963). REPRESSED AND ACTIVE CHROMATIN ISOLATED FROM INTERPHASE LYMPHOCYTES. *Proc. Natl. Acad. Sci. U.S.A.* *50*, 1026–1032.
- Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* *6*, 784–791.

- Ghirlando, R., and Felsenfeld, G. (2008). Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. *J. Mol. Biol.* 376, 1417–1425.
- Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118, 555–566.
- Grindheim, J.M., Nicetto, D., Donahue, G., and Zaret, K.S. (2019). PRC2 proteins EZH1 and EZH2 Regulate Timing of Postnatal Hepatocyte Maturation and Fibrosis by Repressing Gene Expression at Promoter Regions in Euchromatin in Mice. *Gastroenterology*.
- Hashimoto, K., and Ogawa, Y. (2018). Epigenetic Switching and Neonatal Nutritional Environment. *Adv. Exp. Med. Biol.* 1012, 19–25.
- Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491.
- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb Wiss Botanik* 762–818.
- Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T.D., and Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18439–18444.
- Hsu, S.-H., Delgado, E.R., Otero, P.A., Teng, K.-Y., Kutay, H., Meehan, K.M., Moroney, J.B., Monga, J.K., Hand, N.J., Friedman, J.R., et al. (2016). MicroRNA-122 regulates polyploidization in the murine liver. *Hepatology* 64, 599–615.
- Innes, H.A., Hutchinson, S.J., Barclay, S., Cadzow, E., Dillon, J.F., Fraser, A., Goldberg, D.J., Mills, P.R., McDonald, S.A., Morris, J., et al. (2013). Quantifying the fraction of cirrhosis attributable to alcohol among chronic hepatitis C virus patients: implications for treatment cost-effectiveness. *Hepatology* 57, 451–460.
- Ishihara, S., Varma, R., and Schwartz, R.H. (2010). A new fractionation assay, based on the size of formaldehyde-crosslinked, mildly sheared chromatin, delineates the chromatin structure at promoter regions. *Nucleic Acids Res.* 38, e124.
- Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N.K., Pinello, L., Yuan, G.-C., Orkin, S.H., and Shivdasani, R.A. (2016). Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* 165, 1389–1400.
- Jeong, W.-I., Do, S.-H., Yun, H.-S., Song, B.-J., Kim, S.-J., Kwak, W.-J., Yoo, S.-E., Park, H.-Y., and Jeong, K.-S. (2004). Hypoxia potentiates transforming growth factor-beta expression of hepatocyte during the cirrhotic condition in rat liver. *Liver Int.* 24, 658–668.
- Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* 350, aac4383.
- Jung, Y.K., and Yim, H.J. (2017). Reversal of liver cirrhosis: current evidence and expectations. *Korean J. Intern. Med.* 32, 213–228.
- Kirmaz, C., Terzioglu, E., Topalak, O., Bayrak, P., Yilmaz, O., Ersoz, G., and Sebik, F. (2004). Serum transforming growth factor-beta1(TGF-beta1) in patients with cirrhosis, chronic hepatitis B and chronic hepatitis C [corrected]. *Eur. Cytokine Netw.* 15, 112–116.
- Kyrmizi, I., Hatzis, P., Katrakili, N., Tronche, F., Gonzalez, F.J., and Talianidis, I. (2006). Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev.* 20, 2293–2305.

- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.
- Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* 13, 1192–1200.
- Lewis, P.H. (1947). New mutants. *Drosoph. Inf. Serv.*
- Li, D., Cen, J., Chen, X., Conway, E.M., Ji, Y., and Hui, L. (2013). Hepatic loss of survivin impairs postnatal liver development and promotes expansion of hepatic progenitor cells in mice. *Hepatology* 58, 2109–2121.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.
- Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291.
- 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.
- Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13, 1196–1204.
- Ma, W.-L., Lai, H.-C., Yeh, S., Cai, X., and Chang, C. (2014). Androgen receptor roles in hepatocellular carcinoma, fatty liver, cirrhosis and hepatitis. *Endocr. Relat. Cancer* 21, R165-182.
- Mansour, A.A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., Krupalnik, V., Zerbib, M., Amann-Zalcenstein, D., Maza, I., et al. (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* 488, 409–413.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.
- Marullo, F., Cesarini, E., Antonelli, L., Gregoretti, F., Oliva, G., and Lanzuolo, C. (2016). Nucleoplasmic Lamin A/C and Polycomb group of proteins: An evolutionarily conserved interplay. *Nucleus* 7, 103–111.
- Matsuoka, M., and Tsukamoto, H. (1990). Stimulation of hepatic lipocyte collagen production by Kupffer cell-derived transforming growth factor beta: implication for a pathogenetic role in alcoholic liver fibrogenesis. *Hepatology* 11, 599–605.
- Mayhew, C.N., Bosco, E.E., Fox, S.R., Okaya, T., Tarapore, P., Schwemberger, S.J., Babcock, G.F., Lentsch, A.B., Fukasawa, K., and Knudsen, E.S. (2005). Liver-specific pRB loss results in ectopic cell cycle entry and aberrant ploidy. *Cancer Res.* 65, 4568–4577.
- Nakatsukasa, H., Nagy, P., Evarts, R.P., Hsia, C.C., Marsden, E., and Thorgeirsson, S.S. (1990). Cellular distribution of transforming growth factor-beta 1 and procollagen types I, III, and IV transcripts in carbon tetrachloride-induced rat liver fibrosis. *J. Clin. Invest.* 85, 1833–1843.

- Naveau, S., Perlemuter, G., and Balian, A. (2005). [Epidemiology and natural history of cirrhosis]. *Rev Prat* 55, 1527–1532.
- Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*.
- Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* 2, 2006.0017.
- Oksuz, O., Narendra, V., Lee, C.-H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* 70, 1149-1162.e5.
- Pandit, S.K., Westendorp, B., Nantasanti, S., van Liere, E., Tooten, P.C.J., Cornelissen, P.W.A., Toussaint, M.J.M., Lamers, W.H., and de Bruin, A. (2012). E2F8 is essential for polyploidization in mammalian cells. *Nat. Cell Biol.* 14, 1181–1191.
- Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310.
- Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290–1302.
- Piacentini, L., Fanti, L., Berloco, M., Perrini, B., and Pimpinelli, S. (2003). Heterochromatin protein 1 (HP1) is associated with induced gene expression in *Drosophila* euchromatin. *J. Cell Biol.* 161, 707–714.
- Quiévryn, G., and Zhitkovich, A. (2000). Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteasome function. *Carcinogenesis* 21, 1573–1580.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Rao, V.K., Pal, A., and Taneja, R. (2017). A drive in SUVs: From development to disease. *Epigenetics* 12, 177–186.
- Riddle, N.C., Jung, Y.L., Gu, T., Alekseyenko, A.A., Asker, D., Gui, H., Kharchenko, P.V., Minoda, A., Plachetka, A., Schwartz, Y.B., et al. (2012). Enrichment of HP1a on *Drosophila* chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet.* 8, e1002954.
- Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* 42, 330–341.
- Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171, 34–57.
- Schuppan, D., Krebs, A., Bauer, M., and Hahn, E.G. (2003). Hepatitis C and liver fibrosis. *Cell Death Differ.* 10 Suppl 1, S59-67.
- Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139, 1303–1314.

- Slifer, E.H. (1942). A mutant stock of *Drosophila* with extra sex combs. *J. Exp. Zool.* 31–40.
- Snitow, M., Lu, M., Cheng, L., Zhou, S., and Morrisey, E.E. (2016). Ezh2 restricts the smooth muscle lineage during mouse lung mesothelial development. *Development* 143, 3733–3741.
- Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6470–6474.
- Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* 27, 2663–2677.
- Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151, 994–1004.
- Thieringer, F., Maass, T., Czochra, P., Klopčič, B., Conrad, I., Friebe, D., Schirmacher, P., Lohse, A.W., Blessing, M., Galle, P.R., et al. (2008). Spontaneous hepatic fibrosis in transgenic mice overexpressing PDGF-A. *Gene* 423, 23–28.
- Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad. Sci. U.S.A.* 113, E744–753.
- Torre, D., Lolli, F., Ciana, P., and Maggi, A. (2017). Sexual Dimorphism and Estrogen Action in Mouse Liver. *Adv. Exp. Med. Biol.* 1043, 141–151.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13.
- Vakoc, C.R., Mandat, S.A., Olenchok, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell* 19, 381–391.
- Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4672–4677.
- Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602.
- Wells, R.G., Kruglov, E., and Dranoff, J.A. (2004). Autocrine release of TGF-beta by portal fibroblasts regulates cell growth. *FEBS Lett.* 559, 107–110.
- Xu, C.-R., Cole, P.A., Meyers, D.J., Kormish, J., Dent, S., and Zaret, K.S. (2011). Chromatin “prepattern” and histone modifiers in a fate choice for liver and pancreas. *Science* 332, 963–966.
- Xu, C.-R., Li, L.-C., Donahue, G., Ying, L., Zhang, Y.-W., Gadue, P., and Zaret, K.S. (2014). Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. *EMBO J.* 33, 2157–2170.
- Yanger, K., Zong, Y., Maggs, L.R., Shapira, S.N., Maddipati, R., Aiello, N.M., Thung, S.N., Wells, R.G., Greenbaum, L.E., and Stanger, B.Z. (2013). Robust cellular reprogramming occurs spontaneously during liver regeneration. *Genes Dev.* 27, 719–724.
- Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* 337, 971–975.

- Zhou, W.-C., Zhang, Q.-B., and Qiao, L. (2014). Pathogenesis of liver cirrhosis. *World J. Gastroenterol.* 20, 7312–7324.
- Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654.

Appendix A: Murine hepatocyte isolation by liver perfusion

This appendix describes how to isolate mouse hepatocytes from liver from animals using a two-step collagenase perfusion protocol. It is designed for mice at least two weeks old or at least 6-7 g in total body weight. Also included are protocols for confirming enrichment of hepatocytes with RTqPCR and isolating hepatocyte nuclei.

Key to this protocol are the perfusion and digest medias, which are manufactured by Invitrogen. Invitrogen appears to have developed their perfusion and digest medias from a 1987 Ichihara Lab protocol (Nakatani et al., 2002; Shimaoka et al., 1987). While the constituents of these medias are proprietary, it is likely that, per the 1987 protocol, perfusion media is Ca⁺⁺-free Hanks' solution containing 5 mM EGTA at 37°C. Per the catalog, "Liver Perfusion Medium is a buffered, balanced salt solution formulated to cleanse the liver of blood, prevent clotting, and initiate loosening of cell-to-cell contact". Invitrogen describes digest media as containing collagenase and dispase, which digest the extracellular matrix.

I. Objective

Enrichment of hepatocytes from other liver cell types is preferred for studies in which contaminating cell types muddy data interpretation, such as in transcript or chromatin studies. Additionally, isolated and cultured hepatocytes can be used for other studies, albeit the technology maintaining hepatocyte function and identity in culture is currently limited.

II. Key Reagents

Isoflurane (Butler Animal Health Supply, Cat# 029405)

Or other IACUC-approved anesthesia method

22- or 24-gauge catheters (Midwest Veterinary Supply, 381423)

22-gauge for 2-month-old mice (M2), 24-gauge for postnatal day 14 (P14) mice. These catheters have an auto-eject feature, which means that the needle retracts from the catheter when a button is pushed, which is useful because manually retracting the needle can be shaky and can tear the blood vessel.

This protocol is designed for P14 mice or older, which have a total body weight of 6-7 grams on the lower end. Animals younger than P14 are sometimes too small for cannulation with a 24-gauge catheter

37°C liver perfusion media (Invitrogen 17701-038)

25 mL for each P14 animal, 50 mL for each M2 or other adult animal.

Store at 4°C.

37°C liver digest media (Invitrogen 17703-034)

22 mL for each P14 animal, 45 mL for each M2 or other adult animal.

Store at -20°C in 45 mL aliquots. Digest media comes with white particulate. Avoid adding particulate to the aliquots or the tubing because it clogs the bubble trap and can function as a blood clot in the liver.

cell scrapers

swivel head scrapers are easier to work with

70% EtOH

in squirt bottle

PBS

for rinsing tubing

H₂O

for rinsing tubing

William's E Medium (Sigma W4128)

Given the short time in media, the exact media is probably not key unless the hepatocytes are destined for culture.

100 µm cell filter (BD 352360)

Dissecting scissors

curved tip scissors are easier to use

Serrated tweezers

serrated tweezers are better at gripping

Peristaltic pump setup

From input to output: 42°C water bath with media, tubing from media to bubble trap, tubing from bubble trap to threaded adaptor (threaded to screw onto catheter). Peristaltic pump is set up somewhere along the tubing.

The water bath is set at 42°C to mitigate the cooling effects from the media flowing through the tubing.

Bubble trap function: The bubble trap is meant to ensure media but not bubbles reach the catheter since bubbles can act like blood clots. When media is coming into the bubble trap, have it facing input-side down. Gravity creates a reservoir of media inside of the bubble trap. Later, flip the bubble trap to output-side down so that the media falls because of gravity and air rises because of gravity and no bubbles come through the output side of the tube and into the mouse.

10 cm petri dish

Cotton applicators

Surgical qtips

Anesthesia chamber

Surgical setup

Wrap a Styrofoam board, approximately 18 inches by 12 inches, in cellophane for easy cleanup. Prop up the back the Styrofoam on a roll of lab tape or the lid of a tip box. The angled work surface makes it easier to perform surgery and predict where excess liquid will flow. Have a layer of paper towels under the mouse and taller stacks of paper towels framing the left, right, and bottom of the surgical table to soak up excess media.

Perform the perfusion in a chemical hood to avoid exposure to mouse dander and isoflurane.

TRizol Reagent (Thermo 15596026)

iScript cDNA synthesis kit (BioRad 1708891)

or other cDNA synthesis kit

Power SYBR Green PCR Master Mix Set (Thermo 4367659)

RSB

A hypotonic cytoplasmic membrane lysis buffer used in the isolation of hepatic nuclei.

10 mM Tris-HCl, pH 7.4

10 mM NaCl

3 mM MgCl₂

0.5% NP40

Add fresh:

cOmplete Protease Inhibitor Cocktail to a final concentration of 2X (Roche 04693116001)

5 mM PMSF

III. Protocol

Liver perfusion, hepatocyte isolation, and nuclei isolation

See Figure 1 for schematic.

1. Prepare workspace

- a. Warm perfusion and digest media to 42°C in the water bath
- b. Sterilize dissecting scissors and tweezers in EtOH
- c. Rinse tubing in PBS, then EtOH, then PBS.
- d. Turn on peristaltic pump. The tubing has been squished together and may be stuck closed, but ten minutes running without media usually massages out any stuck kinks.

2. Anesthetize the mouse

- a. Per lab IACUC guidelines.
- b. Monitor anesthesia paw pinch. The mouse should continue visibly breathing until sometime after cannulation and blood loss.

3. Cannulation and perfusion

- a. Apply 70% EtOH on the abdomen to keep hair out of the abdominal cavity.
- b. Start the peristaltic pump at ~8 mL/min, with the tubing in perfusion media. The bubble trap should be input-side down.
- c. Using tweezers, grasp the skin of the abdomen, and make a midline incision just above the urethra. Open to the left and right to make a U-shaped incision that exposes the abdominal cavity. Make sure not to cut abdominal organs, like the liver, which often suctions to the diaphragm and body wall.
- d. Push the viscera to the right (left side of the mouse) out of the abdominal cavity with a qtip to reveal the inferior vena cavae (IVC), which runs near the spine.
 - i. Sometimes the bladder will obstruct the IVC but squeezing the bladder with blunt tweezers forces urine out and reduces bladder size.
 - ii. If the lower portion of IVC (near the kidneys) is hidden by fat or connective tissue, as happens often in fat or old mice, gently rub the IVC down towards the tail with the qtip in long strokes, then the connective tissue will rub off, leaving a more exposed IVC.
- e. Flip the liver up and back towards the head to reveal the blood vessels beneath.
- f. Flip the bubble trap to output-side down so media but not air moves through the output tube.
- g. For right-handed people: Have a qtip in the left and the cannula in the right. Block blood flow by using the qtip to apply gentle pressure to the IVC close to

the liver, preferably above the connection of the IVC and the aorta abdominalis. This expands the diameter of the IVC.

- h. Cannulate the IVC with the needle bevel side up.
 - i. For mice weighing 7 grams or more (adult mice and most P14), cannulate the IVC below the connection to the aorta abdominalis
 - ii. For mice weighing less than 7 grams, cannulate the IVC above the connection to the aorta abdominalis because it has a wider diameter
 - iii. Come in at an angle almost parallel to the IVC so as not to poke through the other side of the vein. This step needs to be smooth and sure, because if the needle is shaking, it can tear through the vein.
- i. Press the auto-retract button to retract the needle and leave the cannula in place.
 - i. There should be “flashback”, or blood traveling back up the cannula. Sometimes when this does not happen the perfusion still works.
- j. When the perfusion media is coming out the output side the tubing (and attached adaptor), screw the end onto the cannula.
 - i. The blood vessels should immediately start swelling because media is being pumped into a closed system.
- k. Immediately sever the portal vein.
 - i. The liver should immediately blanch (change color from red to beige) and over the next minutes will slightly swell. If the liver does not blanch, a qtip can be used to gently massage the liver a little to dislodge a clot. Do not over-massage.
- l. After perfusion media switch to digest media.

4. Liver dissociation

- a. Dissect out the liver, transfer to a Petri dish with ~20 mL William's E media.
- b. For “total perfused liver” samples, cut off a small chunk of liver.
- c. Dissociate with a cutting motion, not a scraping motion. Release of cells makes the media cloudy. When all the cells possible have been removed by the cutting motion, then scrape the “husk” of the liver for remaining cells.
- d. Filter dissociated cells through a 100 μ M cell strainer.
- e. Spin at 50 x g at 4°C for 5 min in a swinging bucket centrifuge to form a loose hepatocyte pellet. Hepatocytes are very large and pellet at this slow speed.
- f. For “supernatant” samples, which represents non-hepatocyte liver cell types, pellet remaining cells from the supernatant at 500 x g at 4°C for 10 min.
- g. Resuspend hepatocytes in 10 mL William's E and pellet again. This is the hepatocyte fraction
- h. Count hepatocytes with a hemocytometer. Automated cell counters have trouble with hepatocytes because they look granular.

- i. Rinse tubing with PBS, then 70% EtOH, then PBS, then H₂O.

5. Nuclei isolation

- a. Resuspend hepatic pellet in 10 mL RSB.
 - i. Too much contaminating media will change the tonicity of the solution and the hepatocytes will not swell as needed for efficient cytoplasmic membrane lysis.
- b. Immediately dounce 13 times with a tight-fitting Wheaton dounce attached to spinning drill press.
- c. Check for nuclei.
- d. Count nuclei with a hemocytometer. Automated cell counters have trouble with all the debris.
- e. Pellet nuclei at 100 x g at 4°C for 5 min

Assessing hepatocyte enrichment

This describes a standard RTqPCR analysis of total perfused liver, isolated hepatocytes, and pelleted supernatant from the hepatocyte perfusion protocol. Total perfused liver represents total liver minus cell types lost to perfusion, such as red blood cells. Pelleted supernatant represents non-hepatocyte cell types that were not pelleted with hepatocytes. In this protocol, hepatocyte enrichment is assessed by expression of different liver cell type marker genes in the perfused liver, hepatocyte, and pelleted supernatant fractions. See Figure 1 for results.

Target	Forward, Reverse Sequences
<i>Hnf4a</i> variants 1,2,3	ggcaatgacacgtccccatc cacagatggcgcacaggg
<i>Foxa1</i>	tggctccaggatgttaggga acagggacagaggagtaggc
<i>Krt19</i>	ccctccgagattacaacca ggcagcattgtcaatctgt
<i>Gfap</i>	ctttgacaggacctcggca acgcagccagggtgttctct
<i>Des</i>	gtggagcgtgacaacctgat tcggaaggcagccaagttgt
<i>Pecam1</i> variant 1	aggattcagctgaggtgggc tggatgctgtgatggtgaagg
<i>Acta2</i>	ggagaagcccagccagtcg ccagagccattgtcgcacac
<i>Ptprc</i>	IDT primer assay: Mm.PT.58.7583849
<i>Gapdh</i>	atggtgaaggtcggtgtgaac gcctgactgtgccgttgaat

1. Isolate RNA from hepatocytes, perfused liver, and pelleted supernatant from TRIzol per manufacturer's protocol.
2. Generate cDNA per manufacturer's protocol.
3. Run qPCR on the equivalent of 20 ng RNA that has been converted into double-stranded DNA.
 - a. Primers for markers hepatic cell populations.
 - i. hepatocyte: *Hnf4a*, *Foxa1*
 - ii. cholangiocytes: *Krt19*
 - iii. stellate cells: *Gfap*, *Des*
 - iv. endothelial cells: *Pecam1*, *Acta2*
 - v. immune cells, not including red blood cells or platelets: *Ptprc*
 - vi. Internal control: *Gapdh*
4. Calculate Relative Expression
 - a. Relative expression: $2^{(Ct:target - Ct:Gapdh)}$
 - b. Relative to supernatant: (relative expression supernatant)/(relative expression liver or total perfused liver)

IV. Results

RTqPCR results on total perfused liver, hepatocytes, and pelleted supernatant show that markers of non-hepatocyte liver cell types, including cholangiocytes (*Krt19*), stellate cells (*Gfap*, *Des*), endothelial cells (*Pecam1*, *Acta2*), and immune cells (*Ptprc*) overwhelmingly fractionate into the pelleted supernatant fraction, with some signal in total perfused liver. Additionally, hepatocyte makers (*Hnf4a*, *Foxa1*) are enriched in the hepatocyte fraction. Supernatant shows variable signal for hepatocyte markers because the hepatocyte pellet is very loose and a variable quantity of hepatocytes are taken when supernatant is taken. This could probably be rectified by "clearing" the supernatant with a 50 x g spin to pellet contaminating hepatocytes before pelleting supernatant at 500 x g. We conclude that hepatocytes have been dramatically enriched and non-hepatocyte cell types have been depleted in the hepatocyte fraction.

V. Figures

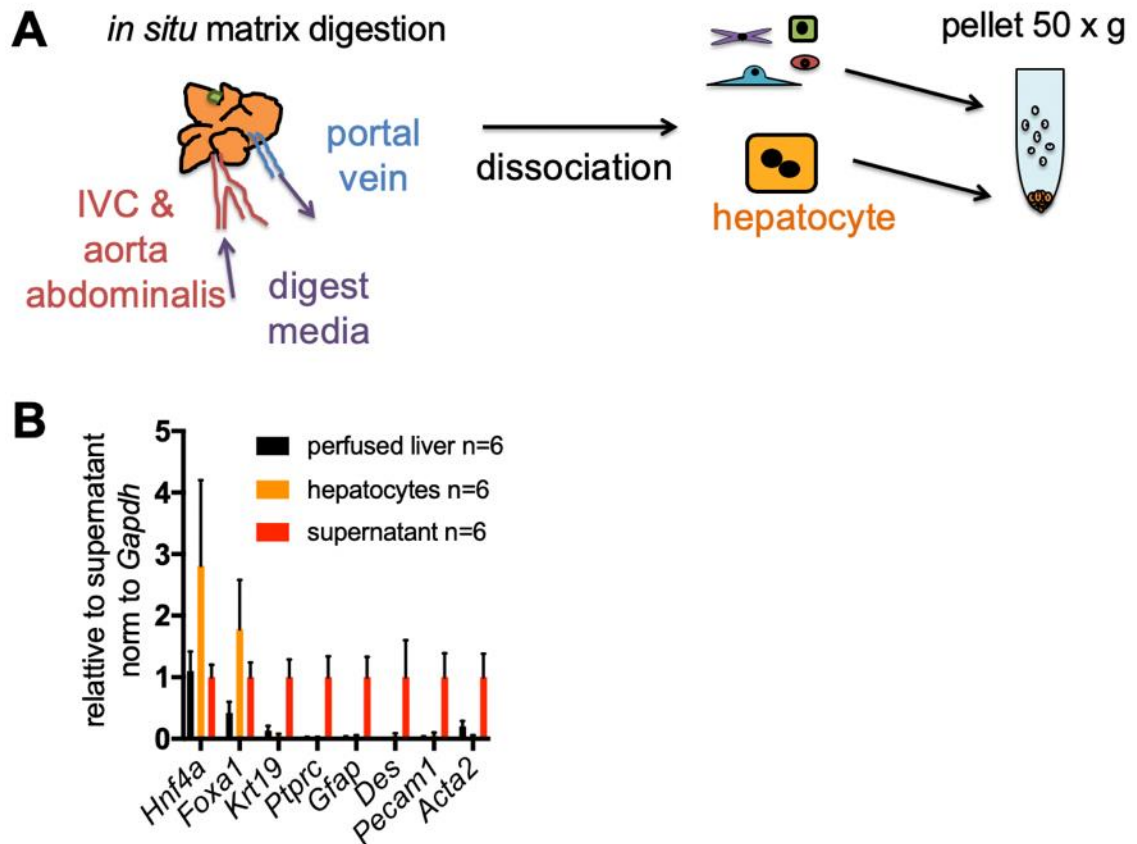


Figure 1: Isolation of murine hepatocytes

- A)** Hepatocyte enrichment strategy. The liver ECM is digested *in situ* by perfusion, the liver is dissociated, filtered through a 100 μ M filter, and hepatocytes pelleted by centrifugation. Remaining cells are pelleted from the supernatant.
- B)** Expression of markers of hepatocytes (*Hnf4a*, *Foxa1*), biliary cells (*Krt19*), blood (*Ptprc*, not red blood cells or platelets), stellate cells (*Gfap*, *Desmin*), and endothelial cells (*Pecam1*, *Acta2*) by RTqPCR from perfused liver (after perfusion and before dissociation), pelleted hepatocytes, or pelleted supernatant. Adapted from (Grindheim et al., 2019).

VI. References

- Grindheim, J.M., Nicetto, D., Donahue, G., and Zaret, K.S. (2019). PRC2 proteins EZH1 and EZH2 Regulate Timing of Postnatal Hepatocyte Maturation and Fibrosis by Repressing Gene Expression at Promoter Regions in Euchromatin in Mice. *Gastroenterology*.
- Nakatani, T., Tsuboyama-Kasaoka, N., Takahashi, M., Miura, S., and Ezaki, O. (2002). Mechanism for peroxisome proliferator-activated receptor-alpha activator-induced up-regulation of UCP2 mRNA in rodent hepatocytes. *J. Biol. Chem.* 277, 9562–9569.
- Shimaoka, S., Nakamura, T., and Ichihara, A. (1987). Stimulation of growth of primary cultured adult rat hepatocytes without growth factors by coculture with nonparenchymal liver cells. *Exp. Cell Res.* 172, 228–242.

Appendix B: Sonication Resistant Heterochromatin Sequencing, a method for mapping sonication-resistant heterochromatin

This appendix provides a detailed explanation of the sonication-resistant heterochromatin and high-throughput sequencing protocol. Dario Nicetto, PhD and I developed protocols in parallel, with my protocol designed for high cell number experiments corresponding to 10 µg isolated DNA (in Chapter 2, (Grindheim et al., 2019)) and Dario's designed for low cell number experiments (30,000 cells) (in Chapter 3, (Nicetto et al., 2019)). The high cell number version isolates large, medium, and small fragments of DNA to increase separation and resolution between small and large populations, while the low cell number version isolates large and small fragments.

This appendix also describes some observations about the relationship of srHC, H3K27me3, transcription, and Type AB compartments. srHC-seq, H3K27me3 ChIP-seq, and RNA-seq data was aligned and processed as described in (Chapter 4, (Grindheim et al., 2019)).

I. Objective

To use the differential sonication sensitivity of crosslinked chromatin to genomically identify euchromatic and heterochromatic loci.

II. Background

It has been observed that heterochromatin is resistant to mechanical shearing (Frenster et al., 1963) and promoters of highly active genes are sonication-sensitive (Auerbach et al., 2009). Genome-wide, the differential sensitivity of crosslinked chromatin to shearing by sonication was confirmed with a technique termed gradient-seq (Becker et al., 2017). The gradient-seq protocol involves fractionation of crosslinked, sonicated heterochromatin via a sucrose gradient, a setup in which heterochromatin is fast migrating and euchromatin is slow migrating (Figure 1). Reflecting the relative crosslinking efficiency and sonication sensitivity of euchromatin and heterochromatin, isolation of DNA fragments from different fractions of the sucrose gradient reveals that DNA from euchromatin corresponds to low molecular weight fragments while DNA from heterochromatin corresponds to higher molecular weight fragments.

While the gradient-seq protocol has already been developed, there are some drawbacks to this method, as is true with any experimental method. Fractionation on a sucrose gradient is laborious and requires analysis of many gradient fractions in each experiment. Additionally, it is likely that a portion of euchromatin fractionates rapidly with heterochromatin because they are crosslinked together (Figure 1). Given that my studies did not include proteomics or transcriptomics of species in euchromatin or heterochromatin, advantages allowed for by gradient-seq, I opted to develop a related technique, srHC-seq, which involves fractionation of low and high molecular weight DNA isolated from crosslinked and sonicated chromatin, corresponding to euchromatin and heterochromatin, respectively.

III. Key Reagents

Crosslinked and sonicated chromatin

This chromatin should be sonicated sufficiently so that the majority of DNA fragments are small molecular weight (≤ 500 bp) (Figure 1). Data in this Appendix and Chapter 2 was from chromatin sonicated on a Diagenode Biorupter for 30 min (30 sec on Hi, 30 sec Off). The tail of larger molecular weight DNA may or may not be visible on a 1% agarose gel with ethidium bromide, depending on gel thickness, quantity of DNA, ethidium bromide concentration, etc. See Figure 1.

Agencourt AMPure XP beads (Beckman Coulter A63881)

SPRI (Solid Phase Reversible Immobilization) beads were developed by the Whitehead Institute (DeAngelis et al., 1995). The paramagnetic property, or magnetic only in a magnetic field, of SPRI beads prevents the beads from clumping and falling out of solution. The beads are coated with carboxyl molecules which reversibly bind to DNA in the presence of the “crowding agent” polyethylene glycol and NaCl. Here “bead slurry” refers to beads in PEG/NaCl buffer.

With an increasing bead slurry to DNA in water or 0.1X TE ratio, progressively more DNA is “crowded” out of solution, with large molecular weight DNA fragments crowding out first. After crowding out of solution, DNA can bind reversibly to the beads. As AMPure XP beads have a high binding capacity (1 μ L beads can bind 3 μ g of DNA, it is unclear whether that is 1 μ L of beads or 1 μ L of bead slurry), the volumetric ratio of bead slurry to DNA in water or 0.1X TE is usually the most important factor for size selection.

magnetic Eppendorf rack

for separation of beads from supernatant

Covaris S220 or other sonicator

for sonication of large DNA fragments to a size sufficiently small enough to be compatible with Illumina library prep sequencing protocols. As of 2018, many high-throughput library preparation protocols involve size selection of small fragments and PCR amplification, which also selects for smaller fragments. Additionally, sequencing on an instrument such as Illumina Nextseq 500/550 models involves a DNA amplification step termed “cluster generation”, which also selects for small molecular weight fragments.

TES buffer

10 mM Tris-HCl, pH 8
10 mM EDTA
1% SDS

TE buffer

10 mM Tris-HCl, pH 8
10 mM EDTA

RNase A

DNase-free. Temperature quoted in the protocol below is based on the manufacturer’s recommendations.

Proteinase K

DNase-free. Temperature quoted in the protocol below is based on the manufacturer's recommendations.

Glycogen (Roche 10 901 393 001)

NEBNext Ultra DNA Library Prep Kit (NEB 7370)

Likely many other commercial kits will also work. This NEB kit has size selection steps for different sized inserts. The variability between small and large-sonicated-to-small fragments is not huge and the goal in the size selection step in the library prep protocol is to select for everything.

NEBNext Multiplex Oligos for Illumina (Index Primers Set 1-4, NEB E7335S, E7500S, E7710S, E7730S)

Agilent Bioanalyzer

The size of srHC size-selected fragments and resulting libraries need to be measured.

NEBNext Library Quantification Kit (NEB E7630)

Libraries need to be quantified prior to sequencing. This is done with information about average library size (from a Bioanalyzer run) and qPCR with this or other kits against standards of known size and concentration. The NEBNext library quantification kit is compatible with the NEBNext Ultra Library prep kit (the qPCR primers will amplify the libraries).

IV. Protocol

Isolation of DNA fragments from crosslinked and sonicated chromatin

This may be scaled up or down as necessary.

1. Decrosslinking

- a. 200 μ L chromatin + 200 μ L TES
- b. 65°C/overnight/shaking

2. RNA and protein degradation

- a. + 200 μ L TE (to dilute the SDS present for enzymatic activity)
- b. + 4 μ L 10 mg/mL RNase A for 2 hours/37°C/shaking
- c. + 4 μ L 20 mg/mL Proteinase K 2 hours/55°C/shaking

3. Phenol:chloroform:isoamyl alcohol fractionation of DNA

- a. + 400 μ L PCI
- b. vortex
- c. Spin \geq 12,000 x g for 15 min at room temperature
- d. Take 350 μ L of aqueous phase

4. Ethanol precipitation

- a. 350 μ L of aqueous phase + 1.5 μ L 10 glycogen + 32 μ L 2.5 M NaCl + 1 mL 100% EtOH
- b. Precipitate at -20°C overnight
- c. Pellet DNA: 20,000 x g for 10 min at 4°C
- d. Wash with 500 μ L ice cold 80% EtOH
- e. Air dry pellet
- f. Resuspend in 50 μ L 0.1X TE
- g. Check DNA molecular weight fragment distribution on a 1% agarose gel with 1 μ L ethidium bromide per 100 mL gel or a Bioanalyzer DNA 1000 ChIP (Figure 1)

DNA size selection of small, medium, and large fragments

Start with 10 µg DNA in 50 µL 0.1X TE or water. This quantity of DNA in the stated volume is important because of the volumetric crowding properties of PEG/NaCl. If this exact amount of DNA in the exact volume of buffer is not available, start with a pilot scaling down the volumes to a minimum of 25 µL because pipetting small amounts is difficult.

- 1. Let the beads come to room temperature and vortex beads to resuspend**
- 2. 10 µg DNA in 50 µL 0.1X TE + 25 µL beads**
 - a. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - b. Take the supernatant to Step 3
 - c. Wash the beads twice with 200 µL freshly prepared 80% EtOH
 - d. Air dry 5 min at room temp
 - e. Elute in 20 µL 0.1X TE. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - f. Expected: 1400 ng large MW DNA fragments
- 3. Supernatant + 10 µL beads**
 - a. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - b. Take the supernatant to Step 4
 - c. Wash the beads twice with 200 µL freshly prepared 80% EtOH
 - d. Air dry 5 min at room temp
 - e. Elute in 10 µL 0.1X TE. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - f. Expected: 400 ng medium MW DNA fragments
- 4. Supernatant + 35 µL beads**
 - a. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - b. Discard the supernatant
 - c. Wash the beads twice with 200 µL freshly prepared 80% EtOH
 - d. Air dry 5 min at room temp
 - e. Elute in 20 µL 0.1X TE. Pipette 10X, incubate 5 min at room temp, incubate 5 min at room temp on a magnetic rack.
 - f. Expected: 1 ug small MW DNA fragments
- 5. Check DNA molecular weight fragment distribution on a Bioanalyzer DNA 1000 ChIP (Figure 1)**

Sonication of large molecular weight fraction DNA

Most library prep protocols cannot accommodate large MW fragments because of a size selection step and a PCR amplification step, nor can the Illumina NextSeq 500/550 sequencer.

1. Sample prep

- a. Dilute large DNA to 130 μ L in 0.1X TE and transfer to a Covaris S220 microtube
- b. Sonication with Covaris S220
 - i. PP: 175 W
 - ii. DF: 10
 - iii. CB: 200
 - iv. 4-9°C
 - v. 5 min

2. Ethanol precipitation

- a. 130 μ L sample
- b. + 70 μ L H₂O
- c. + 17.4 μ L 2.5 M NaCl
- d. + 875 μ L 100% EtOH
- e. Precipitate at -20°C overnight
- f. Pellet DNA: 20,000 x g for 10 min at 4°C
- g. Wash with 500 μ L ice cold 80% EtOH
- h. Air dry pellet
- i. Resuspend in 19 μ L 0.1X TE.
- j. For the rest of the protocol, these large-sonicated-to-small size fragments continue to be referred to as “large” fragments

3. Check DNA molecular weight fragment distribution on a Bioanalyzer DNA 1000 ChIP (Figure 1)

Library Prep

With NEBNext Ultra DNA Library Prep Kit E7370

1. End Prep

- a. Blunts ends and adds an A overhang. This step is per the manufacturer's protocol
- b. Dilute DNA to 55.5 μL with H₂O
- c.

DNA	55.5 μL
End Prep Enzyme Mix	3 μL
10X End Repair Rxn Buff	6.5 μL

- d. Incubate
- e. 30 min @ 20°C
- f. 30 min @ 65°C
- g. Hold @ 4°C (but proceed asap)
- h. With heated lid

2. Adaptor ligation

- a. Adds a stem-loop adaptor with a T overhang. The stem loop prevents concatenation. This step is per the manufacturer's protocol.
- b.

End Prep Rxn	65 μL
Blunt/TA Ligase MM	15 μL
15 μM NEBNext Adaptor	2.5 μL
Ligation Enhancer	1 μL

- c. Incubate
 - i. 15 min @ 20°C
 - ii. 30 min @ 65°C
 - iii. 30°C heated lid because high temperatures can interfere with the ligation reaction
- d. Add 3 μL USER enzyme (cuts at U's in the adaptor)
- e. Incubate 15 min 37°C

3. Size selection of small fraction inserts with AMPure XP beads

- a. This step is adapted to the size of small fragments.
- b. Let the beads come to room temperature and vortex beads to resuspend
- c. 86.5 μL Adaptor Rxn + 13.5 μL H₂O + 55 μL beads
 - i. Pipette 10X
 - ii. incubate 5 min @ room temp

- iii. incubate 5 min @ room temp on a magnetic rack.
- iv. Transfer the supernatant to a new tube
- v. Supernatant + 25 μ L beads
- vi. Pipette 10X
- vii. incubate 5 min @ room temp
- viii. incubate 5 min @ room temp on a magnetic rack.

d. Transfer the supernatant to a new tube

- i. Wash the beads twice with 200 μ L freshly prepared 80% EtOH
- ii. Air dry 5 min at room temp
- iii. + 17 μ L 0.1X TE to elute
- iv. Pipette 10X
- v. incubate 5 min @ room temp
- vi. incubate 5 min @ room temp on a magnetic rack
- vii. Take 15 μ L eluate

4. Size selection of large-sonicated-to-small fraction inserts with AMPure XP beads

- a. This step is adapted to the size of large-sonicated-to-small fragments.
- b. Let the beads come to room temperature and vortex beads to resuspend
- c. 86.5 μ L Adaptor Rxn + 13.5 μ L H₂O + 35 μ L beads

- i. Pipette 10X
- ii. incubate 5 min @ room temp
- iii. incubate 5 min @ room temp on a magnetic rack.
- iv. Transfer the supernatant to a new tube

d. Supernatant + 45 μ L beads

- i. Pipette 10X
- ii. incubate 5 min @ room temp
- iii. incubate 5 min @ room temp on a magnetic rack.
- iv. Transfer the supernatant to a new tube
- v. Wash the beads twice with 200 μ L freshly prepared 80% EtOH
- vi. Air dry 5 min at room temp
- vii. + 17 μ L 0.1X TE to elute
- viii. Pipette 10X
- ix. incubate 5 min @ room temp
- x. incubate 5 min @ room temp on a magnetic rack
- xi. Take 15 μ L eluate

5. PCR enrichment of adaptor-ligated DNA

- a. Per the manufacturer's suggestions for the quantity input DNA, the small fraction has 8 cycles amplification and the large fraction has 7 cycles amplification.
- b. Vortex MM to resuspend

DNA from steps 3 and 4	15 μ L
NEBNext Q5 HotStart HiFi PCR MM	25 μ L
NEBNext Index Primer	5 μ L
Universal PCR Primer	5 μ L

c. PCR amplification and barcoding

Hot Start and Initial Denaturation	30 sec @ 98°C	1 cycle
Denature	10 sec @ 98°C	small: 8 cycles
Anneal/Extend	75 sec @ 65°C	large: 7 cycles
Final extension	5min @ 65°C	1 cycle
Hold	4°C	

- d. During the 4°C hold, run 1 μ L of samples a Bioanalyzer HS DNA ChIP to check for amplification. If no broad hump of amplification is seen around 300 bp, more cycles can be added.

6. PCR cleanup with AMPure XP beads

- a. This step is per the manufacturer's protocol.
- b. Let the beads come to room temperature and vortex beads to resuspend
- c. 50 μ L PCR rxn + 45 μ L beads
- d. Pipette 10X
- e. incubate 5 min @ room temp
- f. incubate 5 min @ room temp on a magnetic rack.
- g. Wash the beads twice with 200 μ L freshly prepared 80% EtOH
- h. Air dry 5 min at room temp
- i. + 22 μ L 0.1X TE to elute
- j. Pipette 10X
- k. incubate 5 min @ room temp
- l. incubate 5 min @ room temp on a magnetic rack
- m. Take 20 μ L eluate

7. Check library molecular weight fragment distribution

- a. on a Bioanalyzer DNA 1000 ChIP (Figure 1)
- b. Possible contaminating factors include:
 - i. 127 bp peak is contaminating adaptor
 - ii. < 85 bp peaks are contaminating primers
 - iii. > 1 kb fragments can be over-amplification
- c. If the area under the curve of these contaminations is small compared to the area under the curve of the library, then that is acceptable.
- d. Note the average fragment size of each library for the library quantification step.

Library quantification, dilution, and sequencing

With the NEBNext Library Quantification Kit (NEB E7630) and the NextSeq 500/550 sequencer.

1. Quantify library concentration

- a. per the manufacturer's protocol (1 in 30,000 and 1 in 60,000 dilutions are a good starting point to use for staying inside the standard curve).

2. Dilute libraries to 2 nM

3. Quantify the 2 nM libraries concentration per the manufacturer's protocol (1 in 10,000)

4. Pool libraries in a 1:1 ratio of molecules to get equal sequencing depth

- a. Example 1: if all the libraries really were diluted to 2 nM, that would be equal volumes of each library.
- b. Example 2: 4 diluted libraries are actually 1 nM, 2 nM, 2 nM, and 4 nM. To get equal representation, pool 10 μ L of the 1 nM library, 5 μ L each of the 2 nM libraries, and 2.5 μ L of the 4 nM library.

5. Quantify the 2 nM pooled library concentration per the manufacturer's protocol (1 in 10,000)

6. Denature and dilute pools

- a. per the current version of the NextSeq 500/550 protocol (or whatever sequencer is currently in use). With the NextSeq, the manufacturer recommends a final concentration of 1.8 pM. The Zaret Lab typically uses more than that without overclustering. 1.5X (2.7 pM) this amount often works well with srHC-seq.

7. Sequence single-end, 75 bp reads

- a. Paired-end might give more information about small fragment size locations, but we have not investigated this.

Alignment and processing of srHC-seq data

Since alignment and analysis software are often updated, the general step is described and particular parameters that are useful. See (Grindheim et al., 2019) for exact parameters.

1. Alignment and processing

- a. Align FASTQs
- b. Convert SAMs to BAMs
- c. Sort BAMs by location
- d. Convert BAMs to BEDs
- e. Get unique alignments to minimize PCR duplication artifacts.
- f. Convert to BAMs
- g. Sort BAMs by location
- h. Convert to BEDs

2. Track creation

- a. Get RPM
- b. Split BEDs into individual chromosomes for computational memory purposes
- c. Make RPM-normalized BGRs
- d. Perform \log_2 ((large fragment RPM-normalized BGR) / (small fragment RPM-normalized BGR))
 - i. \log_2 (large/small) translates into large-/heterochromatic-enriched regions being $y > 0$ and small-/euchromatic-enriched regions $y < 0$. This is easy to interpret with the human eye on a genome browser.
- e. Concatenate chromosomal BGRs
- f. Sort by location
- g. Average biological replicates
 - i. `bedtools unionbedg -filler "N/A"`
- h. Convert to BW

Domain calling

Call domains with the published Zaret Lab algorithm (Becker et al., 2017) plus additional steps. In short, for euchromatic domains, the Becker/Zaret algorithm will perform small/large fragments analysis, and for heterochromatic domains, the algorithm will perform large/small fragment analysis before further processing. See Figure 3 for a simplified schematic.

1. Compute large/small or small/large enrichments for windows and bins:

- a. Script: *computeEnrichments_sliding.mouse.basicChrsNoM.py* (written by Justin Becker, edited by Jessica Grindheim for the mouse genome)
- b. Settings: 10 kb windows, 2 kb bin/slide (“bin” and “slide” used somewhat interchangeably)
 - i. With these settings, the script splits the genome into 10 kb windows with a 2 kb slide between and calculates window and bin division scores for each window and bin, depending on the input files.
 1. The first few windows on chromosome 1 would be, 0 kb - 10 kb, 2 kb - 12 kb, 4 kb - 14 kb, etc.
 2. The first few bins on chromosome 1 would be, 0 kb - 2 kb, 2 kb - 4 kb, 4 kb - 6 kb, etc.
- c. Input files:
 - i. *BEDs of sorted, unique alignments for large and small srHC fractions*
 1. The script uses “ChIP” and “Input” nomenclature.
 2. For heterochromatic domains, “ChIP” = large fragments BED and “Input” = small fragments BED
 3. For euchromatic domains, “ChIP” = small fragments BED and “Input” = large fragments BED
- d. Output files:
 - i. *Windows division scores*: RPM-normalized pileup is calculated for large/small or small/large division in each window for every individual sample.
 - ii. *Bins division scores*: The script also splits the genome up into 2 kb bins. The bins for chromosome 1 would be 0 - 2 kb, 2 - 4 kb, 4 - 6 kb, etc. Each 2 kb bin in the genome is scored for RPM-normalized large/small or small/large division for each individual sample.

2. Get enriched windows with a fixed percentile cutoff

- a. Script: *getEnrichedBins_fixedPercentile.r* (written by Justin Becker)
- b. Settings: 60% cutoff
- c. Input Files: *Windows division scores*
- d. Output files:

- i. *Enriched Windows*: Windows scores for biological replicates are averaged. With a 60% cutoff, the highest 40% scoring windows are defined as enriched windows.

3. Merge neighboring windows

- a. Script: *mergeNeighbors.py* (written by Justin Becker)
- b. Input files: *Enriched Windows*
- c. Output files:
 - i. *Merged enriched windows*: enriched windows that are adjacent are merged into larger windows. For example, if chromosome 1, 2 kb – 12 kb, 4 kb – 14 kb, and 6 kb – 16 kb are all enriched windows, then they become one large window from 2 kb – 16 kb. For another example, if chromosome 1, 2 kb – 12 kb and 6 kb – 16 kb are enriched windows (but not the intervening 4 kb – 14 kb), then they remain two separate enriched windows.

4. Merge replicate bin scores

- a. Script: *mergeReplicateBinScores.r* (written by Justin Becker)
- b. Input files: *Bin division scores*
- c. Output files:
 - i. *Merged bin division scores*: Bins division scores were previously calculated per individual samples. This file has bin scores averaged across the biological group. For example, if three biological replicates have bins scores for chromosome 1, 0 kb – 2 kb of 2, 3, and 4 then the averaged bin score would be 3.

5. Prune merged enriched windows

- a. Script: *pruneDomains.py* (written by Justin Becker)
- b. Settings: 1X prune
- c. Input files: *Merged enriched windows*, *Merged bin division scores*
- d. Output files:
 - i. *Pruned windows*: The 60% cutoff corresponds to an large/small division score. This script takes merged enriched windows and prunes the 2kb edges off if the 2kb edge bins are less than 1X * the 60% large/small cutoff value.

6. Merge overlapping pruned windows

- a. Script: *mergeDomains.py* (written by Justin Becker)
- b. Input files: *Pruned windows*
- c. Output files:
 - i. *Enriched domains*: Adjacent or overlapping pruned windows are merged into “enriched domains”. For example, pruned window 1 is located on chromosome 1 from 4 kb – 20 kb, pruned window 2 is located on

chromosome 1 from 20 kb – 30 kb, pruned window 3 is located on chromosome 1 from 36 kb – 46 kb, and pruned window 4 is on chromosome 1 from 150 kb – 170 kb. Pruned windows 1, 2, and 3 are merged into 1 large domain from 4 kb – 46 kb. Pruned window 4 is separate and the intervening space from 46 kb to 150 kb is not an enriched domain.

This is where the Becker/Zaret algorithm is finished.

7. Determine final heterochromatic, intermediate, and euchromatic domains

- a. Use bedtools to get final domains.
- b. Heterochromatic domains = heterochromatic domains from the Becker/Zaret algorithm minus any regions overlapping with the euchromatic domains from the Becker/Zaret algorithm
- c. Euchromatic domains = Euchromatic domains from the Becker/Zaret algorithm minus any regions overlapping with the heterochromatic domains from the Becker/Zaret algorithm
- d. Intermediate domains = Regions called neither heterochromatic nor euchromatic with the Becker/Zaret algorithm, and regions that were called both heterochromatic and euchromatic with the Becker/Zaret algorithm.

V. Conclusions

srHC-seq is highly reproducible

srHC-seq is a highly reproducible technique by Spearman correlation of replicates and browser views (Chapter 2, Supplemental Figure 3, not reproduced here to avoid redundancy).

The reproducibility of srHC-seq in the hands of different users has not been previously assessed. Dario Nicetto, PhD also performed M2 *Wt* hepatocyte srHC-seq (Nicetto et al., 2019), so I took this opportunity to assess the similarity between his data set and the one generated here. Differences in sample prep include Dario using a Covaris S220 to sonicate crosslinked chromatin while I used a Diagenode Bioruptor UCD-200. Also, Dario used his small cell number/small DNA input protocol which does not isolate medium-size DNA fragments. Different library prep kits were also used. To compare

these protocols as uniformly as possible, both datasets were processed identically from the FASTQ stage on as described in (Grindheim et al., 2019). M2 *Wt* hepatocyte srHC-seq profiles from both datasets are highly similar at the positive controls sites for closed and open chromatin (*Zfp936* and the *Alb/Afp/Afm* locus) and at the dynamic *Cux2* locus (Figure 2).

We also employed a circular shuffling approach on heterochromatic and euchromatic domains to statistically assess similarity (code by Greg Donahue, Zaret Lab). Briefly, heterochromatic and euchromatic domains were called as described in (Grindheim et al., 2019) for both datasets. Grindheim domains were randomly displaced in a downstream direction by X bases on each chromosome, with domains or partial domains dropping off the end of the chromosome re-ordered to the beginning of the chromosome (called “shuffled domains”), and the number of Nicetto domains with 50% or higher sequence overlaps to the shuffled domains was counted. This experiment was repeated 1,000 times, and the p-value was estimated as the number of experiments in which the randomized overlap was greater than or equal to the observed non-randomized overlap, divided by 1,000. Estimated p-values for euchromatic and heterochromatic domains are less than 0.001. These results confirm statistically that Grindheim and Nicetto M2 srHC domains are significantly related.

Robust heterochromatic and euchromatic gene calling with srHC-seq

With increasingly lax enriched window cutoffs for domain calling, an increasing portion of the genome is called as heterochromatic or euchromatic, until the 50% cutoff, where heterochromatic +/-euchromatic + double positive regions that are reclassified as intermediate takeover (Figure 3AB). Despite the increasing portions of the genome which are called as heterochromatic or euchromatic with 10% to 40% enriched window cutoffs,

there is relatively mild change in the number of genes or promoters called as heterochromatic, euchromatic, or intermediate (Figure 3C), reflecting the robustness of this technique for calling gene or promoter compaction states. Also reflecting the robustness of this technique, euchromatic genes have high expression and heterochromatic genes have relatively low expression (Figure 3D). Curiously, genes called as having heterochromatic gene bodies or promoters at P14 have higher levels of expression than that genes called as having heterochromatic gene bodies or promoters at M2, suggesting cellular heterogeneity or incomplete “lockdown” of the genome at the relatively immature P14 timepoint.

srHC-seq reveals a preference for H3K27me3 in Type A compartment heterochromatin

Chromosomal organization is cell-type specific and related to transcriptional control (Dekker and Mirny, 2016; Gorkin et al., 2014; de Laat and Duboule, 2013). The highest level of topological organization segregates the genome into megabase scale compartments termed Type A and B, which are associated with active and inactive chromatin, respectively, with the latter enriched at the nuclear lamina (Boettiger et al., 2016; Lieberman-Aiden et al., 2009; Lin et al., 2012; Rao et al., 2014; Vieux-Rochas et al., 2015; Wang et al., 2016). We found a high concordance between srHC-seq data and adult liver Type AB compartments, with Type A compartments being more euchromatic, B compartments being more heterochromatic (Schwarzer et al., 2017), and also high concordance with lamin-associated domains (Peric-Hupkes et al., 2010) (Figure 4AB). Specifically, nearly three quarters of Type A compartments are called as euchromatic by srHC-seq, and nearly three quarters of Type B is called as heterochromatic by srHC-seq

(Figure 4C). Thus, srHC-seq reflects compaction predictions based on AB compartment calls.

Interestingly, a quarter of Type A compartments are called as heterochromatic by srHC-seq, and nearly a quarter of B compartments are called euchromatic (Figure 4C). Supporting this, gene expression more closely tracks with srHC-seq heterochromatin and euchromatin calls at gene bodies and promoters than with AB compartment localization (Figure 4D). These results confirm the findings of others (Lieberman-Aiden et al., 2009; Rao et al., 2014), that AB compartments are generally transcriptionally active/euchromatic and inactive/heterochromatic, respectively, but also show that regions within a compartment do not always conform to the general euchromatic or heterochromatic nature of the compartment. Interestingly, 70% of heterochromatin in Type A overlapped with M2 H3K27me3 domains, while only 51% of heterochromatin in B (Figure 4B), suggesting that A heterochromatin compartments are more dependent on Polycomb-related compaction. With the srHC-seq assay we can accurately identify the heterochromatic/euchromatic state of annotated genomic features, such as promoters and genes, which in turn predicts the transcriptional states of genes more accurately than AB compartment localization (Figure 4D).

VI. Figures

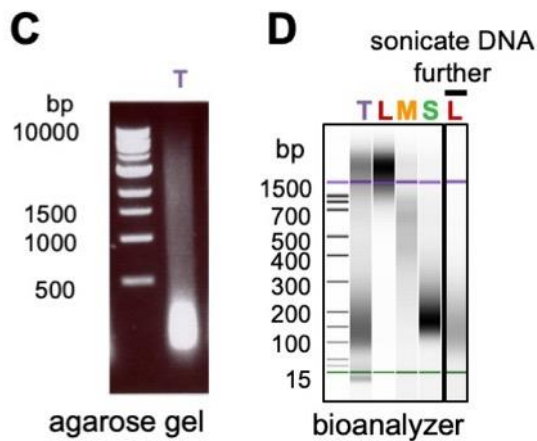
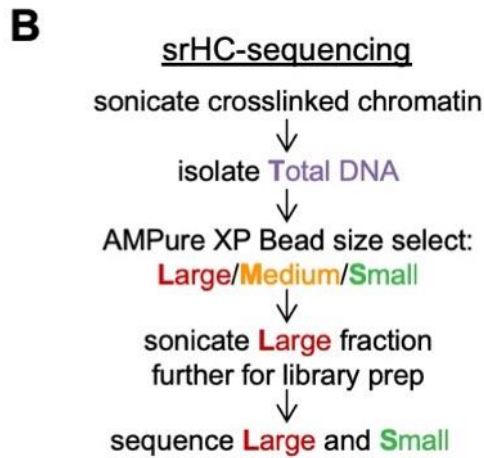
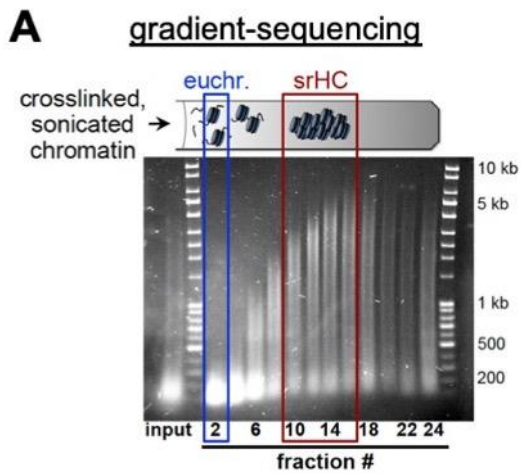


Figure 1: Fractionation of high and low molecular weight DNA for srHC-seq

A) Gradient-seq schematic with DNA from different fractions run on an agarose gel. Adapted from (Becker et al., 2017). Note that low MW DNA fragments correspond to sonication-sensitive euchromatin and high MW DNA fragments correspond to sonication-resistant heterochromatin.

B) High cell number srHC-seq overview.

C) Total DNA from crosslinked, sonicated hepatocyte chromatin is shown on a 1% agarose gel for comparison to

D) Total DNA on a Bioanalyzer DNA 1000 chip. Representative profiles for Large, Medium, or Small DNA fractions and Large DNA that has been sonicated to a smaller size for library prep considerations.

B-D Adapted from (Grindheim et al., 2019)

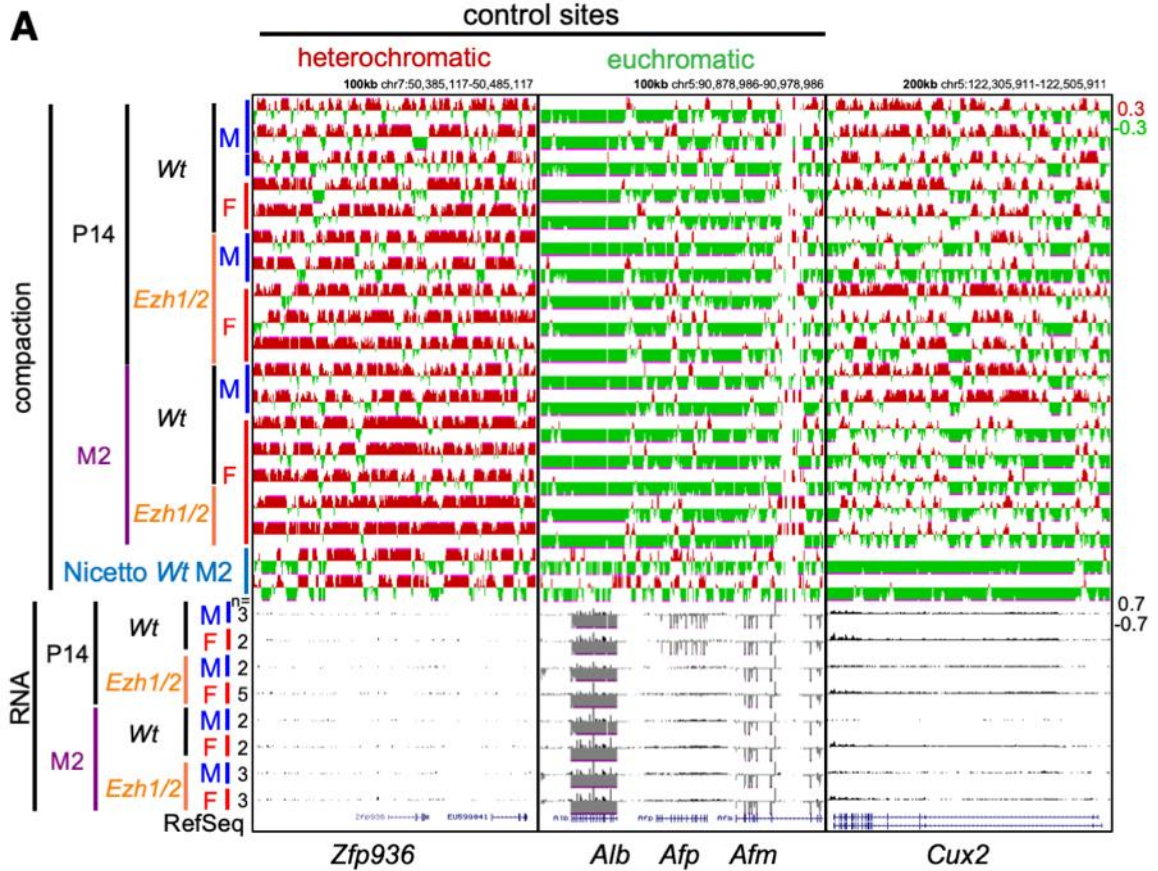


Figure 2: High reproducibility of srHC-seq data

A) srHC-seq (top: Grindheim, bottom in light blue: Nicetto) and RNA signal for positive controls sites for heterochromatin, *Zfp936*, euchromatin, *Alb* locus, or a dynamic gene, *Cux2*.

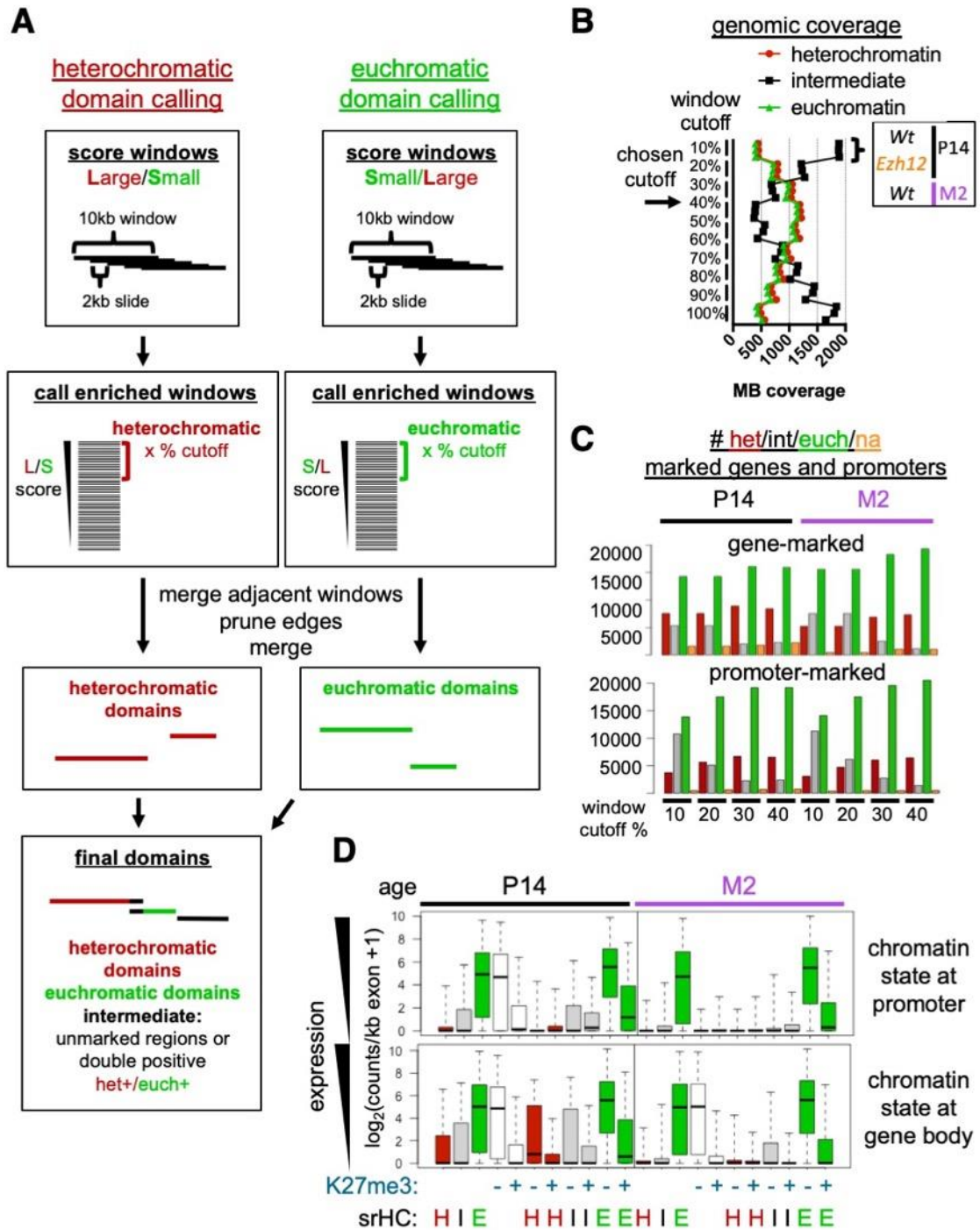


Figure 3: srHC-seq heterochromatin and euchromatin domain, promoter, and gene body calling

- A)** Compaction domain calling overview.
- B)** Portion of the genome called as heterochromatic, euchromatic, or intermediate with variable enriched window cutoffs. Heterochromatic and euchromatic window cutoffs are the same.
- C)** # of genes or promoters called as heterochromatic, intermediate, euchromatic, or NA with variable enriched window cutoffs. Marked genes $\geq 50\%$ of the gene body overlapping with a domain. Marked promoters $\geq 75\%$ of the gene body overlapping with a domain. NA refers to genes that do not meet the definition of a marked gene.
- D)** *Wt* gene expression in P14 (left) and M2 (right) hepatocytes. Genes are divided into srHC **H**eterochromatic, **I**ntermediate, or **E**uchromatic promoters ($\geq 75\%$ of TSS - 1kb/+500bp marked by a domain) or gene bodies ($\geq 50\%$ of gene body marked by a domain) (top 40% enriched window cutoff). Genes are further divided into H3K27me3-marked at the promoter ($\geq 50\%$ of TSS -1kb/+500bp marked by an H3K27me3 domain) or at the gene body ($\geq 25\%$ of gene body marked by a domain) (top 30% enriched window cutoff). Whiskers indicate 5-95 percentiles.

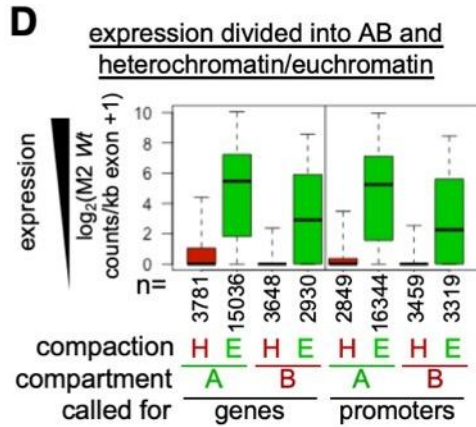
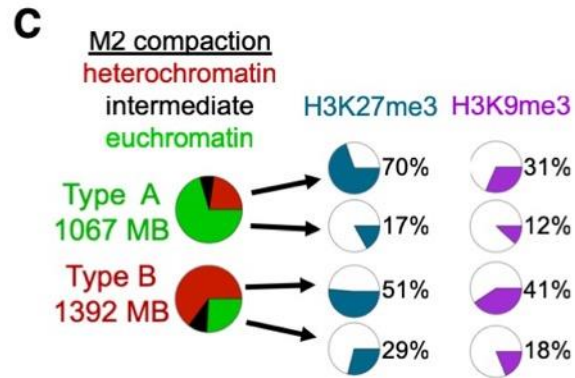
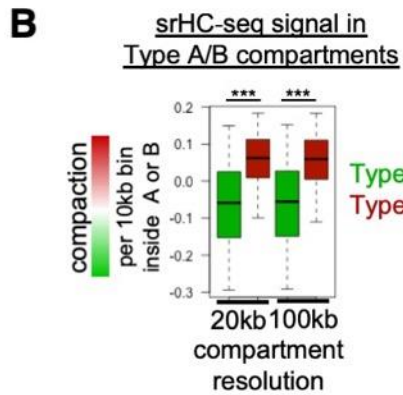
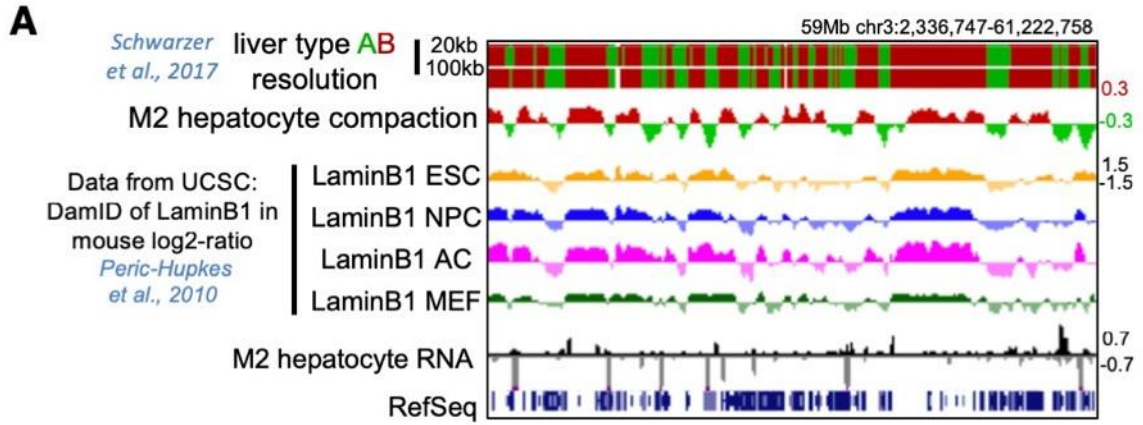


Figure 4: AB compartments

- A)** Type A (green) and Type B (red) compartments in adult mouse liver at 20 kb or 100 kb resolution (Schwarzer et al., 2017), srHC-seq signal in M2 *Wt* hepatocytes, DamID of LaminB1 in mouse log2-ratio (Peric-Hupkes et al., 2010), RNA-seq signal M2 *Wt* hepatocytes, and RefSeq genes.
- B)** srHC-seq signal in non-overlapping 10 kb windows inside of Type A or B compartments (100 kb resolution). Wilcoxon rank sum test with continuity correction, all p-values < 2.2e-16.
- C)** Left: A and B compartments (100 kb resolution) divided into heterochromatin, intermediate, or euchromatic domains. Right: Heterochromatin and euchromatin in A or B compartments further divided into H3K27me3 or H3K9me3 domains.
- D)** Female expression for genes and promoters that have been divided into A/B compartment (100 kb resolution) location and further divided into heterochromatin/euchromatin calls.

VII. Bioinformatics methods

Calling genes in Type AB compartments

The Spitz Lab provided adult mouse liver AB compartment data (Schwarzer et al., 2017).

If a gene body or promoter overlapped at least 90% with a Type A compartment, it was called as in Type A. If a gene or promoter overlapped at least 90% with a Type B compartment, it was called as in Type B. Given the large size of Type A B compartments, this captured the majority of genes. For example, with compartments at a 20kb or 100kb resolution, approximately 5% and 2% of genes were unassigned, respectively

srHC-seq scores in Type AB compartments

To determine srHC-seq scores in Type AB compartments, the genome was first divided into 10 kb windows without overlap. 10 kb windows completely inside of a Type A compartment are “Type A windows” and the same for “Type B windows”. Quantile normalized srHC-seq scores for these windows were used. To test for differences in srHC-seq scores between A and B compartments, the Wilcoxon rank sum test with continuity correction was used.

VIII. References

- Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14926–14931.
- Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* 68, 1023-1037.e15.
- Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.
- DeAngelis, M.M., Wang, D.G., and Hawkins, T.L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* 23, 4742–4743.
- Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. *Cell* 164, 1110–1121.
- Frenster, J.H., Allfrey, V.G., and Mirsky, A.E. (1963). REPRESSED AND ACTIVE CHROMATIN ISOLATED FROM INTERPHASE LYMPHOCYTES. *Proc. Natl. Acad. Sci. U.S.A.* 50, 1026–1032.
- Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14, 762–775.
- Grindheim, J.M., Nicetto, D., Donahue, G., and Zaret, K.S. (2019). PRC2 proteins EZH1 and EZH2 Regulate Timing of Postnatal Hepatocyte Maturation and Fibrosis by Repressing Gene Expression at Promoter Regions in Euchromatin in Mice. *Gastroenterology*.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- 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.
- Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13, 1196–1204.
- Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*.
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* 38, 603–613.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.

- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56.
- Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4672–4677.
- Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602.

Appendix C: Cut and Run, a method for mapping proteins on chromatin

Cut and Run is a method developed by the laboratory of Steven Henikoff for identifying protein-DNA interaction. This appendix describes how I adapted the Elife 2017 version of the Cut and Run protocol (Skene and Henikoff, 2017) to work on fresh hepatic nuclei. In the Cut and Run protocol, a protein A-MNase fusion protein is recruited to chromatin by an antibody against a protein of interest, where the MNase portion of the fusion protein can then cleave DNA. In contrast to ChIP, this Cut and Run protocol is performed on non-crosslinked, unsonicated nuclei. The protocol should also work on non-hepatic nuclei as well.

I. Objective

To determine the genomic location of target, such as transcription factors or histone modifications, without crosslinking or sonication.

II. Background

The genomic location of histone modifications or transcription factor binding is key to understanding transcriptional regulation or other nuclear processes. Therefore, genome-wide chromatin profiles are commonly generated using a technique termed chromatin immunoprecipitation (ChIP), first described in 1985 (Solomon and Varshavsky, 1985). In ChIP protocols, cells are crosslinked with formaldehyde to stabilize protein-DNA interactions. Crosslinking is followed by sonication, which is needed to break

chromatin/DNA down into soluble and mappable smaller fragments. Immunoprecipitation enriches for fragments of interest and DNA is extracted for analysis by qPCR, microarrays, or more recently, high-throughput sequencing. While ChIP has revolutionized the study of chromatin biology and gene regulation, there are some drawbacks, including masked epitopes and false positives from crosslinking (Baranello et al., 2016; Jain et al., 2015; Meyer and Liu, 2014; Park et al., 2013; Skene and Henikoff, 2017; Teytelman et al., 2013). One alternative is native ChIP, or ChIP without crosslinking in ionic conditions that leave electrostatic interactions intact. Native ChIP required the use enzymatic methods, such as MNase treatment, to cleave DNA to smaller, mappable fragments (Kasinathan et al., 2014). Native ChIP can avoid epitope masking, epitope degradation, and can be used on smaller numbers of cells, but can have its own biases related MNase overdigestion or cleavage preferences (Axel, 1975; Iwafuchi-Doi et al., 2016; Rhee et al., 2014; Weiner et al., 2010; Xi et al., 2011).

To circumvent some of these biases, the Henikoff Lab developed Cleavage Under Targets and Release Using Nuclease, or Cut and Run (Skene and Henikoff, 2017). Cut and Run does not require crosslinking or sonication, thus avoiding some epitope masking and false positive biases. The version of Cut and Run (Figure 1) described here involves incubation of nuclei with an antibody against the target of interest, then incubation with a protein A-MNase fusion protein, which is directed to bind the antibody by the bacterially-derived protein A. The cleavage reaction is initiated by the addition of calcium and stopped with the addition of cation chelating agents. DNA fragments released by MNase cleavage of DNA around the DNA-antigen-antibody-pA-MNase complex are released and can be isolated by incubation, nuclei pelleting, and isolation of cleaved DNA fragments from the supernatant (Figure 1). The Henikoff Lab has used their protocol to determine chromatin profiles for both TF and histone modifications and claim that this targeted digestion method

reduces background as compared to standard ChIP methods, thus reducing the cell number and sequencing depth required. In another version of their protocol, robotic automation is possible by the combination of low cell numbers and magnetic beads.

III. Key Reagents

4-6 million nuclei per reaction

Here isolated from fresh hepatocytes as described in Appendix A. We did not test Cut and Run on frozen cells or nuclei, because freezing may cause cellular lysis or DNA breakage and background signal.

3-5 μ g antibody per reaction

Likely more antibody should be used in the cases of histone modifications with broad genomic coverage.

Also do a no antibody control to assess endogenous nuclease and background pA-MNase cleavage.

pA-MNase enzyme

As of 2019, provided directly from the Henikoff Lab.

Cut and Run Buffer 1 (CR1)

20 mM HEPES, pH 7.5

150 mM NaCl

2 mM EDTA

Add fresh:

0.5 mM spermidine (note: short half-life)

0.1% BSA

Protease inhibitors, EDTA-free Roche, # 11836170001

Sterile filter.

Cut and Run Buffer 2 (CR2)

20 mM HEPES, pH 7.5

150 mM NaCl

Add fresh:

0.5 mM spermidine (note: short half-life)

0.1% BSA

Protease inhibitors, EDTA-free Roche, # 11836170001. (note: EDTA-free is very important because it chelates calcium ions, which are needed for MNase cleavage)

Sterile filter.

Cut and Run Buffer 3 (CR3)

CR2 with 2 mM CaCl₂

Calcium ions allow MNase to cleave DNA.

Stop Solution

Cation chelating solution to stop MNase cleavage. Also contains spike-in DNA for sequencing normalization.

150 mM EDTA

300 mM EGTA

75 pg/mL mononucleosomal-sized, heterologous spike-in yeast DNA

0.5 mL Eppendorf tubes

The small volume reduces sloshing and possible nuclear lysis

TES buffer

10 mM Tris-HCl, pH 8
10 mM EDTA
1% SDS

TE buffer

10 mM Tris-HCl, pH 8
10 mM EDTA

RNase A

DNase-free. Temperature quoted in the protocol below is based on the manufacturer's recommendations.

Proteinase K

DNase-free. Temperature quoted in the protocol below is based on the manufacturer's recommendations.

Glycogen (Roche 10 901 393 001)

NEBNext Ultra DNA Library Prep Kit (NEB 7370)

Likely many other commercial kits will also work.

NEBNext Multiplex Oligos for Illumina (Index Primers Set 1-4, NEB E7335S, E7500S, E7710S, E7730S)

Agilent Bioanalyzer

For fragment size analysis

NEBNext Library Quantification Kit (NEB E7630)

For library quantification prior to sequencing. Library quantification is done with information about average library size (from a Bioanalyzer run) and qPCR with this or other kits against standards of known size and concentration. The NEBNext library quantification kit is compatible with the NEBNext Ultra Library prep kit (the qPCR primers will amplify the libraries).

IV. Protocol

Cut and Run Reaction

The hepatocyte nuclei used here were isolated as described in Appendix A.

1. Wash nuclei

- a. Pellet nuclei 200 x g/4 min/4°C in a swinging-bucket centrifuge.
 - i. Faster speeds make nuclei clump together and it becomes difficult to resuspend them.
- b. Resuspend in 1.5 mL ice cold CR1. Mix by gentle pipetting.
- c. Save protein for Western, if desired.
 - i. 100 µL TES per 10E6 nuclei
- d. Incubate for 5 min on ice
- e. Pellet nuclei
- f. Resuspend in about 1.5 mL CR2
- g. Count nuclei with a hemocytometer.
- h. Aliquot 4-6 million nuclei per reaction (the same number of nuclei in each reaction) in 0.5 mL tubes

2. Primary antibody incubation

- a. Pellet nuclei
- b. Resuspend in 500 µL CR2.
- c. +3-5 µg Ab
- d. Rotate/2 hours/4°C
- e. Also do a no antibody control!

3. 3 washes in 500 µL CR2

4. pA-MNase incubation

- a. Pellet nuclei
- b. Resuspend in 300 µL CR2 with 180 ng pA-MNase
 - i. 180 ng corresponds to a 1/1000 dilution pA-MNase of the 360 ng/µL batch of enzyme. Per Pete Skene's recommendations.
- c. Rotate/1 hour/4°C

5. 3 washes in 500 µL CR2

6. Digestion reaction

- a. Place tubes on wet ice. Heat can cause over digestion and off-target digestion.
- b. Pellet nuclei

- c. Resuspend in 300 μ L CR3, including the no Ab control.
- d. Mix by inverting, incubate 5-15 min (determine incubation time in a pilot experiment)
- e. Stop digestion by adding 20 μ L Stop Solution
- f. Mix by inverting

7. Retain desired fraction of cutting reaction

- a. Leach fragments out of nuclei: rotate/1 hour/4°C
- b. Pellet nuclei
- c. Take 300 μ L of leach supernatant.
 - i. Taking too much of the supernatant will also take nuclei, which will lead to background signal.
- d. + 3 μ L 10% SDS
- e. + 5 μ L 5M NaCl
- f. Can freeze here.

8. RNA and protein degradation

- a. + 2 μ L 10 mg/mL RNase A
- b. Incubate for 2 hours/37°C/shaking
- c. + 4 μ L 20 mg/mL Proteinase K
- d. Incubate for 2 hours/55°C/shaking

9. Phenol:chloroform:isoamyl alcohol fractionation of DNA

- a. + 400 μ L PCI
- b. Vortex thoroughly
- c. Spin $\geq 12,000 \times g$ for 15 min at room temperature
- d. Take 350 μ L of aqueous phase

10. Ethanol precipitation

- a. 350 μ L of aqueous phase + 1.5 μ L 10 glycogen + 32 μ L 2.5 M NaCl + 1.1 mL 100% EtOH
- b. Precipitate at -20°C overnight
- c. Pellet DNA: 20,000 $\times g$ for 10 min at 4°C
- d. Wash with 500 μ L 80% ice cold EtOH
- e. Air dry pellet
- f. Resuspend in 15 μ L 0.1X TE

Library generation, quantification, dilution, and sequencing

1. Generate libraries with NEBNext Ultra DNA Library Prep Kit E7370

2. Check library molecular weight fragment distribution

- a. on a Bioanalyzer DNA High sensitivity ChIP
- b. Possible contaminating factors include:
 - i. 127 bp peak is contaminating adaptor
 - ii. < 85 bp peaks are contaminating primers
 - iii. > 1 kb fragments can be over-amplification
- c. If the area under the curve of these contaminations is small compared to the area under the curve of the library, then that is acceptable.
- d. Note the average fragment size of each library for the library quantification step.

3. Quantify library concentration

- a. With the NEBNext Library Quantification Kit (NEB E7630)
- b. A 1/30,000 dilution is a good starting point to use for staying inside the standard curve

4. Dilute libraries to 2 nM

5. Quantify the 2 nM libraries concentration per the manufacturer's protocol (1 in 10,000)

6. Pool libraries in a 1:1 ratio of molecules to get equal sequencing depth

- a. Example 1: if all the libraries really were diluted to 2 nM, that would be equal volumes of each library.
- b. Example 2: 4 diluted libraries are actually 1 nM, 2 nM, 2 nM, and 4 nM. To get equal representation, pool 10 μ L of the 1 nM library, 5 μ L each of the 2 nM libraries, and 2.5 μ L of the 4 nM library.
- c. 5-10 million reads are recommended for proteins with "peaky" profiles, like transcription factors.

7. Quantify the 2 nM pooled library concentration per the manufacturer's protocol (1 in 10,000)

8. Denature and dilute pools

- a. Per the current version of the NextSeq 500/550 protocol (or whatever sequencer is currently in use). With the NextSeq, the manufacturer recommends a final concentration of 1.8 pM. The Zaret Lab typically uses more than that without over-clustering.

9. Sequence paired-end, 37 bp/37 bp reads

- a. NextSeq 500/550 sequencer.

Alignment and processing of Cut and Run data

As softwares are frequently updated, the steps are described in short. Specifically, data was aligned and processed as described for H3K27me3 ChIP data in Chapter 2 (Grindheim et al., 2019).

1. Alignment and processing

- a. Align FASTQs to the target genome, here mouse,
- b. Optional: Align FASTQs to the spike-in control genome, here *saccharomyces cerevisiae*
- c. Convert SAMs to BAMs
- d. Sort BAMs by location
- e. Convert BAMs to BEDs
- f. Filter for unique alignments to avoid PCR duplicates
- g. Convert to BAMs
- h. Sort BAMs by location
- i. Convert to BEDs

2. Track creation

- a. Get RPM
- b. Split BEDs into individual chromosomes for computational memory purposes
- c. Make RPM-normalized BGRs
- d. For traditional, Henikoff RPM-normalized tracks, with spike-in normalization
 - i. Concatenate chromosomal BGRs
 - ii. Sort by location
 - iii. Average biological replicates
 1. `bedtools unionbedg -filler "N/A"`
 - iv. Convert to BW
- e. For Cut and Run minus no Ab control (recommended to control for endogenous nuclease and background pA-MNase cleavage).
 - i. Perform (sample RPM-normalized BGR) - (no Ab control RPM-normalized BGR)
 - ii. Concatenate chromosomal BGRs
 - iii. Sort by location
 - iv. Average biological replicates
 1. `bedtools unionbedg -filler "N/A"`
 - v. Convert to BW

V. Conclusions

Successful transcription factor and histone modification Cut and Run reactions

We performed Cut and Run on a no antibody control sample, H3K4me3, H3K27ac, FOXA2, and several Polycomb group proteins (Figure 2). As positive controls binding profiles in hepatocytes, we compared Cut and Run results to published ENCODE adult liver H3K4me3 and H3K27ac, adult FOXA2 hepatocyte ChIP-exo (Iwafuchi-Doi et al., 2016), and M2 H3K27me3 hepatocyte ChIP (Grindheim et al., 2019). As compared to published ENCODE data, there is robust Cut and Run H3K4me3 and H3K27ac signal over no antibody control background signal. Similarly, FOXA2 signal was robust at FOXA2 binding sites as assessed by adult hepatocyte FOXA2 ChIP-exo. In the case of Cut and Run for the Polycomb components, there is signal at open regions, such as the promoters of expressed genes or FOXA2 binding sites, that disappears when the Cut and Run no antibody control signal is subtracted indicating that Cut and Run was not successful with these antibodies.

While we were successful at performing Cut and Run for several targets, we found that there was significant background signal in the no antibody control samples open sites. Given that the background cutting of pA-MNase in hepatocytes led to profiles that look remarkably similar to FOXA2 binding profiles, a FOXA2 Cut and Run reaction that did not work could easily be mistaken as successful. Consequently, conclusions might be drawn that FOXA2 binds to promoters where it actually does not because open promoters are also sites of pA-MNase background signal. These results contrast with other groups that observe negligible background. We note that other groups are performing Cut and Run

on cell lines, while we are using native cells. It is possible the increased background signal in our studies stems from a difference between cell lines and native cells, such as a difference in the quantity of endogenous nucleases. Hepatocytes may be particularly prone to background cleavage as they contain unusually high quantities of cytoplasmic proteins, with 30-50 times more cytoplasm than nucleus. While these results show Cut and Run to be a promising new technique, they also underline the differences between native tissues and cell lines, and emphasize the importance of controls.

VI. Figures

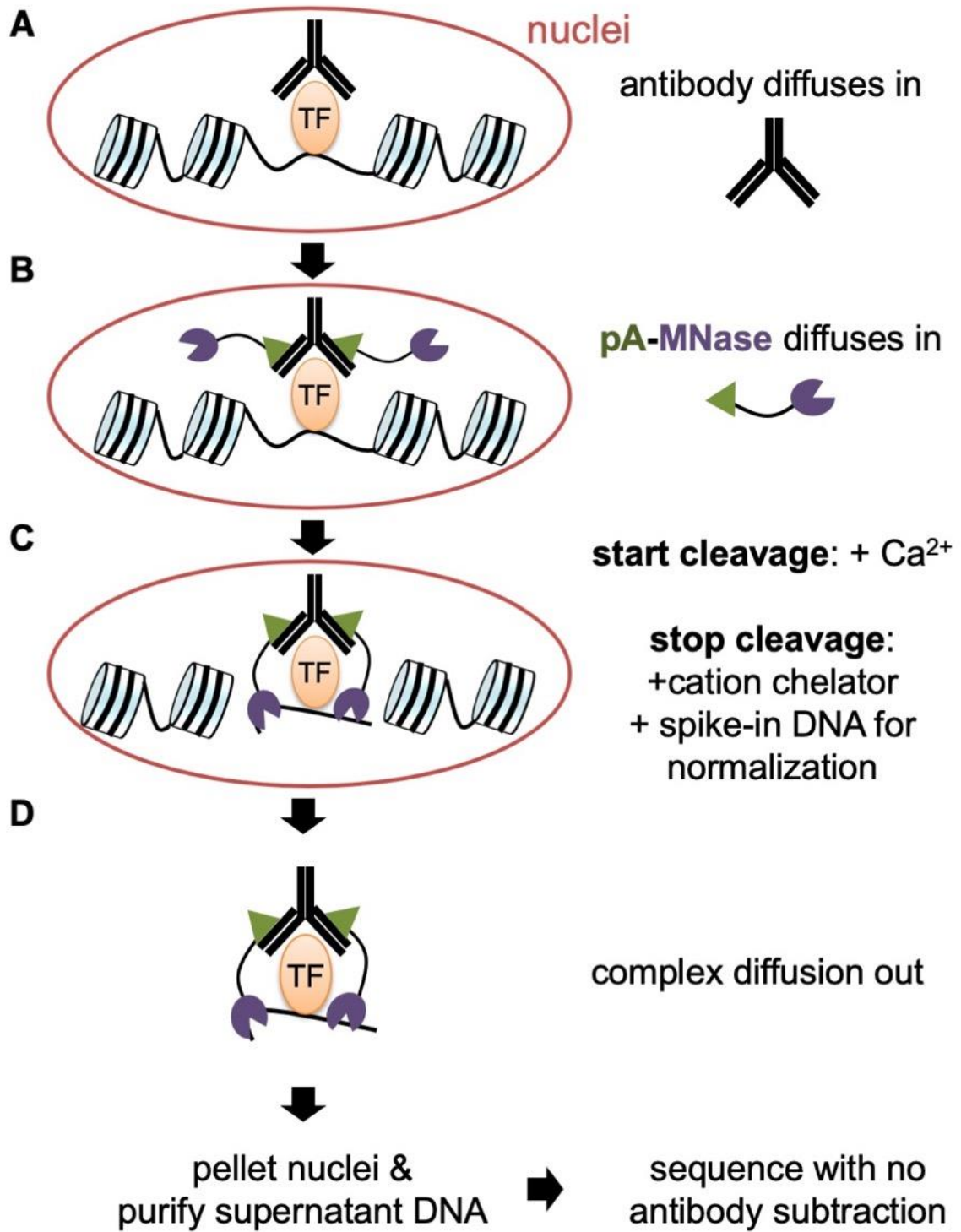


Figure 1: Cut and Run schematic

- A)** Hepatic nuclei are incubated with an antibody against a protein of interest.
- B)** Then, incubate with the pA-MNase fusion protein. The protein A portion of pA-MNase binds to the antibody.
- C)** MNase cleavage is started by the addition of calcium cations. Cleavage is stopped after a set time by the addition cation chelator. In the stop solution is also spike-in DNA for sequencing normalization purposes.
- D)** Nuclei are incubated on a rotator, allowing the cut fragments of DNA to leach out of nuclei. Nuclei are pelleted and DNA isolated from the supernatant.

Figure 2: Successful Cut and Run

All Cut and Run samples have been filtered for unique alignments, were performed on the same adult female, and are shown as RPM-normalized tracks with and without no antibody subtraction.

A) From top to bottom:

No antibody control Cut and Run (black).

H3K4me3: ENCODE adult liver ChIP, Cut and Run (Abcam 8580 GR124346-1) and no antibody control subtraction (brown).

H3K27ac: ENCODE adult liver ChIP, Cut and Run (Abcam 4729 GR3187598-1) and no antibody control subtraction (cyan).

FOXA2: FOXA2 hepatocyte (Iwafuchi-Doi et al., 2016) ChIP-exo (aligned by JMG), Cut and Run (Invitrogen 710730, Invitrogen 720061, Millipore 07-633) and no antibody control subtraction (green).

Polycomb: M2 hepatocyte H3K27me3 (Grindheim et al., 2019), Cut and Run (BMI1: Update clone 6 05-637, CBX2: Bethyl A302-524A, EED: Santa Cruz 30812 Lot G2909) and no antibody control subtraction (blue).

M2 hepatocyte RNA-seq signal from (Grindheim et al., 2019) (black).

VII. References

- Axel, R. (1975). Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. *Biochemistry* 14, 2921–2925.
- Baranello, L., Kouzine, F., Sanford, S., and Levens, D. (2016). ChIP bias as a function of cross-linking time. *Chromosome Res.* 24, 175–181.
- Grindheim, J.M., Nicetto, D., Donahue, G., and Zaret, K.S. (2019). PRC2 proteins EZH1 and EZH2 Regulate Timing of Postnatal Hepatocyte Maturation and Fibrosis by Repressing Gene Expression at Promoter Regions in Euchromatin in Mice. *Gastroenterology*.
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol. Cell* 62, 79–91.
- Jain, D., Baldi, S., Zabel, A., Straub, T., and Becker, P.B. (2015). Active promoters give rise to false positive “Phantom Peaks” in ChIP-seq experiments. *Nucleic Acids Res.* 43, 6959–6968.
- Kasinathan, S., Orsi, G.A., Zentner, G.E., Ahmad, K., and Henikoff, S. (2014). High-resolution mapping of transcription factor binding sites on native chromatin. *Nat. Methods* 11, 203–209.
- Meyer, C.A., and Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* 15, 709–721.
- Park, D., Lee, Y., Bhupindersingh, G., and Iyer, V.R. (2013). Widespread misinterpretable ChIP-seq bias in yeast. *PLoS ONE* 8, e83506.
- Rhee, H.S., Bataille, A.R., Zhang, L., and Pugh, B.F. (2014). Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* 159, 1377–1388.
- Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6.
- Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6470–6474.
- Teytelman, L., Thurtle, D.M., Rine, J., and van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18602–18607.
- Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.* 20, 90–100.
- Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Res.* 21, 718–724.

Appendix D: Chromatin-associated RNA-seq, a method for isolating RNAs bound to chromatin

This appendix provides a description of a collaboration between a graduate student in the Smale Lab, Miguel Edwards, PhD, and I that resulted in the following protocol for chromatin-associated RNA-seq on fresh hepatocytes isolated from 2-month-old mice.

I. Objective

To isolate chromatin-associated RNA mouse hepatocytes. Chromatin-associated RNA isolation has been described (Pandya-Jones and Black, 2009; Wuarin and Schibler, 1994) and expanded for genome-wide studies via high-throughput sequencing technologies (Bhatt et al., 2012). Multiple classes of RNA are captured by chromatin-associated RNA, including nascent RNA and RNA which function in chromatin.

II. Protocol

Fractionation of chromatin-associated RNA

The hepatocyte nuclei used here were isolated as described in Appendix A. Use wide bore tips when handling nuclei.

1. Resuspend the nuclei

- a. 100 μ L glycerol buffer per 10E6 nuclei
- b. Incubate 2 min on ice
- c. Pellet 5 min/15,000 rpm/4°C
- d. Use a microscope and trypan blue staining to check for nuclear lysis.

2. Transfer sup (nucleoplasmic lysate) to new tube.

- a. 10% of nucleoplasmic lysate for Western analysis
- b. 90% of nucleoplasmic lysate in TRIzol

3. Resuspend chromatin pellet in 25 μ L PBS/10E6 nuclei

- a. 10% of chromatin lysate for Western analysis
- b. 90% of chromatin lysate in TRIzol

Assessment of fractionation efficacy by Western blot

To assess the success of fractionation, nucleoplasmic and cytoplasmic lysates were separated on a 12% gel and Western blotted for cytoplasmic, nucleoplasmic, and chromatin markers by Miguel Edwards. Given the limited quantity of sample he normally works with, he blots for all three markers on the same blot without extensive work to achieve equal loading. Therefore, a successful fractionation has high chromatin marker staining without nucleoplasmic or cytoplasmic staining.

Cytoplasmic marker: α -tubulin (Calbiochem CP-06, mouse 1/1000). ~55 kDa

Nucleoplasmic marker: α -SNRP70 (Black Lab antibody. ~ 70 kDa).

Chromatin marker: H3 (Millipore 06-755, 1/2000). ~17 kDa

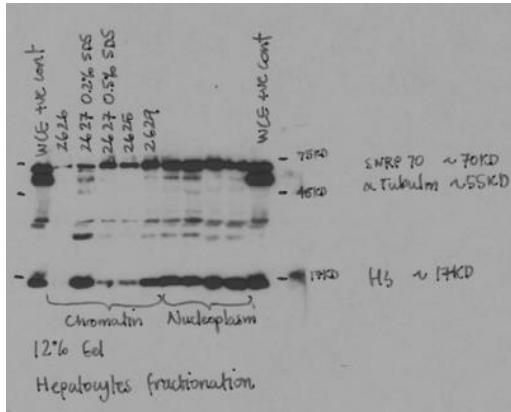
From a pilot fractionation with variable concentrations of SDS, Miguel Edwards and I concluded determined conditions for chromatin fractions in hepatocytes (Figure 1a). These conditions were used on further samples and fractionation confirmed (Figure 1b).

Library prep and sequencing

Isolate RNA from TRizol, deplete ribosomal RNA, and prepare libraries with Illumina's TruSeq RNA Library Preparation Kit v2 with strand-specific reads.

III. Figures

A



B

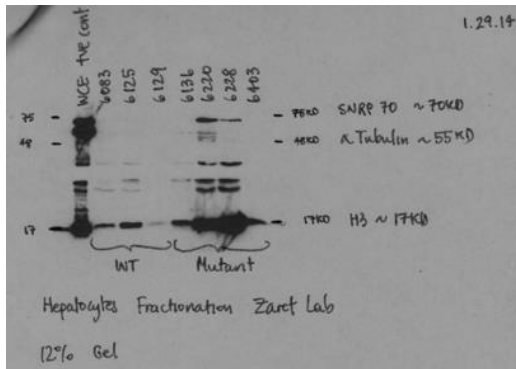


Figure 1: Characterization of chromatin-associated RNA-seq samples

- A)** Western blot results for pilot fractionation of nucleoplasm, chromatin, and a whole cell extract control. Cytoplasmic, nucleoplasmic, and chromatin markers on chromatin and nucleoplasmic lysates.
- B)** Western blot results for fractionation of nucleoplasm, chromatin, and a whole cell extract control. Cytoplasmic, nucleoplasmic, and chromatin markers on chromatin and nucleoplasmic lysates.

IV. References

- Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279–290.
- Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15, 1896–1908.
- Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* 14, 7219–7225.

Appendix E: qPCR primer design and testing

This appendix provides a detailed explanation of how to design and test qPCR primers for reverse transcription qPCR (RTqPCR) and ChIP-qPCR. It also provides a list of validated RTqPCR primer sets for useful targets to the Zaret Lab, primarily involving liver and pancreatic targets.

I. Objective

To describe qPCR primer design best practices and provide a list of useful primers.

II. Background

qPCR with a dsDNA intercalating fluorescent dye, such as the widely used SYBR Green assay, is widely used for both RTqPCR and ChIP-qPCR assays. There are multiple considerations when designing and testing primers for these assays. This is not an exhaustive description of everything that can be done when designing primers to ensure a robust and linearly amplifying assay, but if these practices are followed and three primer sets per target are tested, at least one is usually useable. This does not take into consideration problems such as repeats or polyploid genomes but works well for mice and likely other mammalian genomes.

1. Considerations for all qPCR primers:

- a. **Linearity:** If there is twice as many amplicons in a sample, there should be twice as much fluorescent signal in a SYBR green assay.
 - i. To design primers more likely to have linear amplification, design amplicons to be 80-120 bp.

- ii. To assay for linearity, test primer sets on a 10-fold dilution series. This will test a wide range of concentrations of template.
- b. **Amplification of 1 amplicon:** If there is amplification of multiple products, then you are quantifying the quantity of all of those products together at once and it is not possible to separate which product the signal is coming from.
 - i. To design primers that are more likely to amplify only 1 product, perform *in silico* PCR on the UCSC genome browser.
 - ii. To screen for multiple products, always run a melt curve after PCR amplification. Each amplicon will have a different melt temperature.
- c. **Absence of primer dimer:** Some primers hybridize to themselves and the resulting dsDNA will cause fluorescent signal.
 - i. To screen for primer dimer, run a no template control (NTC). There should be no amplification in the NTC wells.

2. Considerations for RTqPCR primers:

- a. **Absence of gDNA amplification:** Should not amplify genomic DNA (gDNA), as this will cause background if your RNA samples originally contain gDNA contamination. In genes with only 1 exon, this is not always possible.
 - i. To avoid gDNA amplification, design primers that span exon-exon junctions, which increases the amplicon length for gDNA amplicons. Short introns can still be amplified, so try to avoid exon-exon junctions with a short intron between.
 - ii. To screen for gDNA amplification, test the primers against gDNA.

III. RTqPCR

Reagents

cDNA

This cDNA should be from a tissue that should be expressing the RNAs of interest.

Power SYBR Green PCR Master Mix (Applied Biosystems 4367659)

This mastermix contains all the necessary components for qPCR except template and primers. It contains a dye that fluoresces when it is intercalated in dsDNA, which is how PCR amplification is detected by the qPCR machine lasers.

10 ng/μL genomic DNA

unsonicated gDNA for testing whether primers amplify DNA contamination.

ApE plasmid editor software

A free open-source software for manipulating relatively short DNA sequences, such as plasmids. The latest version can be found and downloaded with a simple internet search.

Design

1. Gene structure

- a. Look at the gene on the UCSC genome browser to see what the gene looks like: ie 1 or many isoforms or overlap with another gene. This will inform which mRNA variant you download.

2. Sequence download

- a. Download the mRNA sequence from the NCBI nucleotide dropdown menu. Be aware there often many “variants” which is how NCBI refers to isoforms. There are also partial sequences, so read the description.
- b. Copy-paste the entire NCBI entry from “LOCUS” to the “//” that occurs after the DNA sequence information into ApE. Using the entire entry will allow ApE to automatically annotate sequence features, such as exons.
- c. Choose the exon-exon junction of interest.
- d. Copy the sequence where an amplicon is desired, plus some buffer sequence on either side.

3. Primer selection

- a. Primer3 is an online tool that is available from many sites. Use it.
 - i. 80-120 bp amplicon
 - ii. 58-60°C melting
 1. Do all primers in this range so you can run them all on one plate
 2. This range is based on the annealing/extension preferences of POWER SYBR
 - iii. 18-22 bp primers
 - iv. Use the Primer3 setting to force inclusion of the exon-exon junction in the design

4. *In silico* screening

- a. Use the *in silico* PCR tool on the UCSC genome browser to screen primers for
 - i. Multiple amplicons
 - ii. Amplifying gDNA (make sure to run *in silico* against both UCSC genes, which represent spliced transcripts, and the genome assembly)

IV. ChIP-qPCR

Reagents

10 ng/μL DNA from crosslinked and sonicated chromatin

Power SYBR Green PCR Master Mix (Applied Biosystems 4367659)

This mastermix contains all the necessary components for qPCR except template and primers. It contains a dye that fluoresces when it is intercalated in dsDNA, which is how PCR amplification is detected by the qPCR machine lasers.

Design

1. Locus features

- a. Look at the gene on the UCSC genome browser to see what the feature looks like with respect to known datasets like DNase, histone modifications, TF binding, etc

2. Sequence download

- a. Zoom into the locus where an amplicon is desired, plus some buffer sequence on either side.
- b. Click View → DNA
- c. Mask repeats then get DNA.

3. Primer selection

- a. Primer3 is an online tool that is available from many sites. Use it.
 - i. 80-120 bp amplicon
 - ii. 58-60°C melting
 1. Do all primers in this range so you can run them all on one plate
 2. This range is based on the annealing/extension preferences of POWER SYBR
 - iii. 18-22 bp primers

4. *In silico* screening

- a. Use the *in silico* PCR tool on the UCSC genome browser to screen primers for
 - i. Multiple amplicons

V. Primer Assessment

1. Templates

- a. NTC: **No Template Control**. Use water.
- b. 10-fold dilution series for RTqPCR
 - i. 50 ng/uL cDNA: Not actually 50 ng DNA by a nanodrop, but the equivalent of cDNA had 50 ng/ μ L RNA in it.
 - ii. 5 ng/uL cDNA
 - iii. 0.5 ng/uL cDNA
 - iv. 0.05 ng/uL cDNA
- c. 10-fold dilution series for ChIP-qPCR
 - i. 10 ng/uL input DNA (from crosslinked, sonicated chromatin)
 - ii. 1 ng/uL input DNA
 - iii. 0.1 ng/uL input DNA
 - iv. 0.01 ng/uL input DNA
- d. 10 ng/ μ L gDNA for RTqPCR

2. Template and Primer Mastermixes

MM type	Reagent	1X RXN volume	# primer sets * 1.15 * 2	13.8X (RTqPCR) 11.5X (ChIP-qPCR)
Template MM	2X Power SYBR MM	4		
	Template	1		
Primer MM	2X Power SYBR MM	1		13.8 or 11.5
	20 μ M primer mix	3.9		53.9 or 44.9
	H ₂ O	0.1		1.38 or 1.15

- a. The mastermixes were designed to both add up to 5 μ L, which is a volume most 100 μ L max volume repeat pipettors can do 20 times, making plate loading rapid and precise.

3. Plate loading

a. The first 12 columns and rows A-F of a 384 well plate shown below

		1,2	3,4	5,6	7,8	9,10	11,12
A	10-fold dilution series	Primer set 1	Primer set 2	Primer set 3	Primer set 4	Primer set 5	Primer set 6
B	10-fold dilution series						
C	10-fold dilution series						
D	10-fold dilution series						
E	NTC template						
F	gDNA template (RTqPCR only)						

4. Thermal cycling with melt curve (per manufacturer's protocol)

- a. 95°C 10 minutes: hot start
- b. 40 cycles
 - i. 95°C 15 seconds: denature
 - ii. 60°C 45 seconds: anneal and extend
- c. Melt curve
 - i. 95°C 15 seconds
 - ii. 60°C 15 seconds
 - iii. 95°C 15 seconds

5. Analysis

- a. **Check the amplification plots** (Cycle versus ΔRN) (Figure 1A)
 - i. Sometime the program choses a cutoff (green horizontal line) at the extremes of the y-axis, where the signal is noise or affected by limited primers or nucleotides. A cutoff in the middle of the amplification plot is desired.
- b. **Screen for gDNA amplification (for RTqPCR primers)**
 - i. Select the gDNA wells and then check the melt curve. This should not have signal. The Ct in the table should be undetermined, though ≥ 35 Ct is acceptable because 35 cycles is the max used in real experimental conditions.
- c. **Screen for primer dimer**
 - i. Select the NTC wells and then check the melt curve for amplification. This should not have signal. The Ct in the table should be undetermined, though ≥ 35 Ct is acceptable because 35 cycles is the max used in real experimental conditions.
- d. **Screen for multiple amplicons (Figure 1B)**

- i. Select all templated samples and check the melt curve. There should be one peak. Multiple peaks reflect the different melting temperatures of different amplicons.
- e. **Screen for linearity (Figure 1C)**
- i. With a 10-fold template dilution series and PCR doubling amplicons every cycle, 3.32-fold more signal is expected for every 10-fold increase in template. An ideal slope would therefore be by 3.32, but 3.1 to 3.5 are acceptable.
 - ii. Note the range of Cts that correspond to linear implication. Some primer sets work well from, for example, Ct 20-30, but are no longer linear at higher Ct. The Ct range where primers work is not exact, since the fluorescence cutoff (green horizontal line in Figure 1A) varies between plates.

VI. Figures

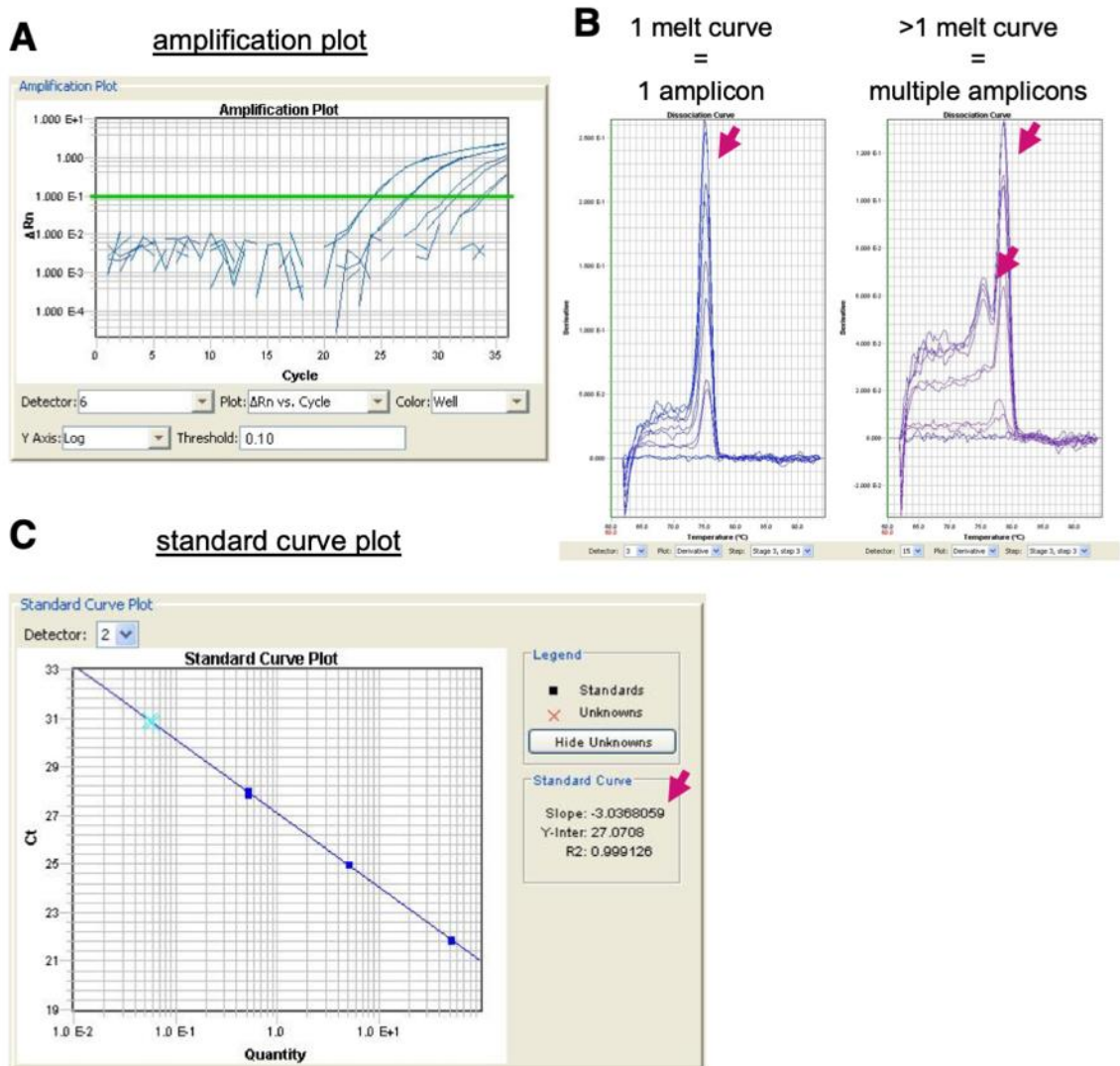


Figure 1: Assessing primer efficacy

Data as presented by Applied Biosystems' SDS software.

- A)** Amplification plot of a 10-fold dilution series.
- B)** Melt curves, also known as dissociation curves. Note on left an example where there is only 1 melt curve, indicating that there was only 1 amplicon. Note on the right, multiple amplicons, possibly from amplification of pseudogenes, multiple splice variants, or amplification of gDNA.
- C)** Standard curve plot. Note to the right (pink arrow) the slope. The slope may be ± 3.3 , depending on the plate setup. In the graph, points off the line may indicate non-linearity at extreme Ct and can be omitted, which will cause re-calculation of the slope, but the linear Ct range should be noted and not considered in future experiments.

VII. Verified *mus musculus* RTqPCR primers

gDNA= amplifies gDNA

Target	Forward	Reverse	Comment
<i>Alb</i>	TTTCCAGGGGTGTGTTTCGC	ACTGGGAAAAGGCAATCAGGA	
<i>Alb</i>	CACCATTGAAAGGCCAGAGG	TGGGGCATAGAAATAAGGATGTC	
<i>Ezh2</i>	GCTCAAGAGGTTCAGAAGAGC	CTGTATCCTCCGCTGCTTCC	Exon 2-3
<i>Ezh2</i>	CTCATTGCGCGGGACTAGG	GCATTCAAGGCTTTAACGGG	Exon 3-4
<i>Ezh2</i>	CCTCTGTCTCACGTGTGGAG	GACGGTGCCAGCAGTAAGT	Exon 15-16
<i>Ezh2</i>	TTACTGCTGGCACCCTCTG	TCTGTCTGCTTCATCCTGAGA	Exon 16-17
<i>Ezh2</i>	GGATGAAGCAGACAGAAGAGGA	CTTTCGGGTTGCATCCACCA	Exon 17-18
<i>Ezh2</i>	GACCACAGGATAGGCATCTTTG	CCCACATACTTCAGGGCATCA	Exon 19-20
<i>Tifa</i>	CCTATCCCCAGGCTTCCA	CCACAGCTAGATCAGGACTCC	
<i>Tifa</i>	GGAGTCTGTACTAGCTGTGG	CAAGTGACCGTCTCCTCTGT	
<i>Ntrk2 var 2</i>	TTGACCCGGAGAACATCACG	GGTTTCTCAGCCCCACGTAA	
<i>Ntrk2 var 2</i>	CTCATTGCAAACCAGAAAAGGC	GCCACAAACTTTAAGCCGGA	
<i>Ntrk2 var 2</i>	GCAGCGACGTCATCCTACA	ACCTGAGGCTTGTCTTCCAA	
<i>Prdm16</i>	GAGAGAGATTCCGCGAGCC	AGGTCCGGGTCAGGTTTATA	
<i>Prdm16</i>	TGCATGTGAAAGAAGGTGCC	TCACACTCATCACAGCGGAA	gDNA
<i>Cbfa2t3</i>	CCATCCTGATCCCCGAGAGC	ATTCTTCTGCCCACTCGCGT	
<i>Entpd2</i>	AGCACTCCACTCTACCTGGGA	GGTACCGTGTGAGCGTCTGT	
<i>Fstl1</i>	GCTCCCACCTTCGCCTCTAA	GCCAGCCATCGTTCCACAT	
<i>Fstl1</i>	TTTGTGGAGCTGGCAGGGAA	TGCCATTACTGCCACACACA	
<i>Cbx6</i>	TGGGAGCCAGAGGAGAACAT	TCTTCTTGGGCCCATACAGCT	
<i>Ngf</i>	TGCCAAGGACGCAGCTTTCT	TCTGCCTGTACGCCGATCAA	gDNA
<i>Ngf</i>	CAGTGAGGTGCATAGCGTAATGT	GCTATCTGTGTACGGTTCTGCC	gDNA
<i>Wnt4</i>	AACCGGCGCTGGAAGTGTTC	ACCTGCTGAAGAGATGGCGT	
<i>Igf2bp3</i>	TCGGTCCCTAAACGGCAGAG	CACAGCTCTCCACCACTCCA	
<i>Fgfr1 var 3</i>	CGGGAGTAAGATCGGGCCAG	CCTCCATTTCTTGTGCGGTGG	
<i>Rgma</i>	CAATACACCTGTGCTGCCGG	TCTGGTCCACACTCTTGGA	
<i>Tulp3</i>	TGCCCTTGACGATGAGACCCT	CAAGGCGCTTCTTCTCTGCT	
<i>Tnfsf12</i>	TTGCAGCCATTATGAGGTCA	GGTCTCTCCAGCCACTCA	
<i>Cdkn2a</i>	CCGACGGGCATAGCTTCAG	GCTGAGGCCGATTTAGCTC	
<i>Krt19</i>	CCCTCCCGAGATTACAACCA	GGCGAGCATTGTCAATCTGT	
<i>Krt19</i>	CCACCTACCTTGCTCGGATT	ACTTCGGTCTTGCTTATCTGGA	
<i>Krt19</i>	ACCACTACTTTAAGACCATCGAG	GCGAGCATTGTCAATCTGTAGG	
<i>Hes1</i>	CCAAGCTAGAGAAGGCAGAC	GGTATTTCCCAACACGCTC	
<i>Hes1</i>	GAAGCACCTCCGGAACCT	GTCACCTCGTTTATGCACTC	
<i>Col1a1</i>	GTTCAAGCTTTGTGACCTCC	TTCAGGATGTCTTCTTGCC	
<i>Foxa3</i>	GTGGAGCTACTACCCGGAG	TGAGTGGGTTCAAGGTCATG	
<i>Cebpa</i>	CAAGAGCCGAGATAAAGCCA	CAGGCGGTCATTGTCACTG	gDNA
<i>Onecut1</i>	TGGAAGTGGCTGCAGGAG	CGTGTCTTGTCTTTCCGT	
<i>Onecut2</i>	CAGGATGTGGAAGTGCTG	CTGCGAGTTGTTCTGTCTT	

<i>Sox9</i>	AGACCAGTACCCGCATCTG	AAGGGTCTCTTCTCGCTCTC	gDNA
<i>Sox9</i>	ATTCTCTCCGGCATGAG	GATCAACTTTGCCAGCTTGC	gDNA
<i>Sox9</i>	ACTCCCCACATTCTCTCTC	GTTTTGGGAGTGGTGGGTG	gDNA
<i>Hnf4a</i> <i>var 7/8/9</i>	TCCTTATGACACGTCCCCATCT	GAGGCTCCGTAGTGTGGCC	embryonic liver
<i>Hnf4a</i> <i>var 1/2/3</i>	GGCAATGACACGTCCCCATC	CACAGATGGCGCACAGGG	adult liver
<i>Hnf1b</i>	AGAGAGCTGCCCTGTACT	AGAGCTCTGGACTGTCTGGT	
<i>Foxa1</i>	TGGCTCCAGGATGTTAGGGA	ACAGGGACAGAGGAGTAGGC	
<i>Sox4</i>	TGGAAGCTGCTCAAGGACAG	GCCGGTACTTGTAGTCAGGG	gDNA
<i>Gfap</i>	CTTTGCACAGGACCTCGGCA	ACGCAGCCAGGTTGTTCTCT	
<i>Acta2</i>	GGAGAAGCCCAGCCAGTCG	CCAGAGCCATTGTCGCACAC	
<i>Des</i>	GTGGAGCGTGACAACCTGAT	TCGGAAGGCAGCCAAGTTGT	
<i>Pecam1</i> <i>var1</i>	AGGATTCAGCTGAGGTGGGC	TGGATGCTGTTGATGGTGAAGG	