Epigenetic chromatin states during embryonic development and adult homeostasis of the mammalian gut

Dissertation

zur Erlangung des Grades

„Doktor der Naturwissenschaften“

Am Fachbereich Biologie

der Johannes Gutenberg-Universität in Mainz

Juri Kazakevych

geb. am 07.03.1987 in Cherson (Ukraine)

Mainz, 2016

Tag der mündlichen Prüfung: 11. Mai

Table of Contents

1.1 Summary

The mammalian small intestine is a complex 3D-structure comprised of various specialized, proliferative and fully differentiated cell types. Although highly specialized in adults, all cellular lineages originate from a monolayered epithelial tube in the early embryo, which consists of a presumably homogenous population. As the fastest proliferating tissue, the murine small intestine is an established model to study embryonic development and adult homeostasis. These processes are mainly regulated in two ways: intracellular signaling, such as the proliferative Wnt-pathway in combination with Notch- Hedgehog- and BMP-signaling, and epigenetic mechanisms, on which this thesis focuses. Here, I assessed the epigenetic marks H3K27Ac, H3K27me3, H3K4me3, H2A.Z and DNA-methylation, which are involved in the regulation of gene activity, providing a cell type and a development specific expression pattern.

The main goal of the projects described in this thesis was to elucidate epigenetic mark deposition and their interaction on the genome-wide scale and to assess their functional impact on gene activity at single genetic loci. Furthermore, single cell *in situ* analysis was employed to visualize tissue heterogeneity of gene expression and epigenetic patterning. Both approaches were used to deepen our understanding of epigenetic mechanisms involved in development and homeostasis as well as in cancer pathogenesis.

Active and repressive epigenetic marks' genomic distribution was analyzed throughout the intestinal development from embryonic to the fully differentiated adult cell types. In combination with RNA-seq data, these ChIP-seq and MBD-seq experiments provided a comprehensive dataset, which allows for the investigation of epigenetic marks' interactions along the intestinal cell development timecourse. Hereby, H2A.Z, a histone variant previously annotated to active genes, was found to be positioned on responsive rather than on active promoters. Moreover, H2A.Z was found to be decoupled from gene activity regulation in fully differentiated cells. The well-known active promoter and enhancer mark H3K27Ac was additionally found to mark poised enhancers, long before the corresponding gene's expression. This novel premarking effect, in combination with the available timecourse gene expression data, could be used for improved genome-wide enhancer prediction and target gene identification.

Numerous potential developmental marker genes were identified through developmental clustering. Single locus analysis of these genes revealed strong heterogeneity of gene expression and epigenetic levels in adult intestinal tissue. Strikingly, similar gene expression patterns were also observed at embryonic stages, which previously were believed to be uniform. This heterogeneity might be the key to adult tissue homeostasis and functionality as well as to the early cell fate commitment and structural patterning in the embryonic epithelium.

1.2 Zusammenfassung

Der Dünndarm besitzt eine komplexe dreidimensionale Struktur, aufgebaut aus mehreren spezialisierten, proliferativen und voll ausdifferenzierten Zelltypen. Trotz des hohen Spezialisierungsgrads im adulten Gewebe stammen alle Zellen des Dünndarmepithels von einer homogen aussehenden embryonalen Population ab. Der sich alle 5 Tage erneuernde Dünndarm ist das sich am schnellsten teilende Gewebe in der Maus und damit ein gängiges Modell für Studien der Embryonalentwicklung sowie der adulten Homöostase. Diese Prozesse werden vor allem auf zwei Arten gesteuert: Intrazelluläre Signalwege und epigenetische Mechanismen. Letzteres steht im Fokus dieser Doktorarbeit. Hierfür wurden die epigenetischen Histonmarkierungen H3K27Ac/me3, H3K4me3 und H2A.Z, sowie die DNA-Methylierung untersucht, die zusammen an der Steuerung der entwicklungsspezifischen Genexpression beteiligt sind.

Die Zielsetzung dieser Arbeit war es, die Verteilung und Wechselwirkung epigenetischer Markierungen genomweit zu bestimmen und ihre Auswirkung auf die Aktivität einzelner Gene zu überprüfen. Zusätzlich wurden bildgebende *in situ* Methoden verwendet, um die Heterogenität adulten und embryonalen Gewebes, insbesondere im Hinblick auf Genexpression und Histonmarkierungen, zu untersuchen. Beide Ansätze dienten dazu, unser Verständnis epigenetischer Mechanismen, die an der Embryonalentwicklung und Homöostase, aber nicht zuletzt auch an der Karzinogenese beteiligt sind, zu vertiefen.

Die Verteilung aktiver und repressiver epigenetischer Markierungen, sowie die Genexpression wurde während der Differenzierung von embryonalen Stadien zum adulten Dünndarmepithelium mithilfe von ChIP- und RNA-seq analysiert. Der hieraus entstandene umfassende Datensatz ermöglicht eine detaillierte Untersuchung epigenetischer Wechselwirkungen während des gesamten Entwicklungsverlaufs. Hierbei wurde H2A.Z, eine Histonvariante, deren Positionierung bisher mit aktiven Promotern und Enhancern assoziiert wurde, als ein Marker von regulationsempfindlichen genetischen Elementen identifiziert und zwar unabhängig von der Genaktivität. Insbesondere in voll ausdifferenzierten Zellen wurde eine Abkopplung der H2A.Z-Positionierung von der Genregulation beobachtet. Die Histonmarkierung H3K27Ac, bekannt für ihre Positionierung auf aktiven Promotern und Enhancern, wurde auch auf bereitstehenden Enhancern lange vor Expression des dazugehörenden Gens gefunden. Dieser neue Vormarkierungseffekt könnte zur verbesserten Identifizierung von Enhancersequenzen sowie der dazugehörigen Zielgene verwendet werden.

Zahlreiche Gene wurden mithilfe der epigenetischen Gruppierung als potentielle Entwicklungsmarker identifiziert. Ihre Einzelanalyse deckte in vielen Fällen eine starke Heterogenität der Genexpression und epigenetischer Markierungen im adulten Gewebe auf. Interessanterweise wurde dies oft auch in den bisher als größtenteils homogen beschriebenen, embryonalen Stadien beobachtet. Diese Heterogenität könnte der Schlüssel zur adulten Homöostase sowie der embryonalen Spezialisierung des Dünndarmepithels sein.

2. Introduction

2.1 Gene regulation

Although most cells of an organism contain the same DNA-encoded information, a large number of distinct cell types are currently known and more continuously discovered and classified[1]. Each of these cell types have distinct functional properties as e.g. their cellular structure, metabolism and proliferation rate. Cell type specifications are established during embryogenesis and adult homeostasis by the differential regulation of genes, the smallest functional units of our genome[2]. On the other hand, mutation-, stress- or age-induced deregulation of gene activity can lead to the loss of cellular functions, cell death or in the worst case to the emergence of cancer[3].

Eukaryotic regulation of gene expression mainly occurs on three stages: Transcription, RNA-processing and, if the gene product is a protein, translation[1,3]. At the first stage, accessibility of the gene for the transcription machinery and the exact composition of the gene binding complexes play major roles, determining transcriptional levels. Gene accessibility is hereby regulated by the overall chromatin architecture with open euchromatin and highly condensed heterochromatin which shuts down the expression of contained genes[4]. A single genes accessibility is further regulated by the histones on its promoter, gene body and regulatory elements, locally condensing or opening the chromatin up for transcription factors and the RNA-polymerase II complex which transcribes messenger RNA (mRNA) as well as microRNA (miRNA)[5,6]. On accessible genes the composition of the RNA-polymerase II complex, which is in turn affected by a wide array of regulatory pathways as well as regulatory factors binding to promoters and *cis*-regulatory elements, can lead to transcription repression, initiation and stalling, yielding a certain level of precursor-mRNA[5,7]. Before leaving the nucleus and being translated, the precursor-mRNA is spliced, polyadenylated and 5' capped[3]. Often several mature mRNA products can be obtained from the same precursor-mRNA by alternative splicing, which does not only affect the genes products size and function but also the lifetime of mRNA[8-10]. The stability and subcellular localization of mRNA strongly depends on post-transcriptional processing, and mRNA-degradation is a crucial regulatory mechanism of gene product levels[3]. Mature mRNA is exported from the nucleus either to the cytosol or the endoplasmic reticulum and is translated to a polypeptide which in turn can undergo a variety of post-translational modifications and/or degradation[3].

Each of the described mechanisms can occur in a developmentally-, tissue- or even allele-specific manner and their combination allows for fine-tuning of gene-expression, ultimately leading to highly controlled cell-type specification and function.

2.2 Epigenetics

Epigenetics generally stands for not DNA-encoded regulatory mechanisms of gene expression, although varying definitions are in use, some of which underline the mitotic inheritability of phenotypes[11,12]. This term was firstly introduced by Conrad Waddington in 1942, naming the connecting mechanisms between a homogenous genotype and a heterogeneous phenotype[13] (see ***Fig. 1***). In embryogenesis and adult cell differentiation the possibility of a genotype to produce several phenotypes (phenotypic plasticity) goes hand in hand with "canalization", the stability of phenotypes. This explains how, beginning from a single stem cell (SC), a variety of irreversibly differentiated cell types can occur without any changes in the DNA-sequence.

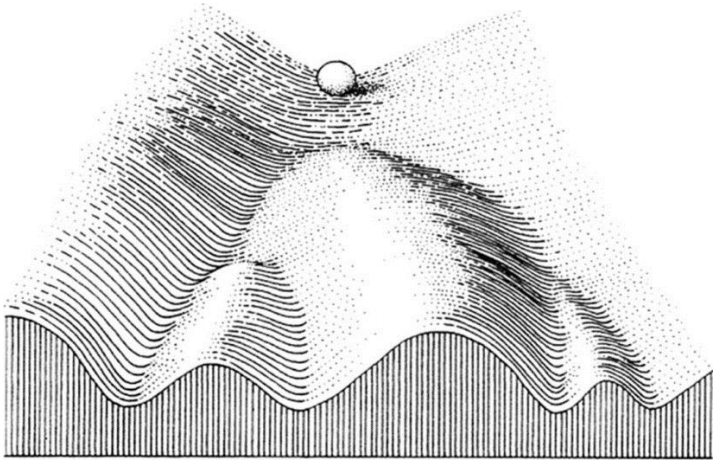


Figure 1: Waddington's model of stemness and differentiation mediated by epigenetic mechanisms. *C.H. Waddington, 1957[14]. As a marble rolling down the depicted hill can reach any of the four depressions at the bottom, a SC can provide any of the downstream cell fates. However, as the marble cannot reach the right depressions once it is half the way to the left ones, the differentiating SC commits to a certain pathway and is restricted in its prospective cell types. The elevations between different cell fates are provided by a set of epigenetic mechanisms, as the genetic information is unchanging throughout differentiation.*

Four main mechanisms are known to contribute to the epigenetic control of cellular functions: the formation of nuclear territories, DNA-methylation, histone variants and a variety of post-translational histone modifications, such as methylation, acetylation, phosphorylation or ubiquitination[15-19]. All four mechanisms are functionally interconnected and allow to adjust gene activity from a shutdown to constitutive expression at the single cell level. This dynamic regulatory network allows to maintain pluripotency of embryonic and adult SC as well as to determine cell fate upon differentiation. The epigenetic mechanisms also enable to some extent the reverse pathway from differentiated to pluripotent cells, although this is occurring rarely in healthy organisms and is mostly associated with cancer[20,21].

It is of special note, that many epigenetic marks are mitotically inherited in somatic cells and in some cases can stay constant throughout the organism lifetime[22]. In contrast, transgenerational inheritance is restricted in sexually reproducing organisms due to several meiotic epigenetic reprogramming steps which nearly completely erase epigenetic marks[22,23], so that inheritance is usually only observed up to the 2nd and 3rd generations, as the embryo (2nd generation) and it's germ cells (3rd generation) can be affected by maternal and environmental factors[24]. For example, the risk of obesity was shown to be increased in two offspring generations if the mother was exposed to undernutrition during the Dutch famine of 1944, which can be explained by changes of DNA-methylation changes of metabolic genes and was confirmed by animal studies[25]. Epigenetic inheritance beyond the 3rd generation is rare but was also observed across many non-mammalian species[22].

Another important field involving epigenetics is aging. Aging phenotypes as well as cancer were for a long time thought to be mostly caused by genetic effects, such as mutations and telomere shortening, however increasing evidence is showing the involvement of epigenetic mechanisms, such as the loss or ectopic gain of epigenetic marks, which can deregulate gene expression in similar ways as genetic aberrations[26].

2.3 Epigenetic transcriptional regulation

2.3.1 Nuclear domains

The formation of accessible active (euchromatic) and condensed inactive (heterochromatic) nuclear territories regulates big stretches of the chromatin, which can be up to a chromosome big, as in the case of Barr bodies – the inactivated X-chromosome in females[27,28]. Additionally, the 3-dimensional chromatin structure is regulating gene activity by physical proximity of the transcription start site (TSS) of a gene with regulatory elements such as enhancers[29]. The concept of "transcription factories" describes euchromatin domains with high transcription levels of co-regulated genes sharing active RNA-polymerase II complexes and strong enhancer elements, so that depending on the movement of a gene in or outside of a transcription factory, its expression varies strongly[29]. Although the effects of nuclear localization are well researched, the molecular mechanisms leading to the 3-dimensional nuclear structure remain mainly unclear.

2.3.2 DNA-methylation

Across all eukaryotic species cytosine methylation (5-mC) is observed at varying levels: from 14 % in *Arabidopsis thaliana*, 7.6 % in *Mus musculus* and <0.1 % in *Saccharomyces cerevisiae*[30,31]. In adult mammals 5-mC is mostly found in the CpG dinucleotide context covering 60-80 % of these loci[32,33], while non-CpG methylation is observed in embryonic SC contributing ~25% to overall methylation in human

embryonic SC[34,35] and at lower levels in skeletal muscles[36] and neuron progenitors[37] while virtually absent in most differentiated cell types.

CpG-methylation is set by the DNA methyltransferase (DNMT) protein family as a symmetric mark, meaning that usually both DNA-strands are either methylated or unmethylated[38]. DNMT1 acts hereby as a "maintenance"-methyltransferase, methylating the second strand of a hemimethylated CpG locus and hence providing somatic heritability of CpG-methylation, while the nonsymmetrical non-CpG-methylations are segregated upon cell division and ultimately lost after several rounds[38,39]. De novo CpG methylation is established by DNMT3A and DNMT3B at early embryonic stages, while de novo non-CpG methylation is performed by DNMT3A/B in a complex with DNMT3L also in adult cells[38,40].

Regions with high CpG content, called CpG-islands, are found throughout mammalian genomes and are especially enriched at promoters, covering 70% of them in humans[41]. While 5-mC is distributed all over the genome, transcription factor binding sites and CpG islands are often hypomethylated and repetitive elements usually hypermethylated[34,42,43].

Methylated CpG dinucleotides in promoter and enhancer regions strongly anti-correlate with gene expression levels[32,44]. One mechanism of gene repression by 5-mC is the direct binding inhibition of transcription factors and activating chromatin remodelers to chromatin[45,46], which can also impair enhancer-promoter interactions[47]. On the other hand, hypomethylated TSS are not necessarily active but can be primed for later induction, e.g. during differentiation[32] or by a stress-response pathway[48].

Opposite to promoter silencing, methylation of the gene body is often found on active genes[49] with highest levels on moderately expressed genes[50,51] and also correlates to constitutive expression in plants and invertebrates[52]. Particularly in exonic regions, it might contribute to correct co-transcriptional splicing by pausing the RNA-Polymerase II[53,54] and it was proposed to suppress transcription from cryptic promoters inside active genes[52].

It was previously shown that DNA-methylation strongly changes during embryonic SC differentiation[32,34] and adult hematopoietic cell differentiation[55]. However it stabilizes in adult stages[56] and can even become completely static, as it was shown in the small intestinal epithelium[32]. This indicates that 5-mC is an important developmental marker while stably maintaining the lineage commitment of adult cells.

Besides its above described roles in cotranscriptional splicing and gene silencing, DNA-methylation is also involved in genomic imprinting, X-chromosome inactivation and the silencing of transposable elements[42,43].

2.3.3 Histone variants

The fundamental unit of chromatin packaging is the nucleosome core consisting of a histone octamer with approx. 147 bp of DNA wrapped around it plus up to 80 bp of linker DNA connecting it to the neighboring nucleosome[57]. While in heterochromatic domains the chromatin is very densely packed and nucleosomes are immobile, in euchromatic regions nucleosomes can shift position along the DNA or be completely displaced leaving a "naked" DNA region, highly accessible for factors involved in transcription or other nuclear processes[58]. Besides the complete removal and repositioning of nucleosome cores, single histone monomers can also be exchanged[59]. In addition to the 4 standard histone variants H2A, H2B, H3, and H4, each contributing as a dimer to the nucleosomic octamer, multiple alternative variants are known[17] (see ***Fig. 2***).

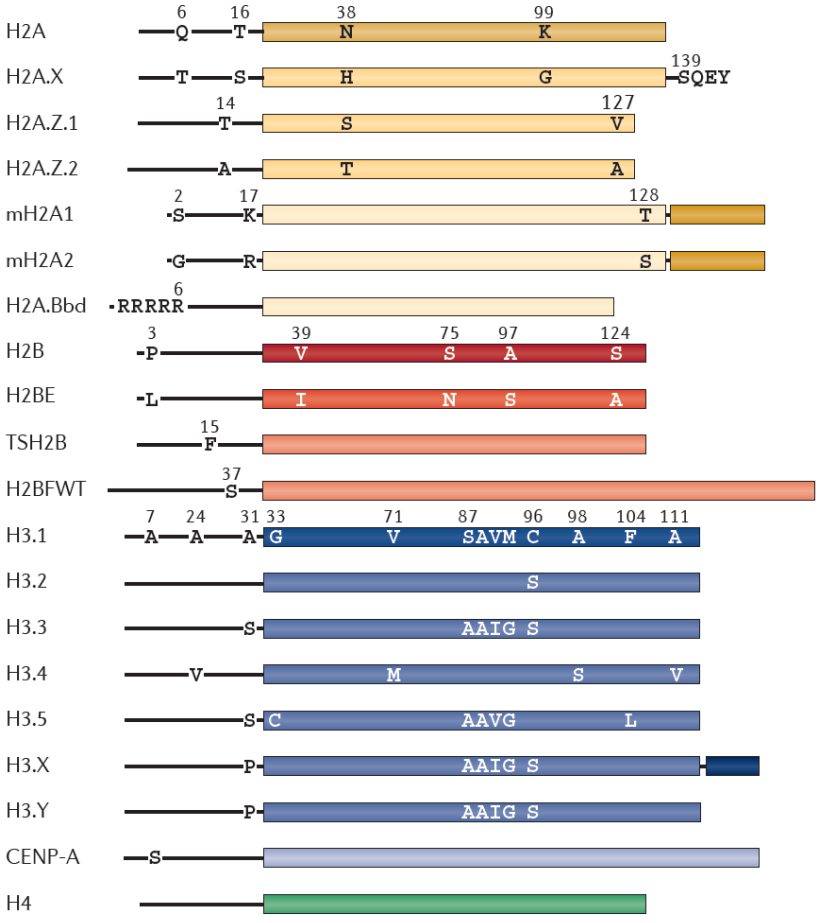


Figure 2: Human core histone variants. *Adapted from Maze et al., 2014[17]. Core domains of histone families are color-coded for similarity. Black depicts unstructured tails that are the prevalent histone modification targets. Capital letters indicate key differences of amino acid sequence inside a histone family.*

Histone variants are known to contribute to a variety of cellular processes, such as chromosome segregation, recombination, transcription initiation, X-chromosome condensation and DNA repair[60]. For example, the H2A.X histone variant, which constitutes up to 25% of all H2.A in mammals, is phosphorylated at DNA-double strand breaks and recruits DNA-repair proteins to the damage site[61-63].

The high conservation of another H2.A variant, H2A.Z, between most eukaryotic species indicates its essential role in nuclear processes[61]. It flanks nucleosome free regions[64,65], is involved in destabilizing the nucleosome structure[66] and nucleosome depletion[67,68] hence increasing accessibility for chromatin remodelers and the transcription machinery. So for example a knockdown of H2A.Z decreases chromatin accessibility on a wide set of genomic loci, including enhancer regions[66]. An especially labile nucleosome fraction is marked by H2A.Z in combination with H3.3. These nucleosomes are found on nucleosome exchange hotspots, mostly transcription start sites (TSS), transcription factor binding sites and cis-regulatory elements[64].

So far H2A.Z is associated with transcriptionally active or poised promoters, especially at the 5' end of the TSS, active enhancer elements[65,66,68-71], exons[68], gene bodies and the 3' end of genes[48]. Overall, H2A.Z is highly correlated with H3K4me3 marking active and poised promoters as well as to H3K27ac together with H3K4me1/2/3 marking active enhancers, while generally anti-correlated to the repressive H3K27me3 mark[65,66] (see section 2.3.4 for details on post-translational histone marks).

In ES cells H2A.Z deviates from these patterns and is linked to Polycomb regulation, colocalizing with PRC2 and the H3K27me3 inactive mark on silent/bivalent promoters, especially marking developmentally important transcription factors, while on differentiation the correlation to H3K27me3 is reversed at most genomic loci[72]. Upon differentiation the linkage of H2A.Z, positioning to PRC2/H3K27me3 localization is only maintained in a small subset genomic loci, now restricted to bivalent sites in combination with H3K4me3, as it was shown for murine neuron progenitors by Ku et al.[65]. The importance of H2A.Z in embryonic differentiation is further underlined by studies showing that H2A.Z knockdown attenuates nucleosome depletion[68] and impairs both the repression of ESC- and the activation of differentiation-specific genes previously kept in the poised state by the bivalent H3K4me3/H3K27me3 marks[66,72].

Further, H2A.Z shows a genome-wide and strong anti-correlation to the mostly repressive 5-mC methylation across eukaryotic species[50,73]. As DNA-methylation generally promotes nucleosome occupation followed by chromatin condensation and H2A.Z destabilizes the nucleosome promoting chromatin accessibility, the mutual exclusion of H2A.Z with 5-mC methylation is a potent epigenetic switch of gene activity. A study involving the DNA-demethylating compound 5-Aza-CDR showed increasing levels of H2A.Z on demethylated loci[67]. On the other hand it was shown, that depletion of H2A.Z has only

minor effects on DNA-methylation, suggesting that the regulatory relationship of these marks is unidirectional[48,67,73].

Interestingly, expression levels were reported to be anti-correlating to H2A.Z positioning at TSS in murine embryonic SC cells and neuronal progenitors, although H2A.Z is completely missing at loci stably repressed by DNA-methylation or H3K27me3[65]. It seems that H2A.Z maintains the full regulatory potential bivalent loci, while on strongly expressed genes it is depleted from the 3'-end of the TSS and only partially replaced by acetylated H2A.Z[65]. On the contrary, the study of Hu et al. shows positive correlation of H2A.Z to expression levels in murine embryonic SC cells but also confirming the H2A.Z-ac enrichment[66].

The deposition of H2A.Z on gene bodies, also strongly anti-correlating to 5-mC distribution, was reported across all eukaryotic species[74,75]. While mostly absent on silent genes, highest H2A.Z levels are found on low expressed genes and decreasing again towards stronger expressed ones[48,52]. H2A.Z was found at TSS and gene bodies of responsive genes, such as stress- or differentiation-induced/repressed ones[48,72]. In *A. thaliana*, genes marked by H2A.Z in the gene body were shown to be induced/repressed several fold stronger on stress stimulus than genes with low gene body H2A.Z, while a depletion of H2A.Z mostly affected those gene's expression[48]. Thus, one mechanism of expression stabilization of constitutively active genes by gene body methylation seems to be the exclusion of H2A.Z.

2.3.4 Post-translational histone modifications

While the inner core histone domains of a nucleosome constitute the chromatin's backbone by binding in a changeless manner, the C'- and especially N'-terminal domains do not have a determined secondary structure and are a target for covalent post-translational modifications (see ***Fig. 3***). Especially histone acetylation, methylation, ubiquitination and phosphorylation on enhancers, promoters and gene bodies were shown to be involved in the regulation of gene expression and chromatin accessibility[76,77] (see ***Table 1***). In some cases this is achieved by changing the binding affinity of histones to DNA, but especially histone tail modifications are crucial for binding of regulatory factors to chromatin[77]. Already a small subset of histone modifications, especially H3K27ac, H3K27me3 and H3K4me3, can indicate gene activity by their localization on TSS and *cis*-acting regulatory elements as enhancers.

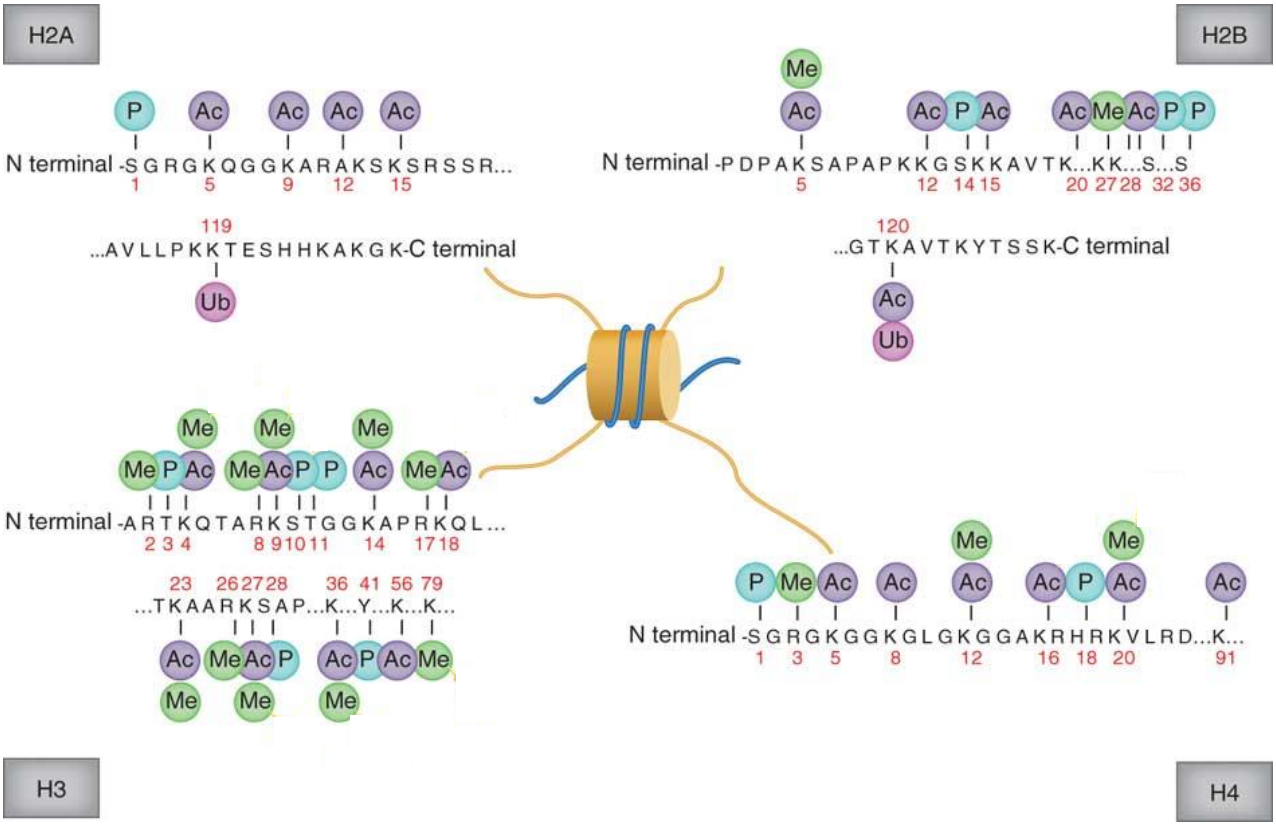


Figure 3: Core histone tail modifications. *Adapted from Rodríguez-Paredes et al., 2011*[78]*. The modifiable residue of N- and C-terminal domains of the four core histones H2A, H2B, H3 and H4 are annotated by their amino acid and position number. The most common modifications are indicated: Ac: acetylation; Me: mono-, di- or tri-methylation; P: phosphorylation; Ub: ubiquitination. Some residues can be modified by both, methylation and acetylation, although never simultaneously. Different patterns of histone modifications are specifically recognized by chromatin binders, affect the general accessibility of the DNA and regulate nucleosome stability.*

Modification	**Function**	**Localization**	**Modifying enzymes**
H3K4me1/2	Active	Enhancers, TSS	MLL1-4, SETD1A/B, ASH1L, PRMD9
H3K4me3	Active	TSS of active and poised genes	MLL1-4, SETD1A/B, ASH1L, PRMD9
H3K36me3	Active	Exons of active genes	SETD2
H3K27ac	Active	Enhancers of active genes	CBP, p300
H2A.Z-ac	Active	TSS, 3' end of active genes	NuA4[79]
H2A.Z-ac,ub		TSS of poised genes	NuA4, PRC1[79]
H3K27me3	Repressive	TSS of poised and silenced genes	PRC2

Table 1: A selection of post-translational histone modifications, their function, localization preference and protein (complexes) involved in their deposition. *Adapted from Elliot et. al, 2015* [16].

2.3.4.1 Acetylation

Lysine acetylation, e.g. H3K27ac, reduces the positive charge of the histone tail and hence decreases the binding affinity to the negatively charged DNA-backbone, which allows an easier access of the transcription machinery to the acetylated locus[80,81]. Additionally acetylated lysine residues provide a binding target for bromodomain containing chromatin regulators[82]. H3K27ac is known to be localized together with H3K4me3 on the promoters of active genes[83] and together with H3K4me1/2 on active enhancers[84-87].

H2A.Z N-terminal acetylation as well as C-terminal ubiquitination are linked to expression levels of the marked gene. While generally associated with TSS, H2A.Z-ac is strongly enriched at the 3' side of highly expressed genes' TSS[65]. Bivalent H3K4me3+H3K27me3 sites are enriched by the double-modified H2A.Z-ac,ub that is suggested to be a target for PRC2[65].

2.3.4.2 Methylation

The state of histone methylation on lysine and arginine residues is established by histone methyl transferases (HMTs), mainly SET-domain proteins such as the trithorax group protein MLL for H3K4me3[88,89], Polycomb repressive complex 2 (PRC2) for H3K27me3[90] and histone demethylases such as KDM1A for H3K4me1/2[91]. Unlike lysine acetylation, methylation can correlate to repression as well as to activation of gene expression, depending on the position on the histone and the number of methyl groups per lysine residue. So e.g. H3K4me3 on promoters and H3K36me3 on exons are established markers for gene expression[89,92,93], while H3K27me3 is a strong repressive mark[94], involved in silencing of embryonic lineage regulators and hence maintaining pluripotency[95]. Interestingly the H3K4me3 and H3K27me3 marks can coexist on the same TSS indicating a "poised" state of the gene, in which it is not expressed but can get activated if the repressive H3K27me3 mark is removed or on the contrary terminally silenced by removal of the active mark and introduction of DNA-methylation at the promoter[96,97]. The repressive function of H3K27me3 also directly opposes the active H3K27ac mark, as the lysine residue can only be either methylated or acetylated and mutual exclusion is generally observed on genomic elements[84].

Mono- and dimethylation of H3K4 at promoters seem only to be intermediates of trimethylation, without any strong correlation to gene activity. However they are marks of enhancer activity, being positioned alone on putative poised enhancers or in combination with H3K27ac on active enhancers, as it could be shown in various cell types[84-87,98]. It was also shown that actively marked enhancers (H3K4me1 + H3K27ac) correlate to actively marked nearby promoters (H3K4me3), while poised enhancers (H3k4me1 only) often neighbor poised promoters (H3K4me3 + H3K27me3)[84].

Both MLL and PRC2 histone methylase targeting of promoters and enhancers seem to be regulated by the positioning of H2A.Z. However this interaction is apparently mutual, as MLL4 activity was shown to

promote H2A.Z deposition at enhancers and the depletion of any of these regulatory components might impair enhancer activity[66].

2.4 Development and homeostasis of the small intestine

A suitable model to study the function and interplay of epigenetic marks in vivo is the murine small intestine. First, the intestinal epithelium is the fastest proliferating tissue in mammals and so it allows studying the fast and sharply controlled development from adult SC to fully differentiated cell types. Secondly, the main cell populations, namely the adult intestinal stem cells (ISC) and the fully differentiated adult enterocytes (AE) are found in different compartments of the monolayered epithelium and hence can not only be tracked and isolated by several specific markers, but also be easily identified after imaging. Finally, yet importantly, the embryonic development and adult homeostasis of the intestinal epithelium is well conserved between vertebrates, allowing comprehensive involvement of scientific results from other organisms, disease model applications and ultimately the translation of acquired results to human clinical research.

In this thesis I am focusing on four consecutive developmental stages of the small intestine: day 12.5 after gestation (E12.5) as an early embryonic stage with a uniform monolayered epithelium, E14.5 as the beginning of crypt-villus axis formation, adult Lgr5+ ISC and finally the non-proliferative terminally differentiated stage with absorptive AE and the Paneth cells, which represent the secretory lineage.

2.4.1 Embryonic gut development

Originating, same as the neighboring stomach and colon, from the endoderm, the embryonic small intestine is closing to a monolayered epithelial tube at day 9.5 of gestation (E9.5)[99]. Although the intestinal epithelium always stays monolayered, it undergoes extensive 3-dimensional rearrangements of its structure, which is supported and partially orchestrated by the surrounding mesenchyme[100]. Around E14.5, villi, worm-like protrusions emerge supported by a mesenchymal core towards the intestinal lumen, while during early postnatal development the intervilli regions are forming indentations into the mesenchyme called crypts[99,101] (see ***Fig. 4***). While at early embryonic stages the cells contributing to the intestinal epithelium look uniform, with the development of the crypt-villus axis the cells in different compartments are increasingly exposed to differential signaling and start to behave differently in terms of gene expression, signaling and morphology[99,101].

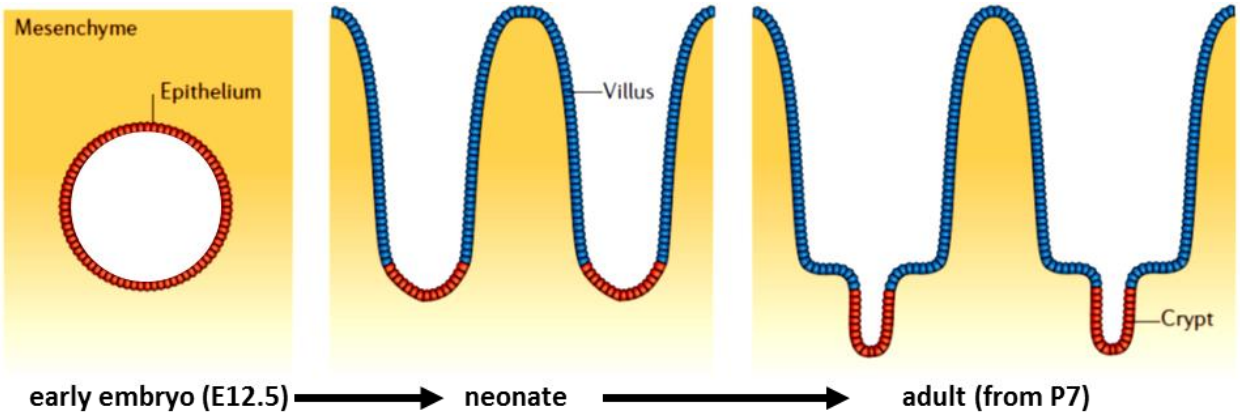


Figure 4: Small intestinal development. *Adapted from Crosnier et al., 2006[101]. Early embryonic gut is a monolayered tube, which develops villi, finger like protrusions into the lumen at later embryonic stages. Finally, until postnatal day 7, intervillus regions invaginate into the mesenchyme forming the crypts.*

Epithelial-mesenchymal communication is mediated from early embryonic stages by the Hedgehog (HH) signaling pathway, which is active at first throughout the epithelium but later concentrates in intervilli regions. This pathway was shown to be necessary for mesenchymal differentiation around the intestinal epithelium and at later embryonic stages for villus formation[102-104]. The development of the embryonic epithelium was reported to depend as the adult intestinal homeostasis on Wnt- and Notch-signaling, although the specific mechanisms, localization and timing of these pathways in embryos remain elusive[99,105,106].

2.4.2 Adult small intestine homeostasis

The adult gut epithelium maintains its function and structure by constant and fast production of fully differentiated absorptive and secretory cells, most of which are shed into the gut lumen 3-5 days after they start to differentiate[99].

The bottom of the crypt compartment harbors hereby the Lgr5-positive ISC which divide approx. once per day and give rise to all cells of the intestinal epithelium[107] (see ***Fig. 5***). Moreover, ISC were shown to give rise to organoids with crypt-villus structures containing all intestinal epithelial cell types *in vitro*[108]. ISC are dividing symmetrically and a possible further differentiation of any one of the daughter cells seems to be mainly regulated by its position in the crypt. Cells positioned in the upper crypt still proliferate for several rounds, but are destined to move outwards and terminally differentiate[101,109]. This positioning effect on the fate of dividing stem cells leads to neutral competition of ISC and finally to monoclonality of each intestinal crypt, although in the embryonic tissue each intervillus compartment is polyclonal[109-111].

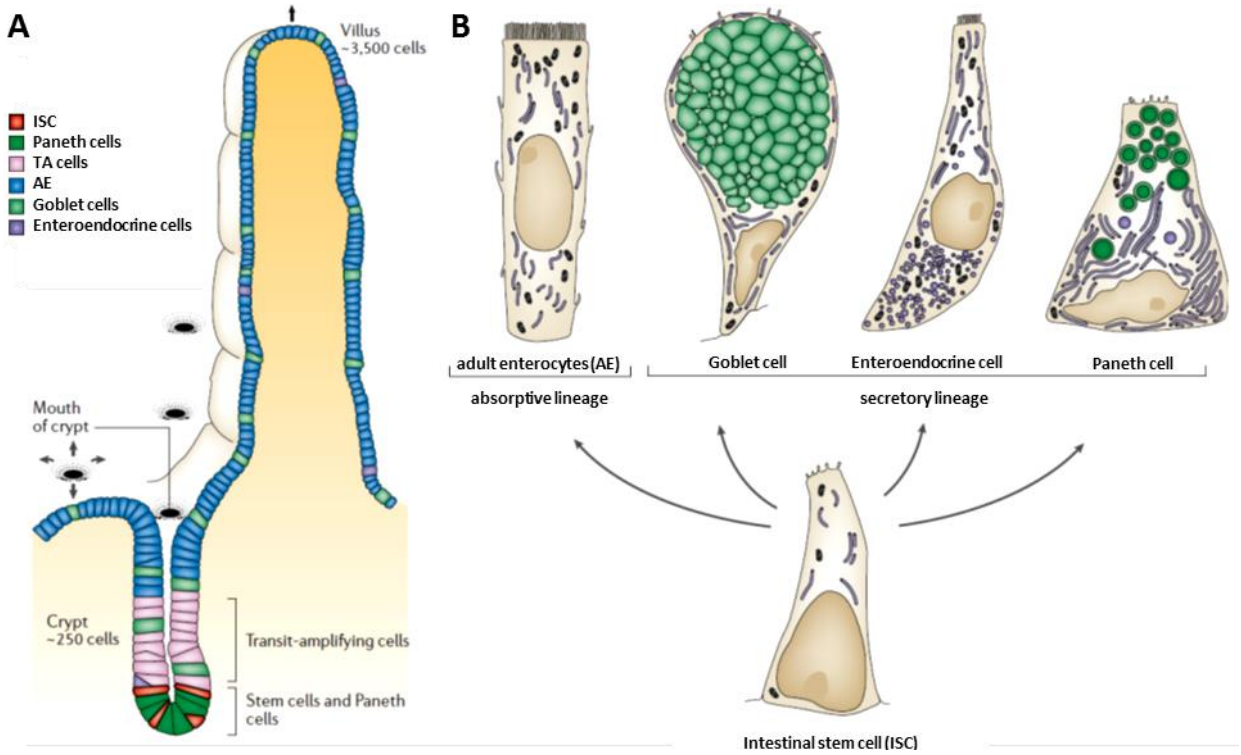


Figure 5: Cell types in the adult intestinal epithelium. *Adapted from Crosnier et al. 2006*[101]*.* ***A:*** *Crypt and villus structure and cell type distribution. Supported by the mesenchyme (orange) a polarized epithelial monolayer surrounds the intestinal lumen. All cell divisions of the adult intestinal epithelium take place in the crypt compartment, providing a steady supply of differentiated cells that move upwards into the villus compartment. In the villus, cells migrate further towards the tip and are discarded into the gut lumen after several days. The intestinal stem cells (ISC) are found at the bottom of the crypt, sharing this site with Paneth cells. Right above the ISC niche and closer to the crypt mound, transit-amplifying (TA) cells are found, the fastest proliferative cells in the intestinal epithelium. These cells differentiate as they move out of the crypt compartment and ultimately most of them become adult enterocytes (blue).* ***B:*** *ISC give rise to all epithelial lineages. As ISC divide, on average every second daughter cell becomes a TA cell and commits either to the absorptive or secretory lineage. The absorptive lineage contains adult enterocytes, which are responsible for nutrient uptake from the gut lumen, as the only cell type. The secretory lineage contains mucus secreting goblet cells, hormone secreting enteroendocrine cells and Paneth cells. The latter migrate not up- but only downwards to the crypt bottom, where they are involved in stemness maintaining signaling to ISC and contribute to pathogen defense by the secretion of antibacterial proteins.*

Stem cell maintenance and lineage commitment are mostly regulated by gradients of Wnt-, Notch-, BMP- and HH-signaling along the crypt-villus axis. Hereby the canonical Wnt-ß-catenin and Notch-signaling pathways at the bottom of the crypt maintain the stemness and self-renewal of the ISC daughter cells, while cells closer to the crypt mound sense decreasing Wnt in contrast to increasing Notch- and BMP signaling and hence develop to highly proliferative transit amplifying (TA) cells, which finally give rise to all non-dividing adult intestinal lineages[101,112-114]. Especially Wnt-signaling is crucial for proliferativity in the crypts, the repression of this pathway leads to cell differentiation and disappearance of crypts, while Wnt-overexpression leads to enlarged crypts[101]. To confine Wnt-induced proliferativity in late embryonic and

early postnatal stages to the crypt compartment, two HH ligands, Sonic (*Shh*) and Indian Hedgehog (*Ihh*), are expressed in the intervillus and diffuse through the mesenchyme to the villus compartment where they counteract Wnt-signaling (see ***Fig. 6A***). Full or partial inhibition of the HH pathway leads to complete disappearance of villi or ectopic Wnt-activation and respectively proliferation[102]. The main antagonist of the Wnt-pathway in neonatal and adult stages is the BMP-signaling from the mesenchyme to the epithelium, which is in turn positively regulated by HH[102]. BMP was shown to block adult cell proliferation and ectopic crypt formation in the villus by inhibition of the Wnt-pathway, while its action is blocked by the antagonist noggin, which is expressed around the crypt[112,115].

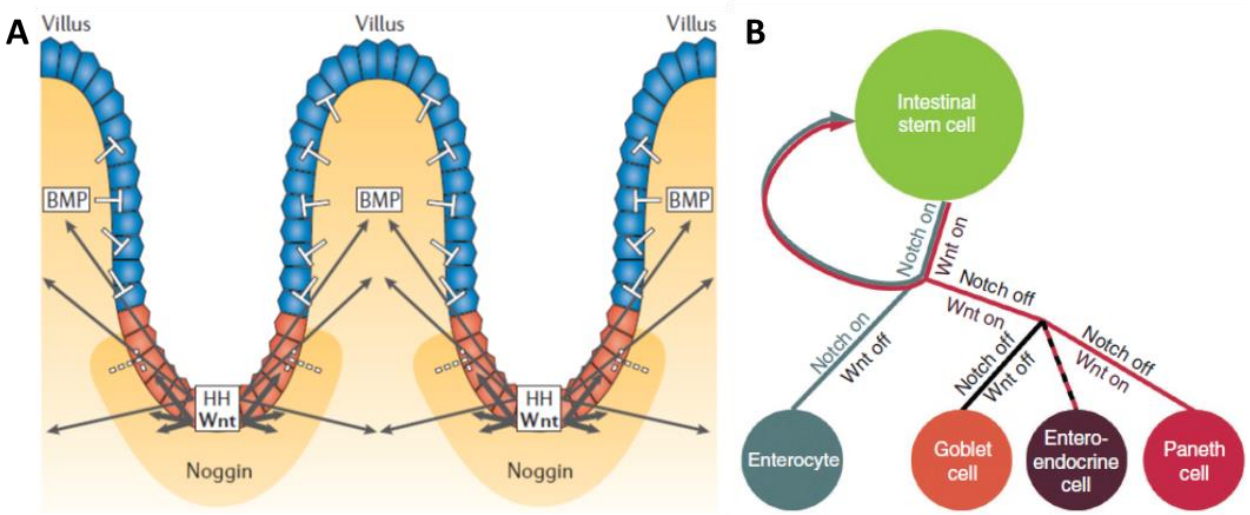


Figure 6: Signaling pathways determine intestinal architecture differentiation. *Adapted from Crosnier et al. 2006 and Yin et al. 2014[101,116]. **A:** Villus-intervillus organization by Wnt-, HH- and BMP-signaling. Short-range Wnt-signaling ensures epithelial proliferativity and stem-cell maintenance in the intervillus. While HH is highest expressed at the same position as Wnt, it is diffusing into the surrounding mesenchyme and activates BMP-expression. BMP in turn counteracts Wnt-signaling and restricts proliferativity. The suppressing effect of BMP- on Wnt-signaling is antagonized by Noggin, which is expressed around the intervillus pocket, hence preserving Wnt-signaling effects in this compartment. **B:** Notch- and Wnt- signaling regulate stem cell maintenance and cell fate commitment. Self-renewal of ISC requires the activity of both signaling pathways. Cells, which only sense Notch signaling, commit to the absorptive lineage and differentiate after several divisions to adult enterocytes. On the opposite, if Notch-signaling decreases, cells commit to the secretory lineage. The terminal differentiation to Goblet, Enteroendocrine or Paneth cells is hereby regulated by the level of Wnt-signaling.*

Besides their involvement in ISC self-renewal, Notch-signaling plays a major role in cell fate commitment (see ***Fig. 6B***). Proliferative cells evicted from the crypt bottom rapidly loose intestinal stemness markers, such as the Wnt-receptor Lgr5 (see ***Fig. 11***), and differentiate after approx. 5 division rounds[117]. Driven by Notch-signaling, most cells differentiate to adult enterocytes (AE) after migration upwards out of the crypt mound[113,114]. These cells are responsible for the absorption of nutrients from the gut lumen and their transport across the epithelium into the vascularized mesenchyme.

Around 10% of TA cells have low levels of Notch-signaling and differentiate to the secretory lineage, consisting of enteroendocrine, goblet and Paneth cells[113] (see ***Fig. 6B***). These cells are responsible for local and endocrine signaling, sustain the mucosal layer and produce defensins to protect the epithelium from infections[99]. This lineage is not restricted to upward migration as AE, and especially Paneth cells localize at the crypt bottom between ISC supporting their self-renewal by intercellular signaling, including the Wnt- and Notch-pathways[118]. As Paneth cells do not migrate to the villus tip and hence are not shed with AE into the gut lumen, their live expectancy is with several weeks substantially higher[119]. It is of special interest how Paneth cells remain fully differentiated in the presence of strong proliferative Wnt-signaling at the intestinal crypt bottom[120]. Paneth cells were shown to be essential for intestinal stem cell function by demonstrating that ISC can only efficiently form organoids *in vitro* in the presence of Paneth cells[109].

An open question remains, from which embryonic cell population the adult Lgr5+ ISC originate. While early embryonic stages look like a uniform monolayered cell population, only few cells give rise to the adult crypts, which ultimately become monoclonal[99,109,110]. One possibility is that *Lgr5* might be expressed in embryonic SC precursors all along, thus their cell fate predetermined from early embryonic stages. Contradicting this, unpublished data from our lab shows that the ISC-marker *Lgr5* is not expressed until e13.5 and hence cannot premark lineage commitment of early epithelial cells. Although the mechanisms of Lgr5 patterning in the early and late embryo are not yet fully understood, our results highlight the heterogeneity of the embryonic small intestinal epithelium and hence the necessity to start intestinal development research not just at the villus formation but already at early embryonic stages.

When investigating the development and homeostasis of the small intestine it is important to note that the posterior and anterior parts of it are similar but not equal, the anterior intestine shares properties with the stomach, while the posterior gut partially resembles the colon. For example the posterior small intestine harbors shorter villi then in anterior regions, which resembles the short and broad villi of the colon[121]. Also cell numbers of different cell types vary, with most Paneth cells in the intermediate and posterior small intestine and most enteroendocrine cells in the anterior part[121,122]. Previous studies[122,123] as well as our unpublished data show differential gene expression patterns along the anterior-posterior axis. This has to be remembered when studying cell populations, which are collected from the whole small intestine and hence show the average of this tissue. At the same time, single cell studies or imaging experiments should be performed in parallel on material from different intestinal regions to account for the mentioned differences.

3. Aims of this study

The elucidation of epigenetic mechanisms in gene regulation is crucial for the understanding of embryonic development, adult tissue homeostasis and cancer pathogenesis. As both, epigenetic mechanisms and the processes, they are involved in, are not only complex in themselves but also highly interconnected, this study was designed to deepen the understanding of these interactions.

The global aim of this study is to pinpoint the epigenetic states of genetic loci to their developmental functions and to unravel the involved epigenetic pathways' interactions.

In particular, in the following described projects I investigate the combined effects of the histone marks H3K4me3, H3K27me3, H3K27Ac and H2A.Z as well as DNA-methylation on gene activity. My goal is to understand three key points: Firstly, the genome-wide distribution and changes of active and repressing epigenetic marks throughout developmental stages of the murine small intestine. Secondly, the correlation of these marks to each other and their effect on gene activity depending on their position on the gene body or regulatory elements. And thirdly, the linkage of those marks to cell fate determination.

The toolkit of molecular biology is continuously expanding and diversifying, however for the raised questions it can be classified in two main approaches applied in this thesis: The genome-wide population approach employs pure, FACS-sorted or culture grown cell-populations, which are analyzed at the genome-wide scale for expression levels and several epigenetic marks in parallel. Complementary to this, the single-locus approach is suitable to determine the specific effect of an epigenetic mark at a clearly defined site and is especially powerful in combination with *in situ* single cell analysis.

4. Whole epigenome approach

To shed light on epigenetic mechanisms, a detailed genome-wide overview of epigenetic marks and mRNA-levels from early embryonic to terminally differentiated stages was generated. In the used murine model, the embryonic stages E12.5 and E14.5 were selected, because they represent the transition from a morphologically uniform to a specializing tissue. The proliferative ISC and two fully differentiated cell types, AE and Paneth cells, were used as well defined adult stages of the intestinal epithelium differentiation. FACS-purified populations of these two embryonic stages and three adult cell types were analyzed by chromatin immunoprecipitation (ChIP)-seq for histone marks, methyl binding domain (MBD)-seq for DNA-methylation and RNA-seq.

The previously available RNA-seq data allowed me to investigate differential gene expression and the identification of developmental markers, which in turn are targets for single locus approaches described below. The genome-wide distribution data of epigenetic marks allowed statistical correlation analysis with genomic features such as promoters, exons, regulatory elements, repeats etc. More so, combinatorial association of several marks with genetic and the (anti-)correlation to the targeted gene's expression were investigated.

Although by now multiple whole-genome datasets were published for an increasing number of epigenetic marks in different organisms, a comprehensive set for the embryonic and adult gut development of the small intestine is still unavailable. So for example previous data on DNA-methylation of ISC and AE showed nearly no differential methylation between these stages[32], however data for embryonic stages or the secretory differentiation pathway are yet to be published. It is also of high importance to compare genome-wide data of consistent samples, as the combination of data from different sources usually comes with substantial changes in sample origin, isolation and the following analytical performance, which introduces a strong bias. The here presented research overcame these problems by standardized tissue preparation protocols as well as a maximally unified sample processing and follow-up analysis (see Method section for details).

4.1 Whole genome datasets

All next generation sequencing datasets analyzed in this thesis were generated in the research group of Dr. N. Soshnikova, partially by me and partially previously by my colleagues with the kind support of the IMB core facilities cytometry, genomics and bioinformatics. For details see ***Table 2***. All input material for the used datasets was purified from the whole embryonic or adult small intestine by FACS as described in section 4.3.1. RNA-seq and MBD-seq inputs were not formaldehyde fixed in contrast to ChIP-seq material, but the remaining isolation by FACS was highly similar between the protocols.

dataset	**material**	**replicates**	**performed by**
RNA-seq	E12.5, E14.5, ISC, AE, Paneth cells	3	N. Soshnikova, S. Berl, L. Lavitas
MBD-seq	E12.5, E14.5, ISC, AE, Paneth cells	2	N. Soshnikova, L. Lavitas
ChIP-seq H3K4me3	E12.5, E14.5, ISC, AE, Paneth cells	2	C. Krienke, L. Lavitas
ChIP-seq H3K27me3	E12.5, E14.5, ISC, AE	5	J. Kazakevych, C. Krienke
ChIP-seq H3K27Ac	E12.5, E14.5, ISC, AE, Paneth cells	2	J. Kazakevych
ChIP-seq H2A.Z	E12.5, E14.5, ISC, AE	2	J. Kazakevych

Table 2: Genome-wide next generation sequencing (NGS) datasets.

4.2 Bioinformatic approaches

After ChIP-seq peak calling, three main analysis pathways were used. Firstly, the association of a chromatin mark with annotated genetic elements, namely TSS, introns, exons, 3'UTR, 5'UTR, intergenic regions and repeat elements was assessed. This allowed me to check if the epigenetic marks localization corresponded to previously reported loci, if the analyzed mark had novel positioning preferences and to select loci of interest for a closer look by the single locus approach (see section 5). Secondly, TSS were analyzed for the colocalization of different epigenetic marks and additionally correlated to the expression levels of the corresponding gene. While the active marks combination containing H3K4me3, H3K27Ac and H2A.Z correlated well to active TSS, genes with contradictory marks, e.g. H3K27Ac + 5mC at their promoter were selected for a closer examination by statistical methods and single gene analysis. And thirdly, the marks were followed through the four analyzed developmental stages and clustered for called promoter peaks based on their changes between the four stages. This differential analysis revealed not only genes that might be directly involved in intestinal stem cell maintenance, proliferation and differentiation to the secretory and absorptive lineages, but also gave clear indication at what stage which of the epigenetic marks are dynamic and hence are likely playing a role in cellular regulation.

4.3 Results: Whole epigenome approach

4.3.1 Purification of cell populations

Embryonic intestinal epithelium cells from the stages E12.5 and E14.5 as well as the adult ISC, AE and Paneth cell populations, were purified by careful dissection and tissue digestion, followed by FACS with specific fluorescently labeled antibodies, either targeting the cell type of interest, or the contaminant cell types to be excluded. Here I show the optimized sorting strategies and the validation of all used FACS-antibodies by single cell and tissue section imaging. The detailed protocols are described in section 7.2.

4.3.1.1 FACS strategies

Fixed (for ChIP-seq) or unfixed (RNA-seq) cell suspensions were FACS sorted to obtain pure populations with a minimal amount of contaminant cells, cell debris and cell doublets. Gating was performed on SSC-A and FSC-A to exclude cell debris and to focus on correct cell size. Subsequent gating for SSC-W and FSC-A excluded doublets. In case of unfixed cells, DAPI was used to stain and exclude dead cells. Finally, up to 5 fluorescently labeled antibodies were used to gate for the correct cell population (see ***Fig. 7-9***). Embryonic stages and adult enterocytes were sorted from wild type material based on the epithelial marker Epcam (see ***Fig. 7***). ISC were isolated from *Lgr5-GFP* mice. As shown in section 4.3.1.2, this fluorescent marker expression is very specific so that no additional antibody (Ab) staining was necessary (see ***Fig. 8***).

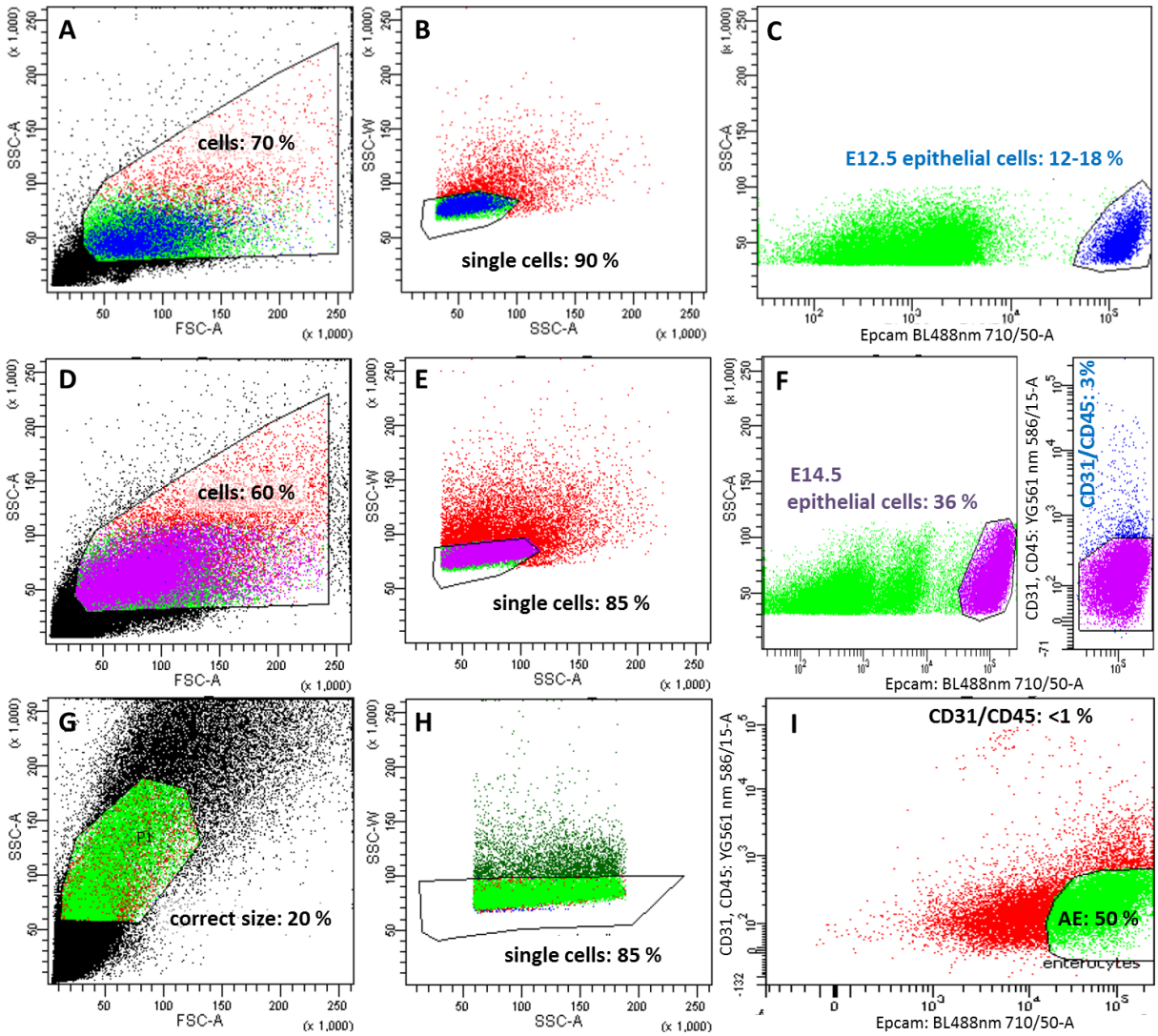


Figure 7: Epcam based FACS gating strategies. *Percentages indicate fraction of gated cells compared to parent gate.* ***A-C:*** *E12.5 epithelial cell sort.* ***D-F:*** *E14.5 epithelial cell sort.* ***G-I:*** *Adult enterocyte sort.* ***A,D,G:*** *cell debris and signals of too high/low size were excluded.* ***B,E,H:*** *Single cell gating to exclude doublets.* ***C,F,I:*** *Epcam positive epithelial cells were sorted to PBS for further analysis.* ***F,I:*** *CD45/CD31-positive lymphocytes and blood vessel epithelial cells were excluded. Markers used: Epcam – epithelial cells; CD31 – blood vessel epithelium; CD45 – lymphocytes.*

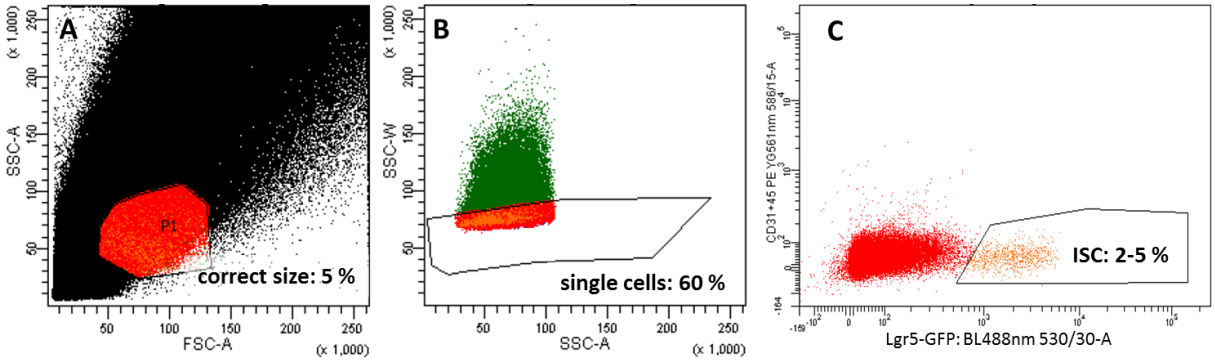


Figure 8: Lgr5-GFP based sorting of ISC. *Percentages indicate fraction of gated cells compared to parent gate.* ***A:*** *Cell debris and signals of too high/low size were excluded.* ***B:*** *Single cell gating to exclude doublets.* ***C:*** *GFP positive ISC were sorted to PBS for further analysis.*

Paneth cells were sorted with two positive markers CD24 and UEA1, as none of them was sufficiently specific by itself (see section 4.3.1.2 and ***Fig. 9***). Additionally, as for AE, cells were sorted for high Epcam and low CD31/45 signals. In comparison to other adult cell types, Paneth cells were stronger attached to neighboring (stem) cells and required harsher digestion prior to FACS. To further increase sorted population purity, an additional SSC-A / FSC-A gating for correct size was added, as the SSC-W / SSC-A gating did not efficiently recognize Paneth-ISC doublets due to the big difference in their cell size. Altogether, a population purity of 90-95 % was achieved for Paneth cells and > 98% for all other cell types.

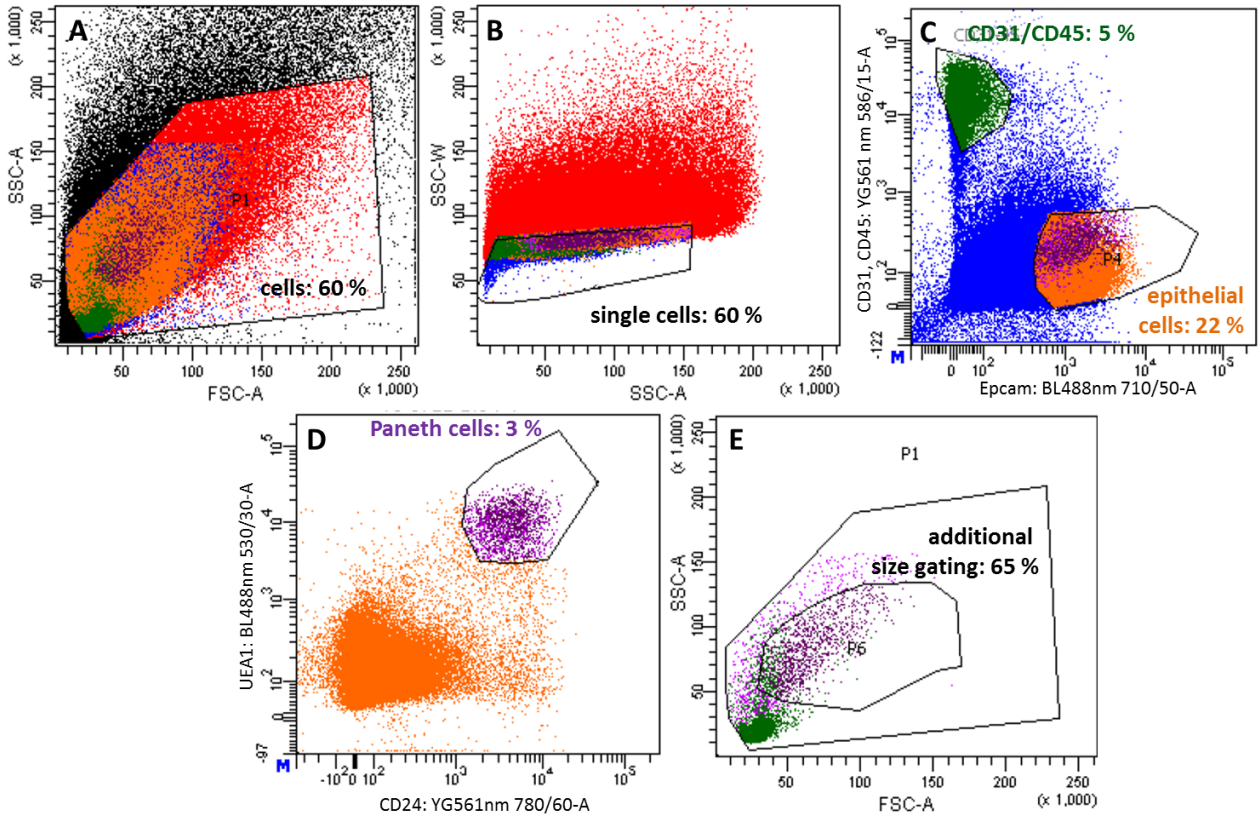


Figure 9: Paneth cell FACS. *Percentages indicate fraction of gated cells compared to parent gate.* ***A:*** *Exclusion of cell debris and preliminary size selection.* ***B:*** *Single cell gating to exclude doublets.* ***C:*** *Gating for Epcam positive and CD31/45 negative epithelial cells.* ***D:*** *Double positive labeling of Paneth cells.* ***E:*** *Additional size gating was applied to increase population purity. In green, CD45-positive lymphocytes are overlaid for size comparison. Markers used: Epcam – epithelial cells; CD31 – blood vessel epithelium; CD45 – lymphocytes; UEA1, CD24 – Paneth cells.*

4.3.1.2 FACS validation by microscopy

All fluorescently labeled antibodies for FACS as well as *Lgr5-GFP* expression, were validated by fluorescent microscopy to ensure specificity. Immunofluorescent stainings were performed either on cryosections or on intact, fixed and unfixed crypts, extracted by the usual cell harvesting protocol (see

section 7.2.1) without single cell digestion. Purity quantification and additional validation of Paneth cell sorting was performed by periodic acid-Schiff (PAS) staining of sorted single cells.

Epcam antibodies were tested on embryonic and adult cryosections by a secondary IF staining (see ***Fig. 10***). Embryonic epithelial cells were strongly stained, while the surrounding mesenchyme yielded very low signal. Adult epithelial cells in crypts and villi were also stronger stained than mesenchyme, however the signal ratio was not as good as for embryonic tissues. This corresponds to substantially better separation of Epcam+ from Epcam- populations in embryonic sorts in comparison to adult ones (see ***Fig. 7 and 9***).

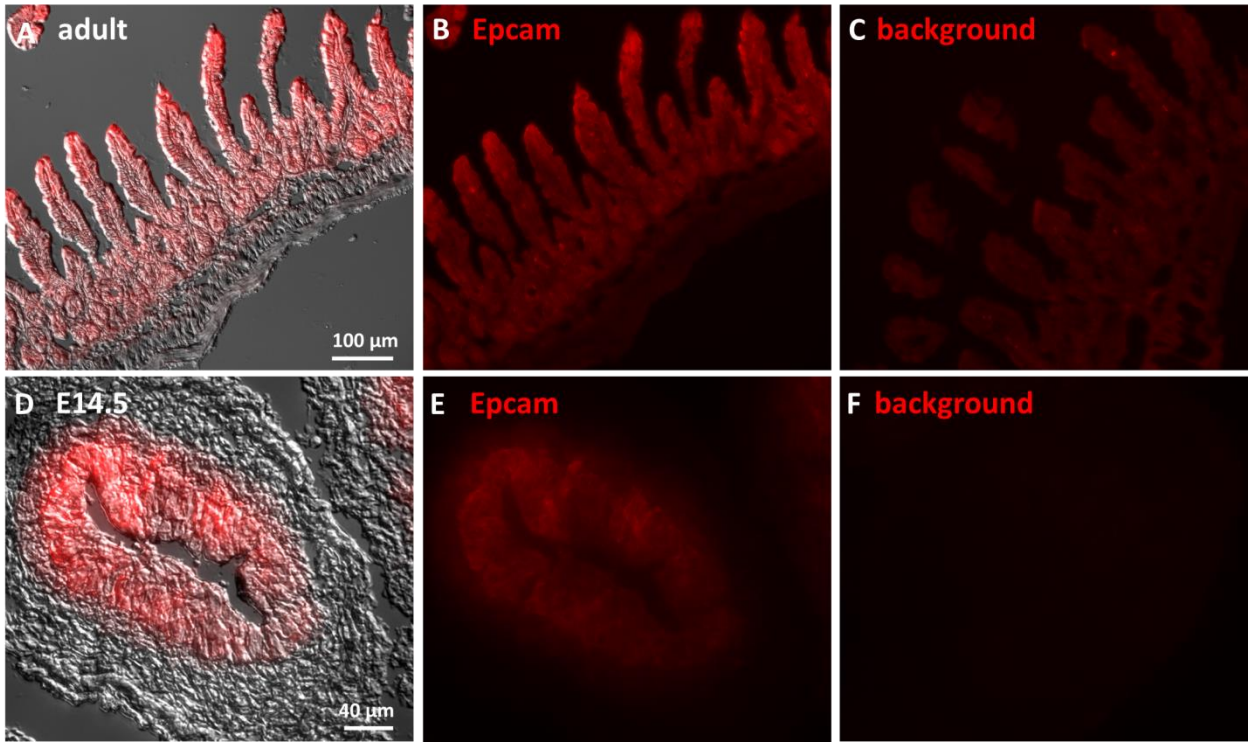


Figure 10: Epcam, secondary IF staining. *Cryosections, imaged by AF7000 widefield fluorescent microscope.* ***A-C:*** *Adult epithelial (top) and mesenchymal (bottom right) tissue.* ***D-F:*** *E14.5 epithelial (center) and mesenchymal (surrounding) tissue.* ***A,D:*** *Overlay of DIC channel with enhanced contrast fluorescent channel as in B and E.* ***B,E:*** *Secondary IF staining against Epcam. Primary Epcam FACS-Ab (eBioscience, 46-579182), secondary AF568 conjugated Ab (Invitrogen, A11077).* ***C,F:*** *Background controls for B and E without primary antibodies.*

ISC were sorted solely by GFP, expressed under the *Lgr5*-promoter. Specificity of this approach was confirmed by imaging of GFP in wholly isolated crypts (see ***Fig. 11***). Fixation of the crypts prior acquisition increased imaging quality. Additionally, secondary IF staining against GFP was performed, which was more sensitive, also targeting crypts with low GFP-levels. While direct GFP imaging visualized only ~10 %, secondary staining confirmed *GFP*-expression in at least 70 % of all crypts. As expected from crypt monoclonality, GFP levels were equal between the same crypt's base cells (except the negative Paneth cells)

and gradually diminished towards the TA cell region. Heterogeneity in GFP intensity between crypts explains the wide distribution of GFP+ signal in FACS, where low GFP signals blend in with the negative population (see ***Fig. 8***).

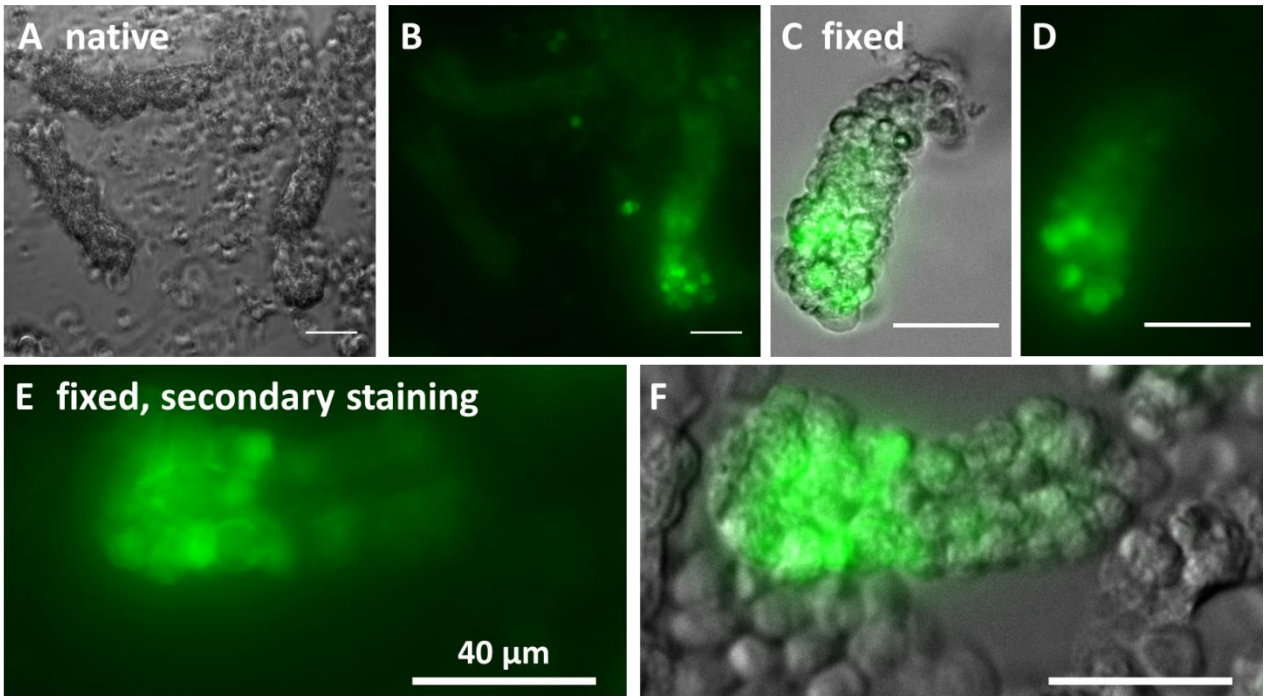


Figure 11: Lgr5 GFP localization in crypts. *Isolated crypts, imaged by AF7000 widefield fluorescent microscope. GFP signal in ISC at crypt bottom and decreasing signal toward the transit amplifying cell compartment in the upper crypt.* ***A***,***B****: DIC and GFP channels with 3 unfixed crypts, only one displays detectable GFP signal at crypt bottom.* ***C-F:*** *Formaldehyde fixed crypts.* ***B***,***D****: Fluorescent GFP signal.* ***E****: Secondary IF staining against GFP. Primary rabbit anti-GFP Ab (LifeTechnologies, A11122), secondary anti-rabbit Ab conjugated with AF488 (LifeTechnologies, A11070).* ***C***,***F****: overlay of DIC- and fluorescent channels.*

Paneth cell specific antibodies were validated both, by staining of intact crypts and on cryosections. Fixed crypts and adult tissue cryosections were directly stained by fluorescently labeled FACS-Ab against CD24 and UEA1 (see ***Fig. 12***). As expected, whole crypts showed strong fluorescent signal in the bottom compartment for both targets. On cryosections, UEA1 also yielded very specific crypt bottom signal with a few strong signals on villi, probably enteroendocrine or goblet cells. CD24 showed besides the crypt bottom signal also a comparatively high number of mostly mesenchymal, positive cells. In combination, both marks label Paneth cell very specifically. Additionally, the sorted cell population's purity is increased by CD31/45 exclusion, especially as cell harvesting protocols for FACS already remove most cells from villi and the mesenchyme (see ***Fig. 9*** and section 7.2.1).

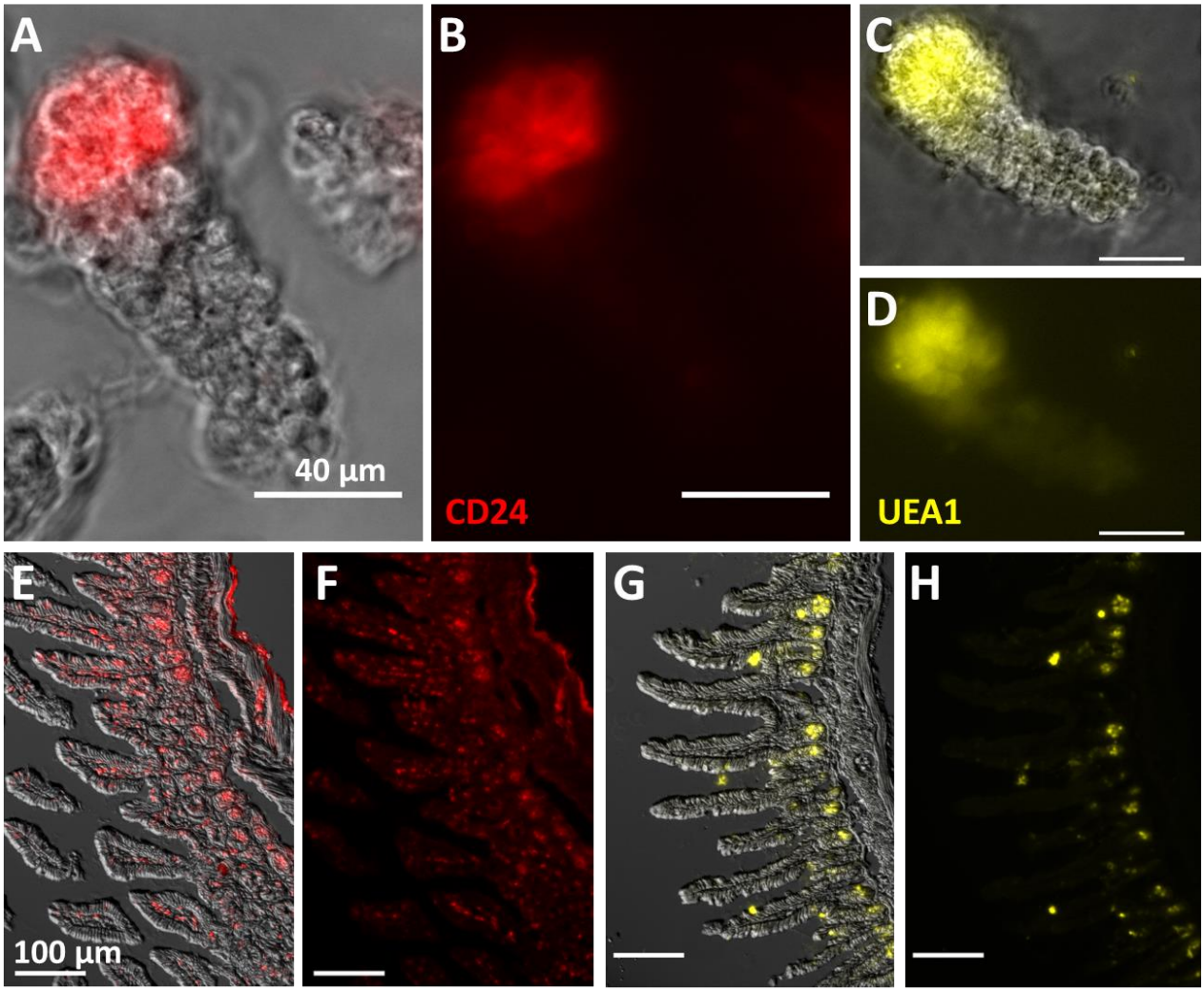


Figure 12: Paneth cell labeling with UEA1 and CD24 Ab. *Imaged by AF7000 widefield fluorescent microscope. Both CD24 and UEA1 signals localize predominantly at crypt bottoms.* ***A-D:*** *Isolated, formaldehyde fixed crypts.* ***E-H:*** *Adult wild type cryosections. Epithelial (left, center) and mesenchymal (top right) tissue.* ***B,F:*** *Primary IF labeling with anti-CD24 Ab conjugated to PE-Cy7 (Sigma, #560536)* ***C,H:*** *Primary IF labeling with anti-UEA1 Ab conjugated to TRITC (Sigma, L4889)* ***A,C,E,G:*** *overlay of DIC- and fluorescent channels.*

FACS-sorted Paneth cell populations were optimized for purity by numerous protocol adjustments. Population purity was quantified by PAS staining and single cell imaging, as exemplified in ***Fig. 13A***. Alternatively, sort purity was assessed on IF-labeled sorted cells, by comparing fluorescent (UEA1-FITC) and widefield channels (see ***Fig. 13B, C***). In the final protocol, a population purity of up to 95 % could be reproducibly reached.

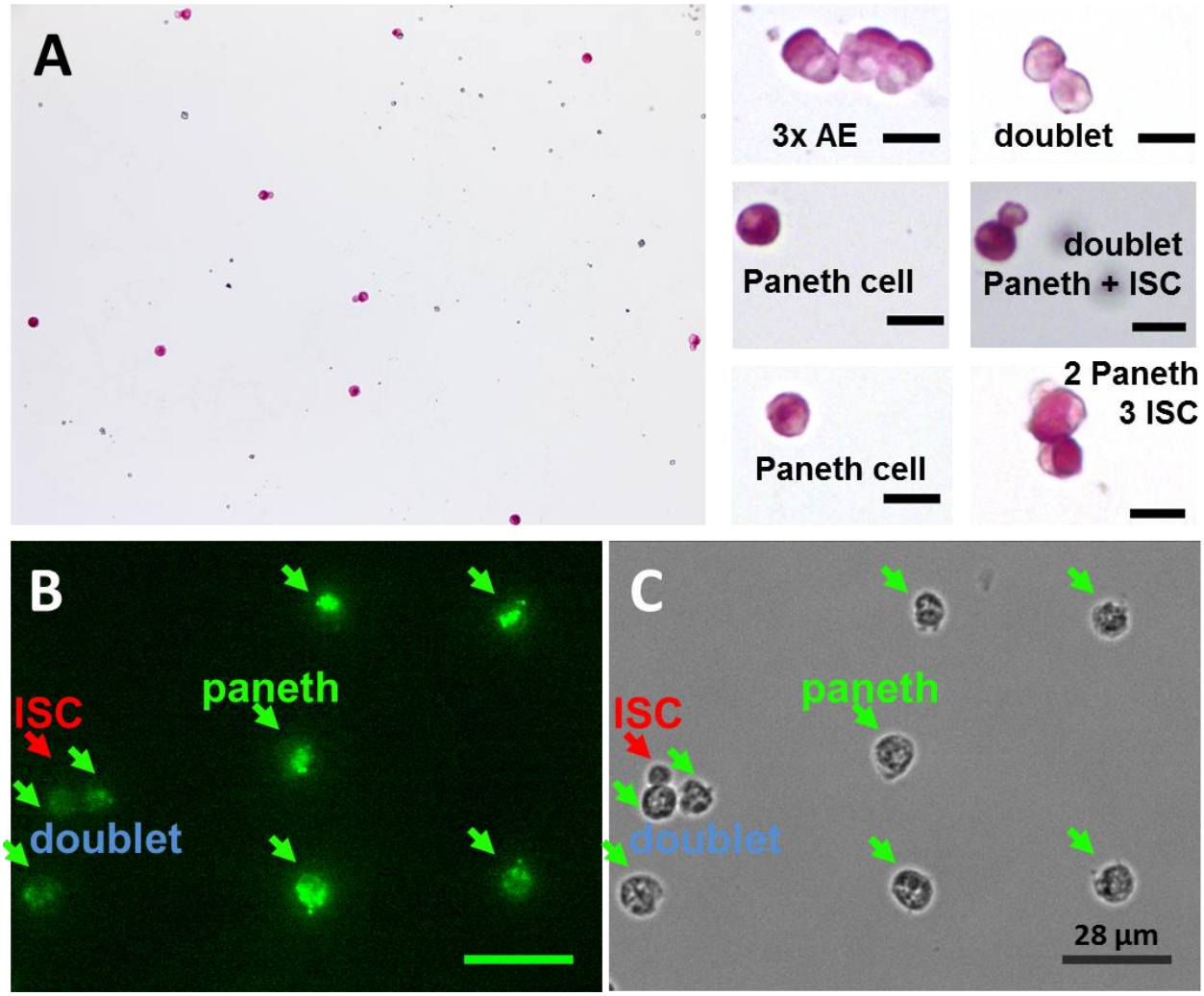


Figure 13: Population purity quantification. A: *PAS staining is positive in Paneth cells and brush borders at the luminal side of AE. left: Paneth cell sort output. Strongly diluted cells are imaged and quantified for cell type and singlet/doublet properties by shape and PAS-staining intensity. right: Examples of imaged singlets/doublets of different cell types.* ***B,C:*** *Fluorescence based quantification of Paneth cell population purity.* ***B:*** *FACS output, labeled with Paneth specific UEA1-FITC Ab (Sigma, L9006).* ***C:*** *Widefield channel, UEA1-negative cells (red) indicate contamination of the sorted cell population.*

4.3.2 ChIP-seq

In addition to the previously available datasets for H3K4me3 and H3K27me3 at the stages E12.5, E14.5, ISC and AE, I performed ChIP-seq for the H2A.Z and H3K27me3 marks at four stages and for the H3K27acetyl mark at five stages, including Paneth cells.

In general, each ChIP-seq experiment was performed in duplicates with one sequenced input per replicate. An exception is hereby the H3K27me3 ChIP-seq with 5 fully sequenced replicates, only 1-2 of which are suitable for peak calling and quantitative analysis.

4.3.2.1 H3K27Acetyl ChIP

Per replicate, 300,000 formaldehyde fixed and sorted cells were processed. The ChIP protocol was optimized for buffer compositions, antibodies, incubation times and sonication intensity. The major optimization was the addition of NaButyrate to most buffers, which preserves histone acetylation by inhibition of histone deacetylases (HDAC)[124]. Although long exposure might even change the acetylation pattern[125], the comparably quick initial steps of the used ChIP protocol, followed by chromatin isolation, prevent aberrant histone acetylation. While signal distribution was maintained, 20 mM NaButyrate treatment improved enrichment ratios measured by ChIP-qPCR from 2-fold (untreated) to 25-fold (see ***Fig. 14***). Hereby, total DNA recovery after ChIP of up to 25 % was reached, meaning that at a strongly H3K27acetyl marked locus, one fourth of the chromatin was precipitated. Hence, the H3K27acetyl ChIP reached very high efficiency (DNA-recovery) and specificity (fold-enrichment to negative controls).

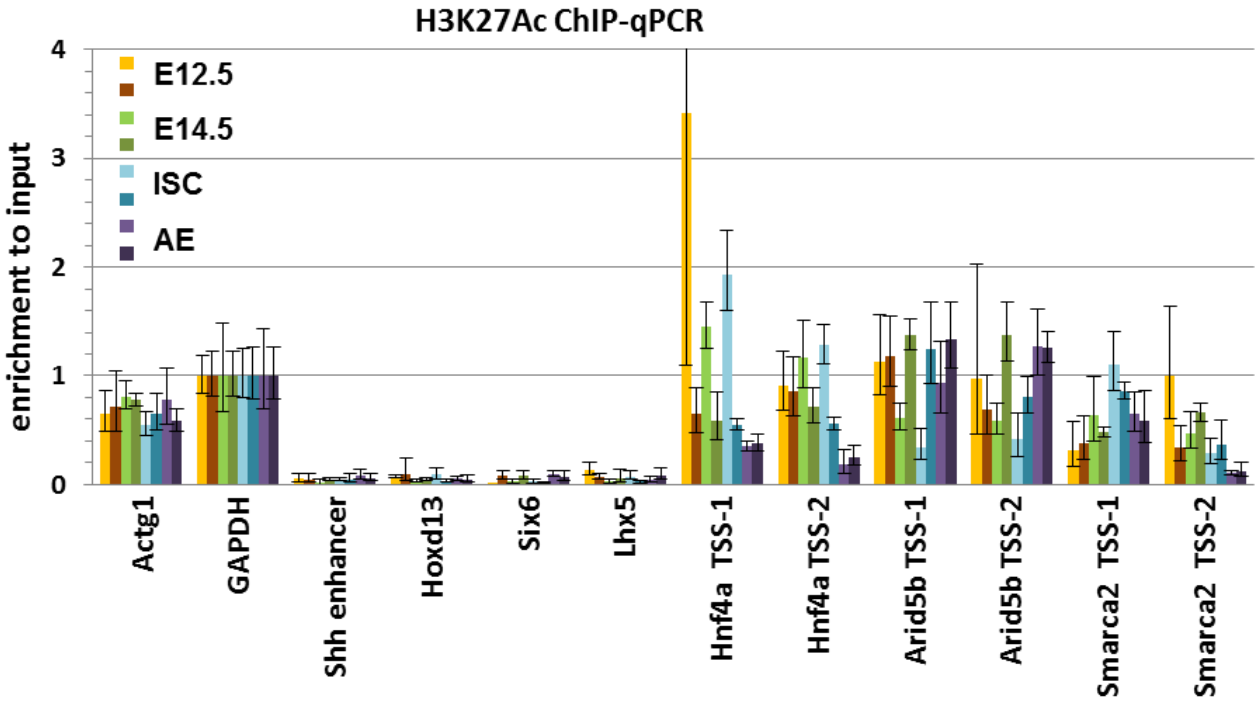


Figure 14: qPCR of H3K27Ac ChIP-seq replicates. *Enrichment against input, normalized to GAPDH. 300,000 FACS-purified cells were used per ChIP replicate of the developmental stages E12.5, E14.5, ISC and AE. Positive controls: promoters of Actg1, GAPDH (housekeeping genes). Negative controls: Shh-enhancer and promoters of Hoxd13, Six6, Lhx5 (not active in gut). Additional targets: Long(1) and short (2) TSS of differentially expressed and asynchronously epigenetically marked genes (see section 5.3). Pos.ctrl. enrichment to input is 15-fold higher in comparison to neg. ctrl. and additional targets are enriched up to 25-fold over neg. ctrl. Error bars show combined triplicate SD of ChIP and corresponding input samples.*

Paneth cell H3K27Acetyl ChIP duplicates were performed separately from the other 4 developmental stages and reached 12- and 15-fold enrichment over negative controls with over 10 % recovered DNA on strongly marked loci.

4.3.2.2 H2A.Z ChIP

Based on the H3K27Acetyl ChIP protocol, H2A.Z ChIP was established and sequenced at four developmental stages. A 12-fold enrichment of positive controls over negative controls with up to 4 % precipitated DNA on strongly marked loci was achieved (see ***Fig. 15***).

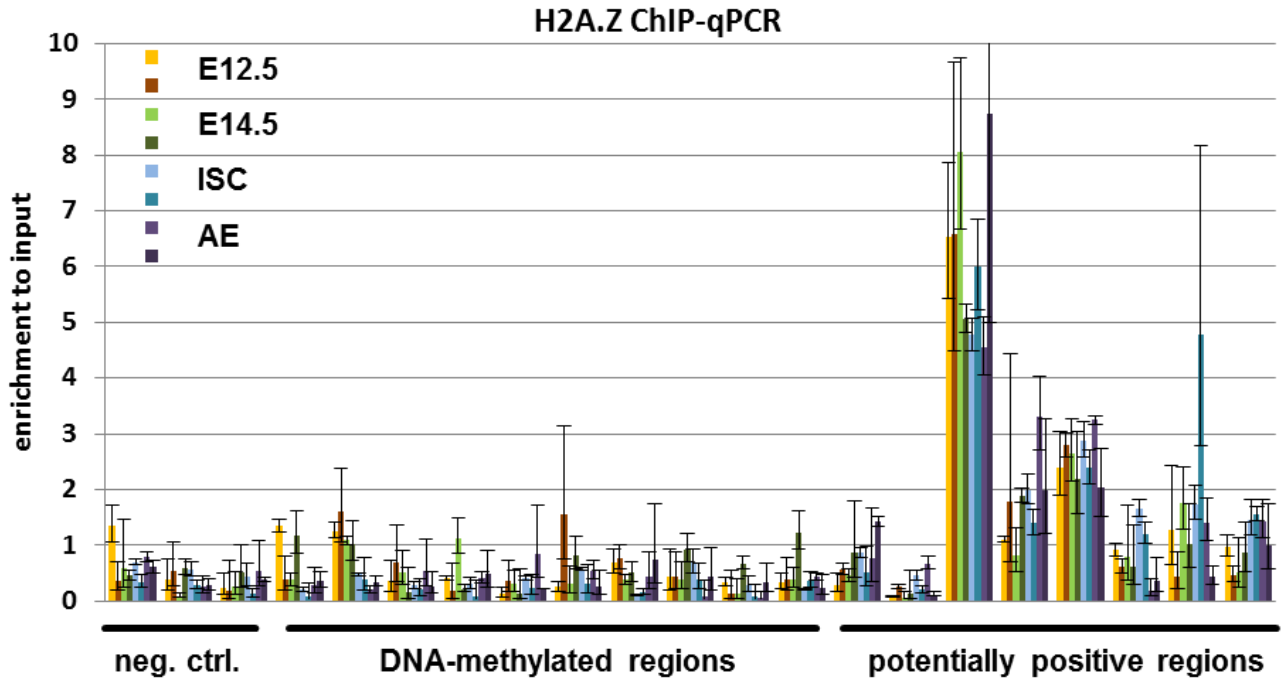


Figure 15: qPCR of H2A.Z ChIP-seq replicates. *Enrichment against input, normalized to all targets average. 500,000 FACS-purified cells were used per ChIP replicate of the developmental stages E12.5, E14.5, ISC and AE. Negative controls: promoters of Hoxd13, Six6 (silent genes), 5mC-positive regions of Shh, Mdk, Meis1, Foxa1, Axin2, Olfm4, Kcnq1, Prelp, Rnf43, Sema7 (5mC generally anti-correlates to H2A.Z). Potentially positive regions: Long and short TSS of differentially expressed and asynchronously epigenetically marked genes Hnf4a, Arid5b, Smarca2 and the promoters of Nbs1 and Stag1 (absence of 5mC and presence of active chromatin marksH3K27Acetyl and H3K4me3 indicate possible H2A.Z localization). Enrichment of positive target Arid5b TSS-1to negative controls' average is 12-fold. Error bars show combined triplicate SD of ChIP and corresponding input samples.*

4.3.2.3 H3K27me3 ChIP

Previously available data from our lab included one replicate of E12.5, ISC, AE and two replicates of E14.5 H3K27me3 ChIP-seq. In a first attempt to provide good quality duplicates for the whole dataset, I used the same protocol as used for the first replicates in a slightly optimized version. As can be seen in ***Fig. 20*** the ChIP-seq results were insufficient, showing only the strongest H3K27me3 peaks, such as Hox-clusters. Therefore, I conducted a more extensive H3K27me3 protocol optimization, including the successful ChIP protocols of H3K27Ac and H2A.Z ChIP, and sequenced one more replicate for all four developmental stages.

For the first two replicates 300,000 formaldehyde fixed and sorted cells were processed by the NP-buffer based protocol. On average, a 9-fold enrichment of positive controls over negative controls with up to 1.5 % precipitated DNA on strongly marked loci was achieved (see ***Fig. 16***). However, low signal controls, especially *Arid5b* showed half the signal intensity of strong controls, such as *Hoxd13*.

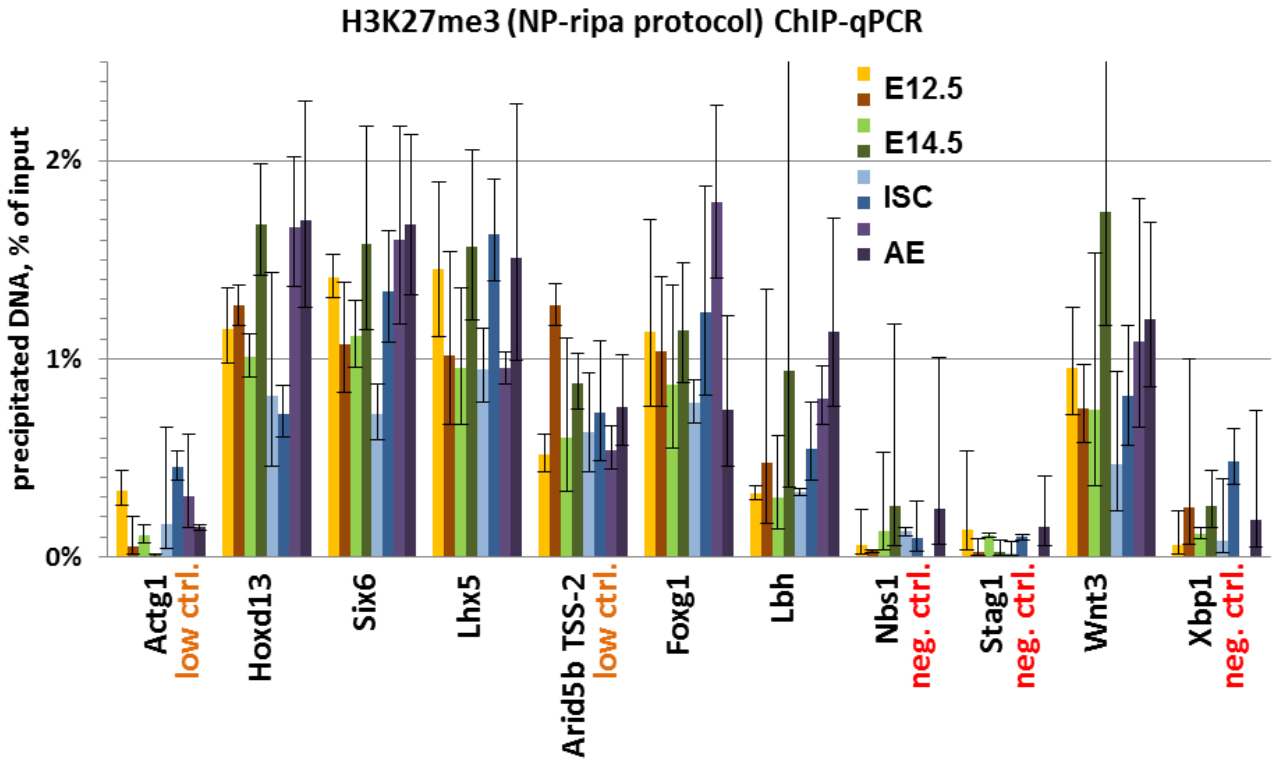


Figure 16: qPCR of H3K27me3 ChIP-seq replicates 1 and 2 prepared by NP-buffer based protocol. *Precipitated DNA as % of original input. 300,000 FACS-purified cells were used per ChIP replicate of the developmental stages E12.5, E14.5, ISC and AE. Positive controls: promoters of Hoxd13, Six6, Lhx5, Foxg1, Lbh, Wnt3 (H3K27me3 mark expected according to previous ChIP-seq results). Negative controls: promoters of Nbs1, Stag1, Xbp1 (no H3K27me3 mark detected previously, stably repressed genes). Low signal controls: promoters of Actg1 and Arid5b short (only low H3K27me3 signal detected previously). Enrichment pos.ctrl. vs. neg. ctrl. is 8-fold for 1st replicates and 11-fold for 2nd replicates, except of ISC 2nd replicate with 4-fold enrichment. Error bars show combined triplicate SD of ChIP and corresponding input samples.*

The next attempt for a 2nd high quality H3K27me3 ChIP-seq replicate was performed after extensive troubleshooting and protocol optimization. Different buffer compositions, antibodies, sample preclearing steps and bead amounts were tested. The usage of another antibody batch and minimal bead amounts turned out to be crucial for a reduced background signal and hence improved enrichment of positive regions vs. neg. ctrl. This time, 500,000 cells were used for each stage. The optimization was performed especially aiming to increase the ratio of positive regions vs. low signal promoters of *Arid5b* and *Actg1*. A 4-fold enrichment of positive regions vs. *Arid5b* for E14.5, ISC, AE and up to 12-fold enrichment in the E12.5 sample was observed (see ***Fig. 17***). This is an at least twice better enrichment than the first two H3K27me3

replicates. Enrichment to neg. ctrl. was 8-15 fold. The total DNA yield was lower, with <1 % precipitated DNA on strongest marked loci.

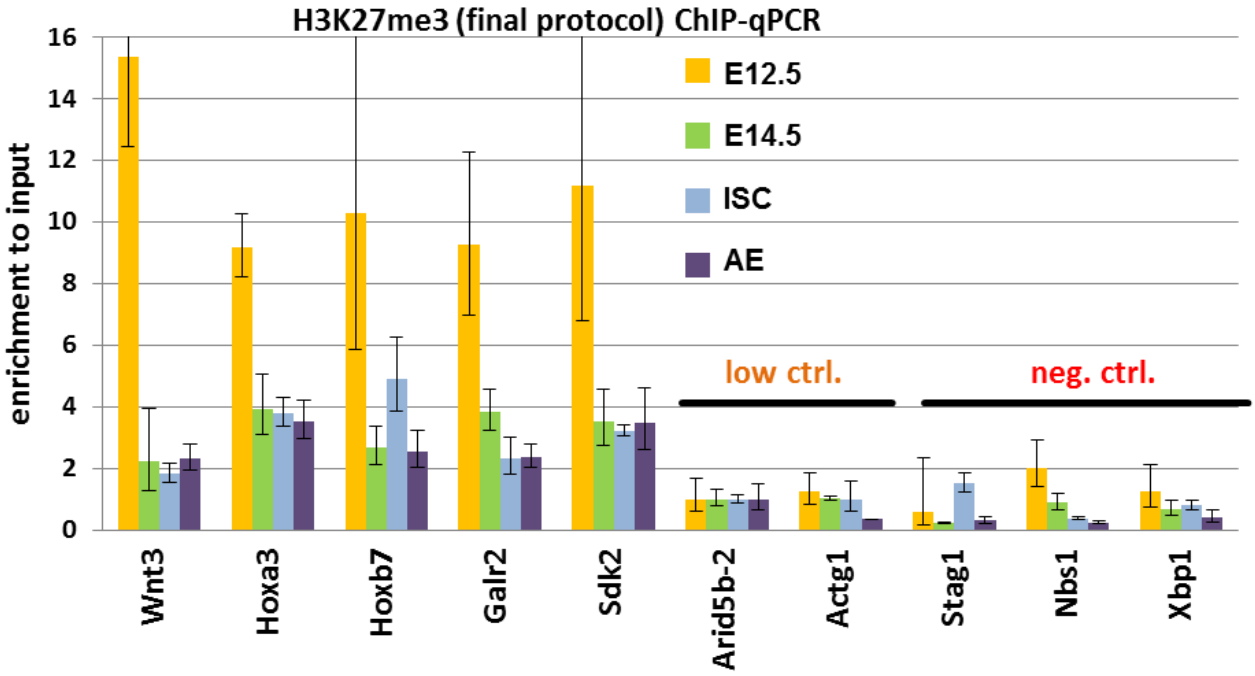


Figure 17: qPCR of H3K27me3 ChIP-seq 3rd replicate, prepared by the optimized protocol. *Precipitated DNA normalized to Arid5b signal. 500,000 FACS-purified cells were used per ChIP replicate of the developmental stages E12.5, E14.5, ISC and AE. Positive controls: promoters of Wnt3, Hoxa3, Hoxb7, Galr2, Sdk2 (H3K27me3 mark expected according to previous ChIP-seq results). Negative controls: promoters of Stag1, Nbs1, Xbp1 (no H3K27me3 mark detected previously, stably repressed genes). Low signal controls: promoters of Actg1 and Arid5b short (only low H3K27me3 signal detected previously). Error bars show combined triplicate SD of ChIP and corresponding input samples.*

4.3.3 Initial ChIP-seq processing

4.3.3.1 H3K27Acetyl ChIP-seq

Sequencing of H3K27Acetyl ChIP samples yielded 52-85 million reads per replicate. On all five stages, E12.5, E14.5, ISC, AE and Paneth cells, at least 40 million reads were successfully mapped with an average of 77 % mappable reads per replicate. Tracks show a very high reproducibility of ChIP and input replicates (see ***Fig. 18***). In general, input samples were interchangeable, so that peak calling could as well be performed with any single one of them without qualitative or quantitative loss.

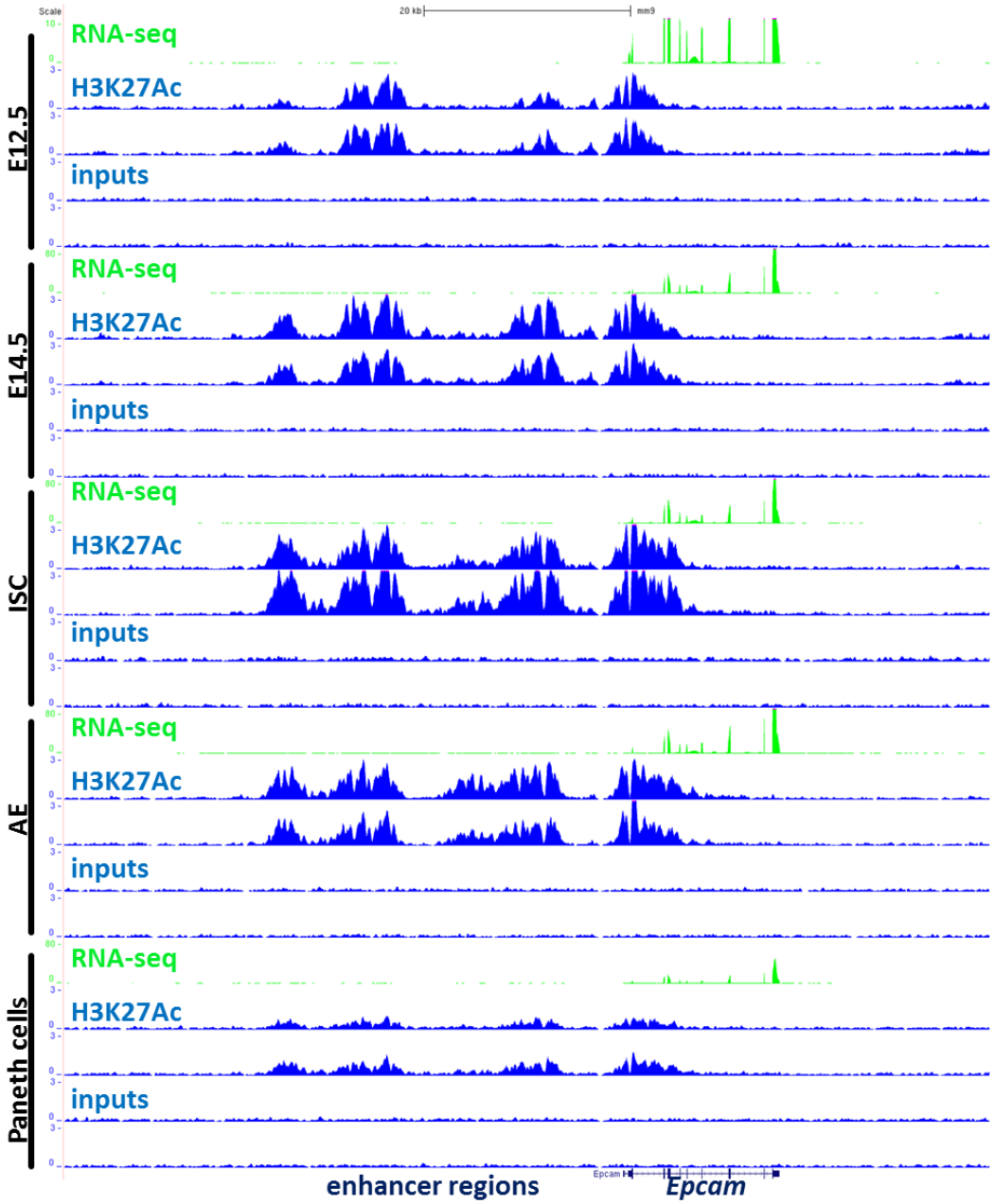


Figure 18: H3K27Ac ChIP-seq tracks. Epcam genomic region, including upstream enhancers. *For each stage (E12.5, E14.5, ISC, AE and Paneth cells) duplicates of ChIP and inputs were sequenced. RNA-seq tracks indicate gene expression levels, which generally correlate with the H3K27Ac mark at promoters and enhancers. Scale bar: 20 kb.*

Peak calling returned 36-53 thousands peaks per replicate with the exception of Paneth cell replicate 1, where only 5.6 thousand peaks were called. On visual inspection of the tracks, the low number of called peaks is linked to a lower signal to noise ratio of the replicate, so that while most peaks are still detectable by eye, only a fraction can be automatically called. In accordance to ChIP-qPCR results, Paneth cell ChIP-seq yielded lower enrichment to unmarked regions in comparison to the other 4 stages. However, at least in the 2nd replicate, the enrichment was sufficient to call the same amount (37 thousands) of peaks as in both AE replicates (36 thousands each).

4.3.3.2 H2A.Z ChIP-seq

Sequencing of H2A.Z ChIP samples yielded 47-65 million reads per replicate. On all four analyzed stages, E12.5, E14.5, ISC and AE, at least 34 million reads were successfully mapped for both ChIP replicates and at least one input replicate with an average of 75 % mappable reads per replicate. Tracks show a very high reproducibility of all ChIP and stage E12.5/E14.5 input replicates (see ***Fig. 19***). The inputs ISC_2 and AE_1 failed (see section 4.3.3.4 for details).

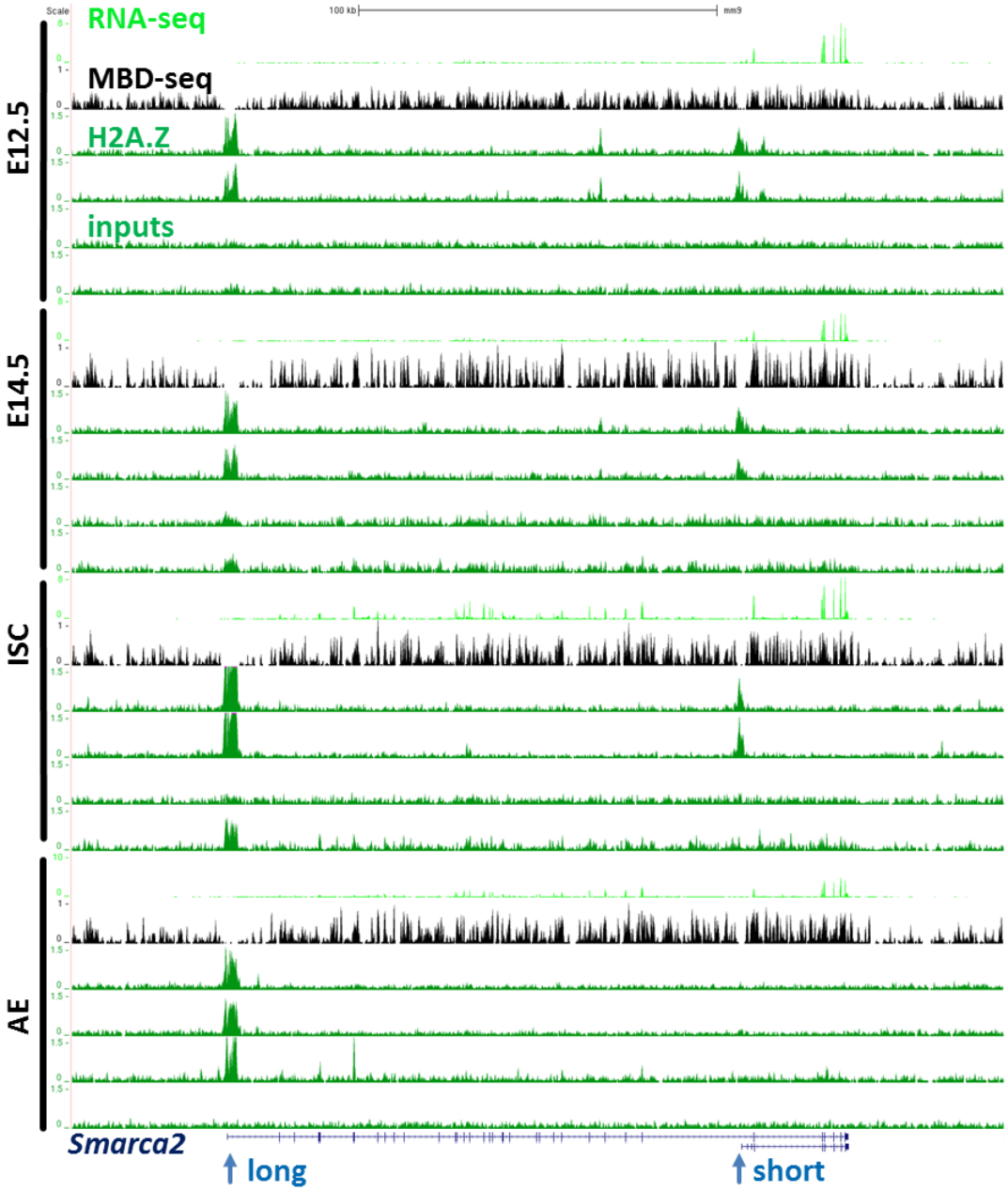


Figure 19: H2A.Z ChIP-seq tracks. *Smarca2 genomic region, including both TSS. For each stage (E12.5, E14.5, ISC and AE) duplicates of ChIP and inputs were sequenced. RNA-seq tracks indicate gene expression levels. DNA-methylation is shown by MBD-seq tracks, which generally anti-correlate with H2A.Z. Input replicates ISC_2 and AE_1 failed, showing a ChIP-like pattern on tracks (see section 4.3.3.4 for details). Scale bar: 100 kb.*

Peak calling returned 18-33 thousands peaks per replicate with the exception of the 2nd E14.5 replicate, where only 8 thousand peaks were called. On visual inspection of the tracks, the low number of called peaks is linked to a lower signal to noise ratio of the 2nd E14.5 ChIP and input replicates, so that while most peaks are still detectable by eye, only a fraction can be automatically called.

4.3.3.3 H3K27me3 ChIP-seq

Sequencing of H3K27me3 ChIP samples yielded 41-61 million reads in the 1st/2nd replicates and 36-41 million reads in the 3rd replicate. On all four analyzed stages, E12.5, E14.5, ISC and AE, at least 32 million reads for ChIP and input of the 1st/2nd replicates and 24 million reads for the 3rd replicate were successfully mapped, with an average of 74 % mappable reads per replicate. Exceptions to that were the failed input samples, E14.5_2, E14.5_3 and ISC_3, where only 43-60 % of reads could be mapped (see section 4.3.3.4 for details). Tracks show varying enrichment to background but reproducible peak positioning at each stage (see ***Fig. 20***).

In accordance with varying ChIP enrichments measured by qPCR and visualized by the tracks (see ***Fig. 16, 17 and 20***) as well as due to inconsistent input quality, peaks could be called only on a subset of samples. No peaks could be called for any samples from the 1st and 2nd replicates. In the 3rd replicate only sample E12.5 returned 8.300 peaks, which is an improvement in comparison to previous data, where at this stage 6.400 peaks were called. Other replicates in previous ChIP-seq experiments had yielded 7.3 and 11 thousand peaks for E14.5, 23 thousand for ISC and 5 thousand peaks for AE. Although the peak calling of H3K27me3 ChIP-seq is mostly disappointing, the presence/absence of this mark can be visually assessed on the generated tracks of at least one replicate per stage.

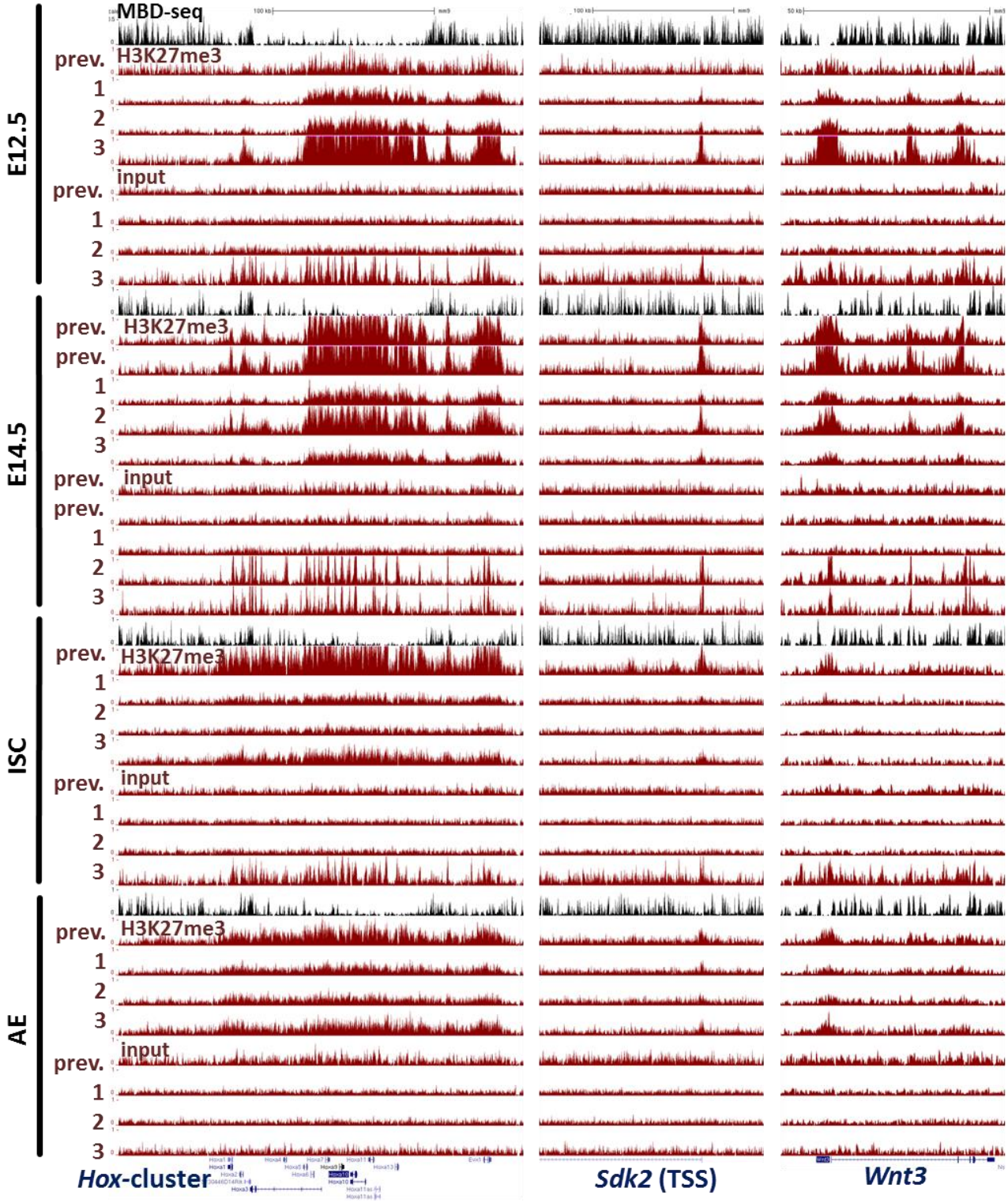


Figure 20: H3K27me3 ChIP-seq tracks. *HoxA-cluster, promoter region of the Sdk2 gene and Wnt3. For each stage (E12.5, E14.5, ISC and AE) 3 replicates of ChIP and inputs were sequenced. The 3rd replicate was prepared by an improved protocol. Additionally, previously available H3K27me3 ChIP-seq data (prev.) for all four stages is shown (only successful replicates). DNA-methylation (black) generally anti-correlates with H3K27me3. Failed input replicates E12.5_3 E14.5_2, E14.5_3 and ISC_3 show a ChIP-like pattern on tracks (see section 4.3.3.4 for details). Scale bars: 100, 100, 50 kb.*

4.3.3.4 Failed input replicates

In the H2A.Z and H3K27me3 ChIP-seq experiments, several input samples (H2A.Z: ISC_2 and AE_1, H2K27me3: E12.5_3, E14.5_2, E14.5_3 and ISC_3) failed. Instead of yielding an evenly distributed background signal, the reads mapped in a ChIP like signal, with mostly promoter localizing peaks (see ***Fig. 19 and 20***). Also, most of the affected input replicates showed with 43-60 % a substantially lower read mappability in comparison to >71 % for all other inputs. Taken together, this indicates a contamination with mammalian but non-murine DNA, possibly human cell line IP material. With regard to the reduced mappability, promoter region peaks on the failed inputs' tracks might emerge due to higher genomic conservation of these regions and hence higher mappability of non-murine reads.

As ChIP material and the successfully processed input replicates were processed in parallel with the affected input samples throughout the 3 completely independent ChIP-seq experiments, contamination during the sample preparation is very unlikely. No contamination of any involved reagents was detected and no non-murine eukaryotic DNA was handled during the time period of these experiments.

As a possible contamination source library preparations were identified by Bioanalyzer profiles. While input samples showed a regular size distribution before and after library preparation, all affected input samples showed an additional 350 bp peak after library preparations (see ***Fig. 21***). However, same reagents, including sequencing adaptors, were also used for unaffected samples and hence their contamination could not be confirmed.

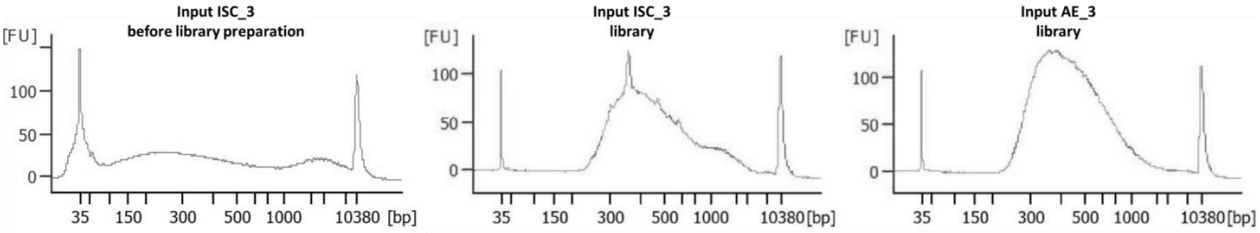


Figure 21: Input size profiles before and after library preparation. *Measured by Bioanalyzer HS DNA chip. H3K27me3 ChIP-seq input material.* ***Left/center:*** *failed input replicate ISC_3.* ***Right:*** *before library preparation. Center: after library preparation. An additional 350 bp peak is observed across failed input replicates and indicates contamination of the sample. Right: successfully processed input replicate AE_3.*

Although all steps of ChIP and library preparation protocol were checked for possible contaminations and partially also repeated independently, the contamination source could not be tracked back. As only input but no ChIP samples were affected, most bioinformatic processing could instead be performed with parallel replicates of the affected inputs. So e.g. peak calling of H2A.Z samples AE_1 and AE_2 were both performed with the input replicate AE_2. As processed input samples represent the general mappability and library preparation efficiency across the genome, no substantial differences are expected between input replicates. This was confirmed by all mentioned ChIP-seq experiments with two or more successful input replicates per stage (see ***Fig. 19-21***).

4.3.4 Genome-wide distribution of epigenetic marks

All assessed histone marks, H3K27Ac, H2A.Z, H3K4me3 and H3K27me3, showed predominant signal and peak localization on genic regions, especially promoters. The preferences for specific regions varied strongly depending on the epigenetic mark, while staying mostly constant throughout the developmental stages (see ***Fig. 22***). While >75 % of H3K4me3 peaks were found at promoters, over 50% of all H3K27Ac peaks are distributed on intergenic and intronic regions. This is in line with expectations, as the H3K4me3 mark is only known as an active promoter mark, while H3K27Ac is also associated with active enhancer elements. H2A.Z is with approx. 50 % promoter peak localization between the first two marks, which corresponds to its positioning on TSS but also enhancers and gene bodies. Because of varying ChIP-seq signal quality, H3K27me3 peak distribution is least consistent over developmental stages. Not more than 50 % of peaks localize at TSS ±3 kb and intergenic peaks yield a similar fraction to H2K27Ac and H2A.Z. This indicates not only promoter limited but widespread positioning of H2K27me3, such as on Hox-clusters.

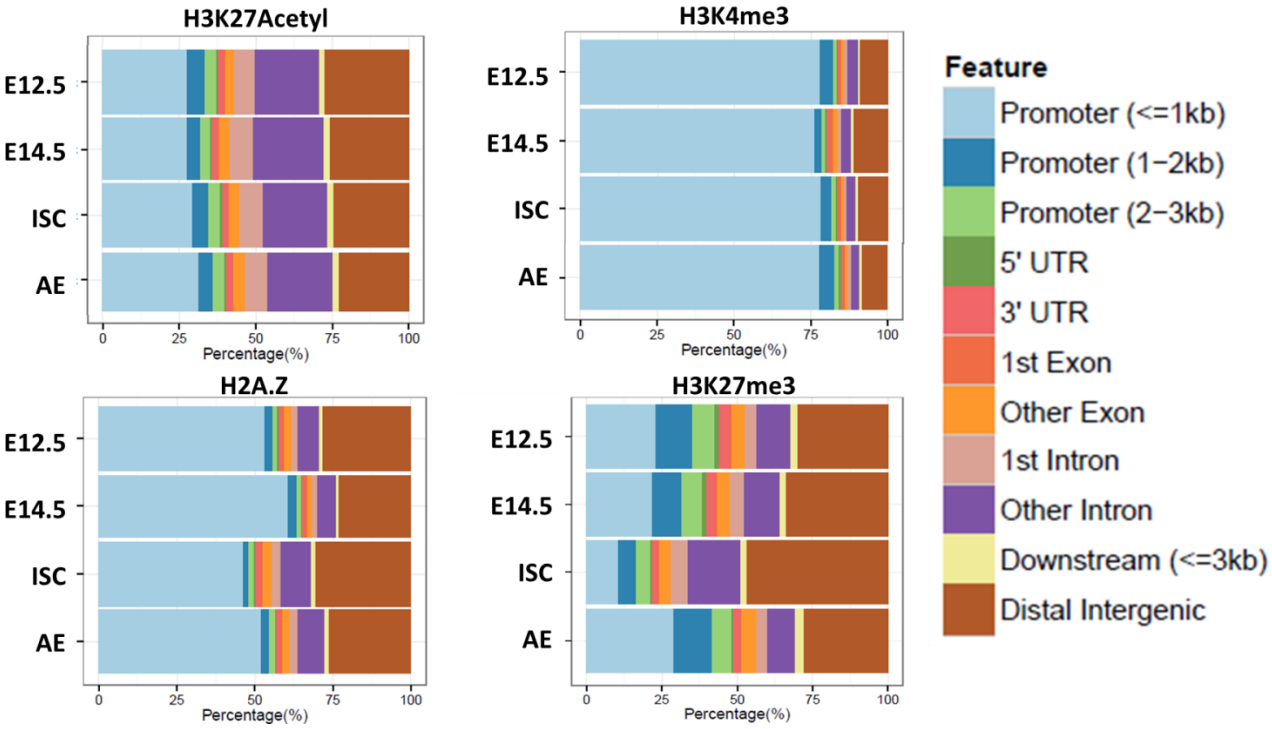


Figure 22: Peak distribution of epigenetic marks by genomic features. *One replicate per epigenetic mark and developmental stage is shown. Shown replicates were selected for best ChIP-seq quality and called peak number.*

All 4 histone marks showed a symmetrical distribution at TSS in all developmental stages. Hereby, H2A.Z and H3K4me3 had sharpest peaks, while H3K27Ac and H3K27me3 distributed broader around TSS (see

Fig. 23). All marks, especially H3K27Ac, often distributed far into the gene body, however with much lower ChIP-seq intensities in comparison to the TSS proximal region. TSS peaks of all marks consist of 2 summits neighboring the TSS symmetrically, however often they merge into one central peak due to limited resolution (see ***Fig. 23C***). The variation in resolution is presumably based on technical replicate variations, such as sonication efficiency, however, this should not affect downstream analysis, as two-summit TSS were generally called as one peak.

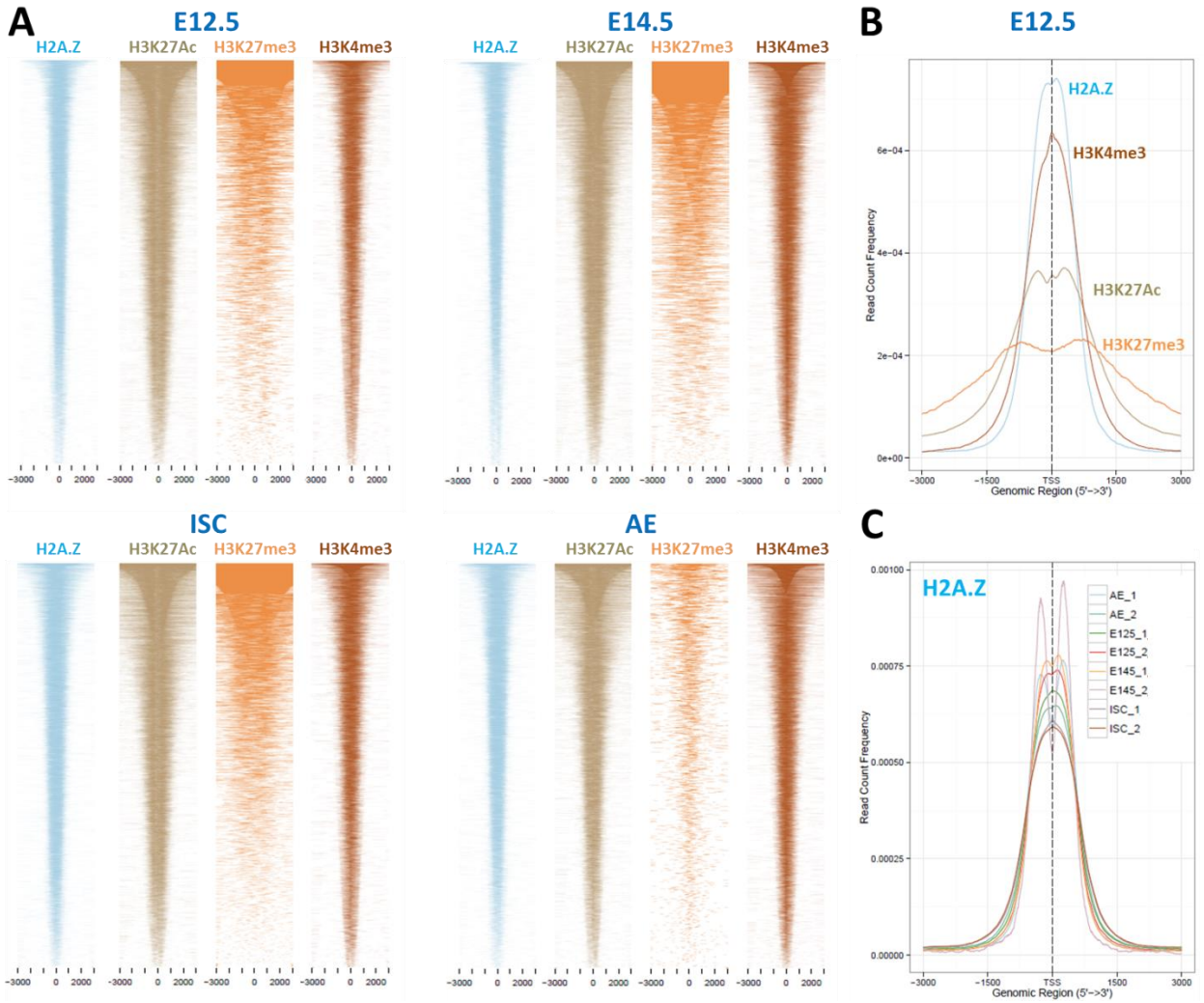


Figure 23: Histone mark distribution at TSS. A: *H2A.Z, H3K27Ac, H3K27me3 and H3K4me3 intensity maps centered at peak called TSS at 4 developmental stages. One replicate shown for each mark and stage.* ***B:*** *Average ChIP-seq signal density around TSS at stage E12.5.* ***C:*** *H2A.Z ChIP-seq signal density at TSS of all replicates show either single- or double-summit peaks.*

4.3.5 Correlation and exclusion of epigenetic marks

4.3.5.1 Active histone marks colocalization on active TSS

With RNA-seq of all assessed developmental stages, quantitative data on gene activity was available. After setting a threshold to cut off non- or low-expressed genes, lists of top 10,000 active genes' TSS were

generated and compared with H3K27Ac, H3K4me3 and H2A.Z TSS top 10,000 peaks (see ***Fig. 24***). Generally, in the top 10,000 TSS marked with any single one of those active epigenetic marks, 74-81 % would simultaneously be marked by both other active marks. This strong correlation indicates cooperative action of H3K27, H3K4me3 and H2A.Z as well as possible interconnected deposition and regulatory mechanisms of these marks. Furthermore, 85 % of triple positively marked TSS belong to the top expressed genes at the given developmental stage.

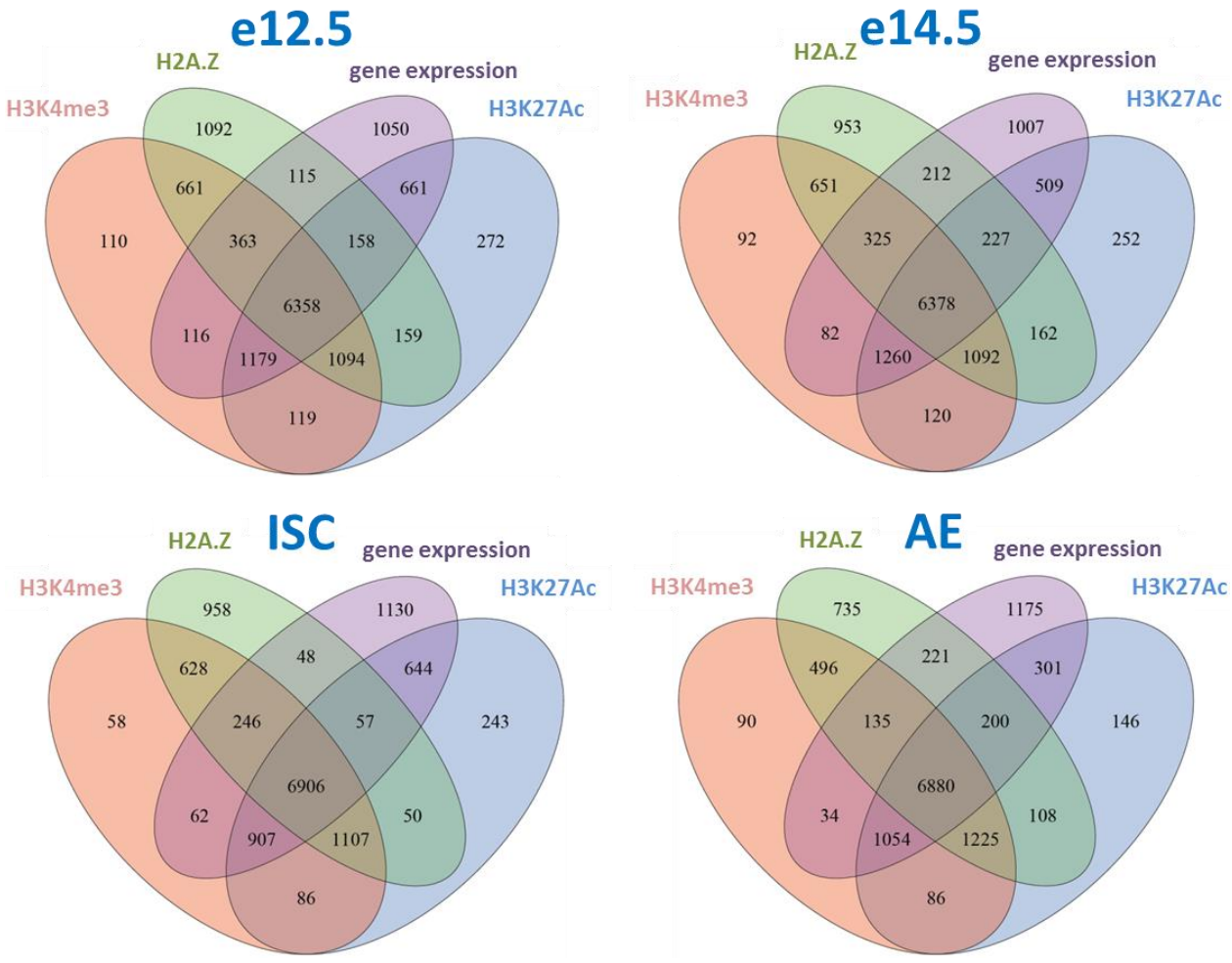


Figure 24: Histone marks and gene expression. *Lists of 10,000 strongest TSS peaks and highest expressed genes were compared independently for each developmental stage. No weighting for position inside the list was performed. Numbers show absolute size of gene subsets with the annotated marks and/or high gene expression.*

In the top 10,000 expressed genes 64-70 % are marked with all active histone marks. Out of them H2A.Z shows least contribution to high expression, as additional 9-13 % expressed genes are found with the H3K4me3 + H3K27Ac combination, while the combinations H2A.Z + H3K4me3 and H2A.Z + H3K27Ac contribute only 1-4 % and 1-2 % of expressed genes respectively. H3K27Ac mark seems to be the only single mark contributing substantially to gene activation: On average 5 % of all expressed genes are marked

by H3K27Ac only, while only 2 % of active genes are marked by any other single mark. Interestingly, >10 % of all active genes have none of these histone modifications/variants at their TSS.

Regarding the shown correlations, it should be kept in mind that they are threshold based. While the overall correlations are qualitatively very clear, thresholding introduces a certain level of statistical noise, which might mask or artificially increase minor effects, because "close to threshold" genes fall into subset lists near randomly. Quantitative analysis of TSS marks and gene expression correlation would rather require a weighted and probability based but hence also less comprehensible approach.

4.3.5.2 Active histone marks' deposition in accordance to gene activity

On aligning ChIP-seq promoter peaks to expression levels of the corresponding genes, the histone marks showed preferential occupation. H3K27Ac and H3K4me3 mark promoters of strong and intermediate expressed genes, while H2A.Z mark intermediate and weaker genes, being mostly absent at the topmost expression gene subset (see ***Fig. 25***). All 3 histone marks strongly anti-correlate with DNA-methylation, which is mostly absent in strongest expressed genes, gradually increases towards weaker promoters and is especially present on completely repressed ones. At these silent promoters only a small fraction of the 3 active marks is found and interestingly, also here the active histone marks anti-correlate strongly with DNA-methylation.

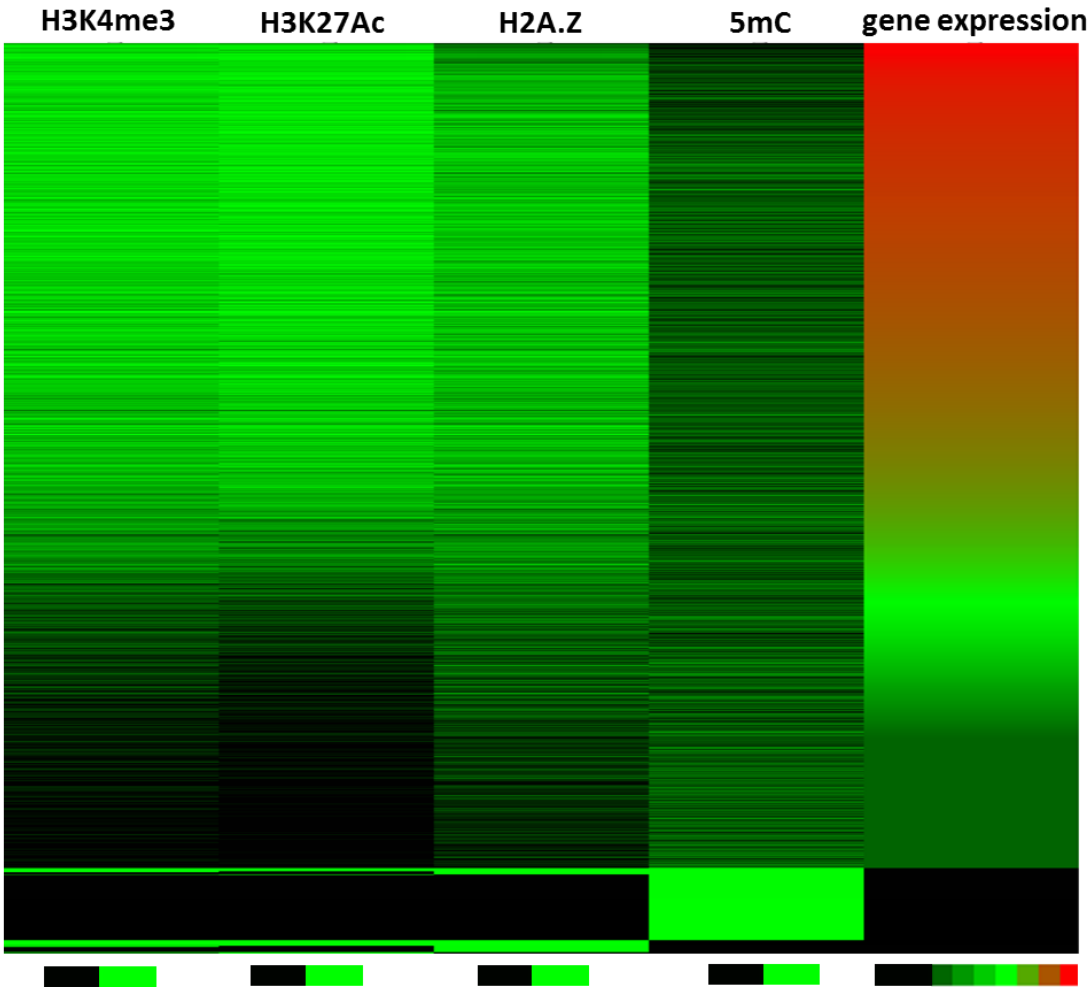


Figure 25: Histone marks and DNA-methylation preferential deposition in regard to promoter activity. *The heatmap shows binary distribution of top 10,000 ISC promoter peaks of H3K4me3, H3K27Ac and H2A.Z, as well as all called promoter MBD-seq peaks, sorted by rpkm (reads per kilobase per million mapped reads)-normalized gene expression levels. Other developmental stages show a similar distribution. Columns 1-4: black – absent, green – present. Column 5: black – silent gene, green/red gradient – gene expression levels. TSS without detected expression or any of the 4 epigenetic marks are not shown.*

4.3.5.3 H2A.Z colocalizes with H3K27Ac and excludes DNA-methylation genome-wide

Strong colocalization of H3K27Ac and H2A.Z, as well as mutual exclusion of H2A.Z and DNA-methylation at TSS, were observed as expected due to their (anti-)correlation to active genes' promoters. Additionally to this, ChIP- and MBD-seq showed also the same behavior of these three marks in intergenic regions (see ***Fig. 26A***). In genes, i.e. on TSS, exonic and intronic regions, 5'- and 3'-UTRs, 87 % of H2A.Z peaks coincide with H3K27Ac. In intergenic regions this is still true for 64 % of all called H2A.Z peaks. In contrast, the mutually exclusive marks H2A.Z and 5mC overlap only at 11 % of all gene localized peaks and 20 % in intergenic regions (see ***Fig. 26B***).

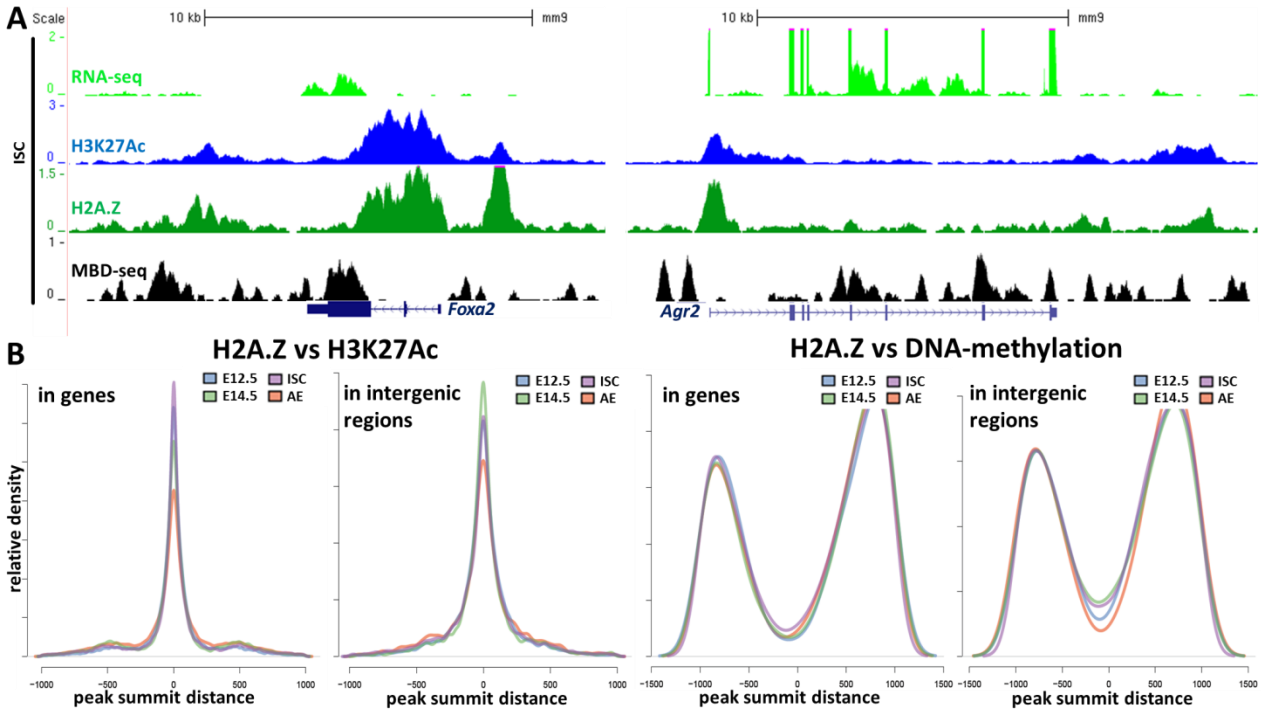


Figure 26: H2A.Z, H3K27Ac and 5mC genomic localization. A: *ISC RNA-seq, ChIP-seq and MBD-seq distribution on and around the Foxa2 and Agr2 genes. On both, weakly and strongly expressed genes as well as in intergenic regions, H3K27Ac and H2A.Z peaks overlap, while DNA-methylation is excluded. Scale bars: 10 kb.* ***B:*** *Peak summit distances of H2A.Z to H3K27Ac (left) and DNA-methylation (right), on gene bodies including promoters (left) and intergenic regions (right). All 4 developmental stages show consistent distribution.*

4.3.6 H3K27Ac marks poised enhancers

In addition to the expected deposition of H3K27Ac at active enhancers and promoters (see ***Fig. 18***), ChIP-seq performed with the HDAC-inhibitor NaButyrate showed intermediate levels of this mark also at poised enhancers (see ***Fig. 27***). The ISC marker *Olfm4* is was shown to be not active until postnatal development (unpublished data from our group), which means that H3K27Ac is deposited at the *Olfm4* enhancer at least 2 weeks before its activation. The enhancer region shows already a mild exclusion of DNA-methylation, which however is not completely removed until the full enhancer/promoter activity in ISC. A similar positioning at poised promoters was not observed.

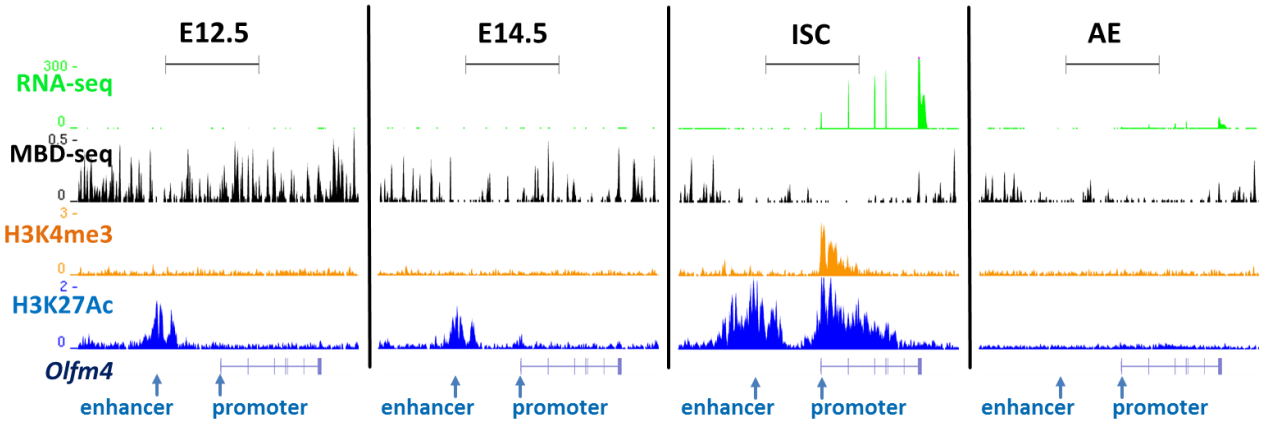


Figure 27: H3K27Ac marks enhancer prior activation. *E12.5-AE RNA-seq, ChIP-seq and MBD-seq tracks of the Olfm4 locus. As indicated by RNA-seq and the active promoter mark H3K4me3, The ISC marker Olfm4 is strongly expressed in ISC cells, while virtually absent at other stages. At embryonic stages the locus is repressed by DNA-methylation, which is removed completely from enhancer and gene body until ISC stage. In addition to a strong enhancer and promoter deposition at the active ISC stage, H3K27Ac is also found at intermediate levels on the poised enhancer at least from stage E12.5. Scale bars: 20 kb.*

4.3.7 H3K27Ac and H3K4me3 mark active transposable elements

H3K27Ac, H2A.Z and H3K4me3 showed a constant distribution on repetitive elements throughout the developmental stages. An exception of this was observed with H3K4me3, which showed an increasing localization preference on LTR elements at later stages (see ***Fig. 28***). The increasing number of H3K4me3 marked transposable elements corresponds to their DNA-demethylation and hence activation at adult stages. Although no substantial increase of the H3K27Ac peak fraction on LTR elements was observed, peak intensities clearly rise for both marks with LTR activation. Based on these observations, the combination of H3K4me3 and H3K27Ac on non-promoter sites might generally predict the activity of transposable elements.

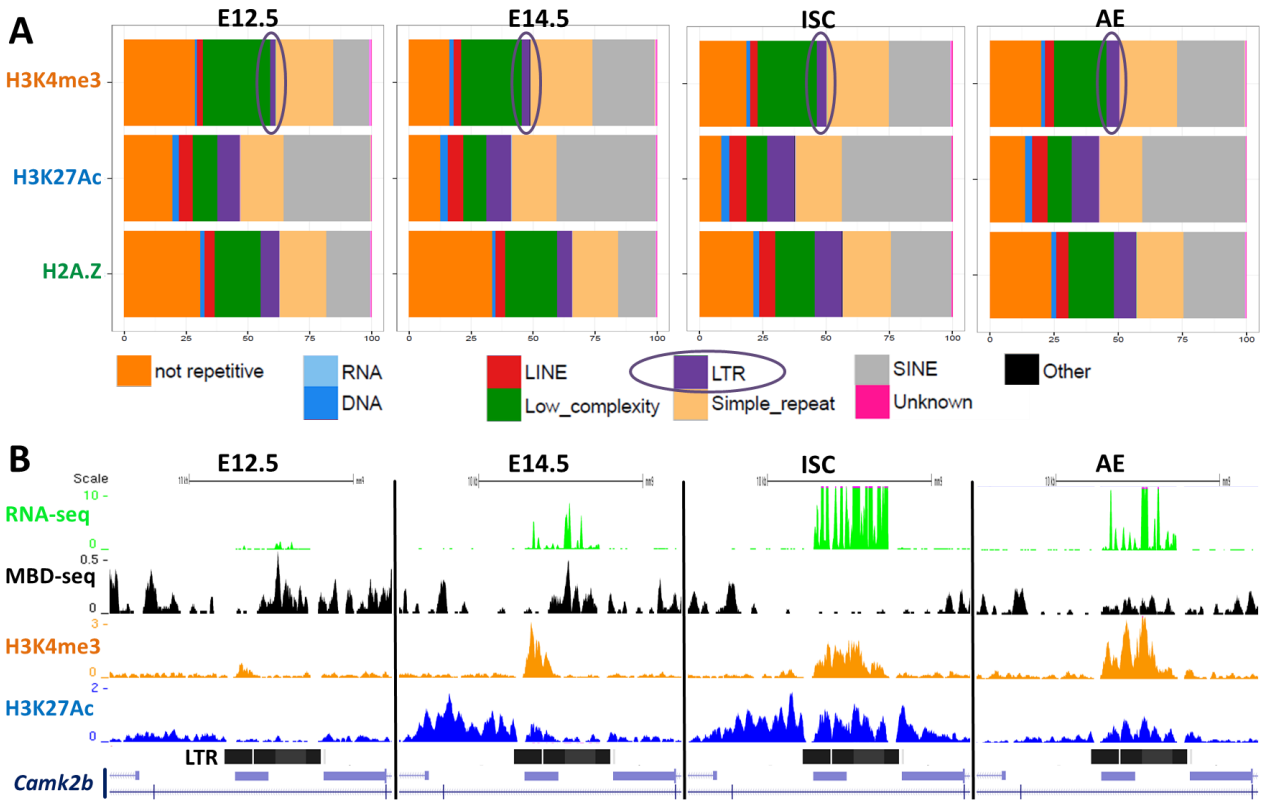


Figure 28: H3K4me3 marks in combination with H3K27Ac active transposable elements. A: *Peak distribution of epigenetic marks on repetitive elements. One replicate per epigenetic mark and developmental stage is shown. Repetitive elements are color-coded. LTR elements are indicated as the only repetitive element substantially changing in occupation by H3K4me3.* ***B:*** *E12.5-AE RNA-seq, ChIP-seq and MBD-seq tracks surrounding the LTR locus inside the Camk2b gene. At embryonic stages the locus is repressed by DNA-methylation, which is removed in ISC. In ISC and to lesser extent in AE, the LTR is active, which is shown by high RNA-levels, and marked by the active chromatin marks H3K27Ac and H3K4me3. Scale bars: 10 kb.*

4.3.8 Dominant epigenetic enhancer states

Combined RNA-seq, MBD-seq and histone ChIP-seq data clearly show, that active marks at the promoter region are not necessarily sufficient for gene expression and can be completely overruled by the enhancer state, as exemplified in ***Fig. 29*** on the *Foxa1* gene. The active promoter marks H3K4me3 and H3K27Ac persist until the fully differentiated AE stage and are even complemented from the ISC stage on with H2A.Z, also an active promoter associated mark. However, gene expression ceases at adult stages, matching not the epigenetic promoter state but the replacement of active enhancer marks H3K27Ac and H2A.Z by DNA-methylation. The decreasing expression of *Foxa1* at stage E14.5, before enhancer changes are observed, points to the involvement of additional epigenetic mechanisms in gene activity regulation. Although the correlations of epigenetic marks on promoters and enhancers to gene activity are clear, it is unlikely a one-way regulatory pathway, so that differential positioning of epigenetic marks might well be not the cause of but caused by gene expression changes.

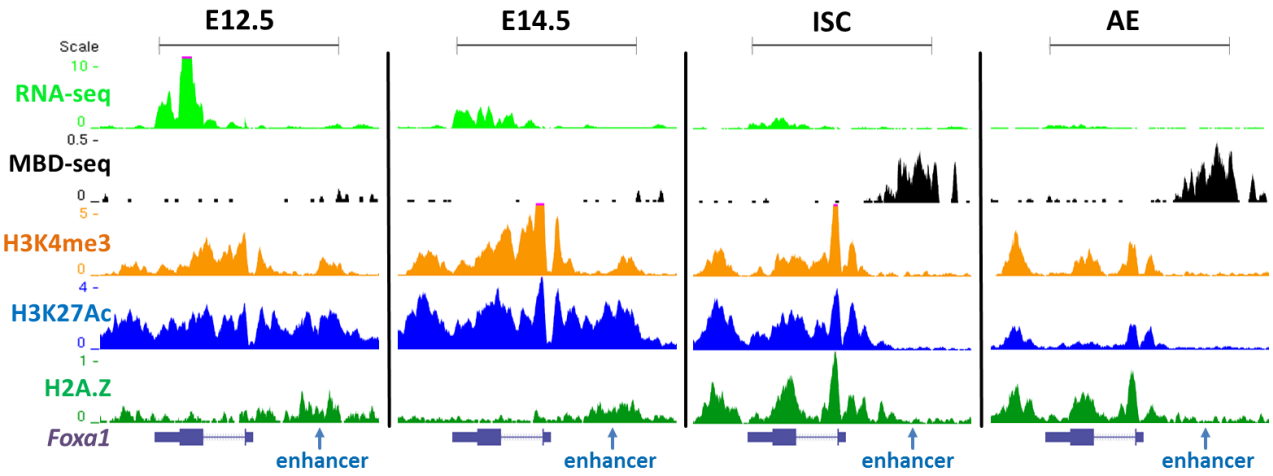


Figure 29: Contradicting enhancer/promoter states. *E12.5-AE RNA-seq, MBD-seq and ChIP-seq tracks of the Foxa1 gene and enhancer locus. At embryonic stages the gene is active, no DNA-methylation is observed on the promoter and enhancer regions, active histone marks are found on both. At adult stages the enhancer is repressed by DNA-methylation, while active enhancer marks are depleted. This matches the decreased gene activity, although the promoter is still actively marked. Scale bars: 10 kb.*

4.3.9 Gene clusters

For a comprehensive analysis of developmental gene activity and epigenetic state changes, the genome-wide datasets were clustered. Although the ChIP-seq data sets contain also intronic and intergenic peaks, the clustering was performed only on promoter regions to allow correlation analysis with the clustered RNA-seq data. All datasets were clustered for fold-changes (increasing, decreasing or equal) between the 4 developmental stages E12.5, E14.5, ISC and AE, resulting in 27 clusters. To simplify the analysis and to reduce cluster assignment noise, 27 clusters were combined to a 9 cluster system, depicting only changes between the stages E14.5, ISC and AE.

4.3.9.1 RNA-seq clustering

RNA-seq clustering was based on 2-fold changes in gene expression. As expected, the largest cluster contains silenced genes and stably expressed housekeeping genes, which together sum up to over 15,000 or 67 % of all annotated genes in the analysis (see ***Fig. 30***). The clustering method was confirmed by the annotation of well known cellular markers, as e.g. *Lgr5* and *Olfm4* to the stemness clusters E12.5<E14.5<ISC>AE and E12.5=E14.5<ISC>AE, *Shh* to the embryonic cluster E12.5>E14.5>ISC=AE or *Vil1*, an AE marker to the differentiation cluster E12.5<E14.5=ISC<AE. The clustering allowed identification of potential cell fate markers, which were further used for the single locus approach projects described in section 5.

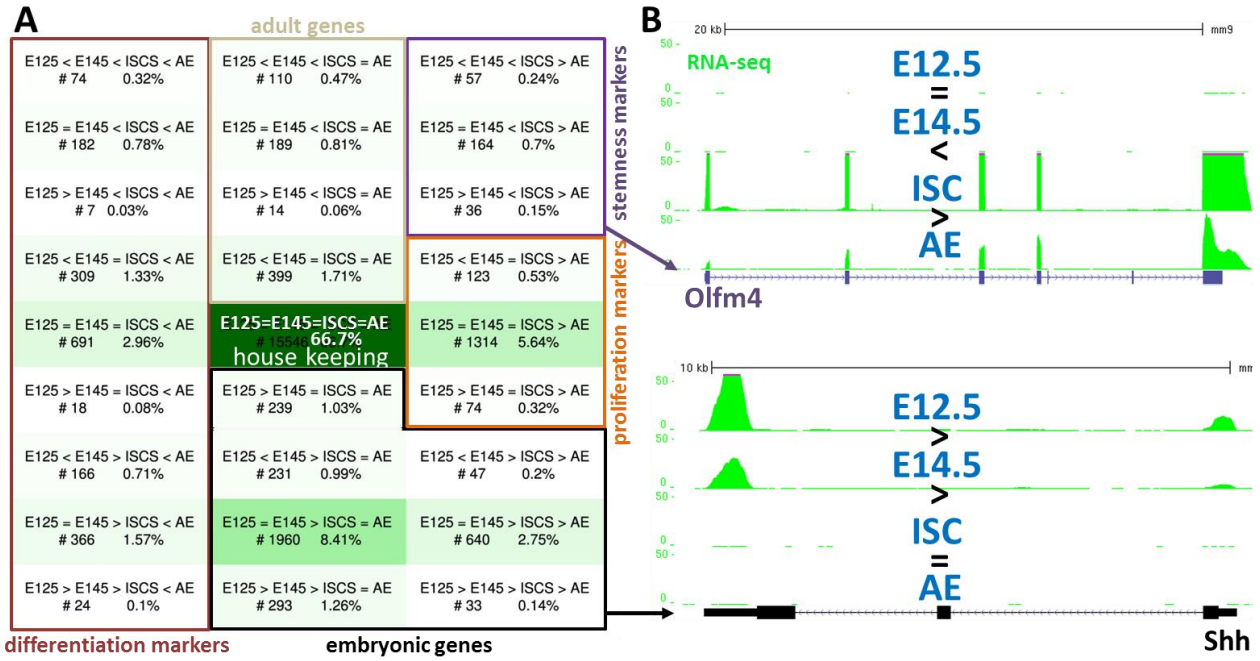
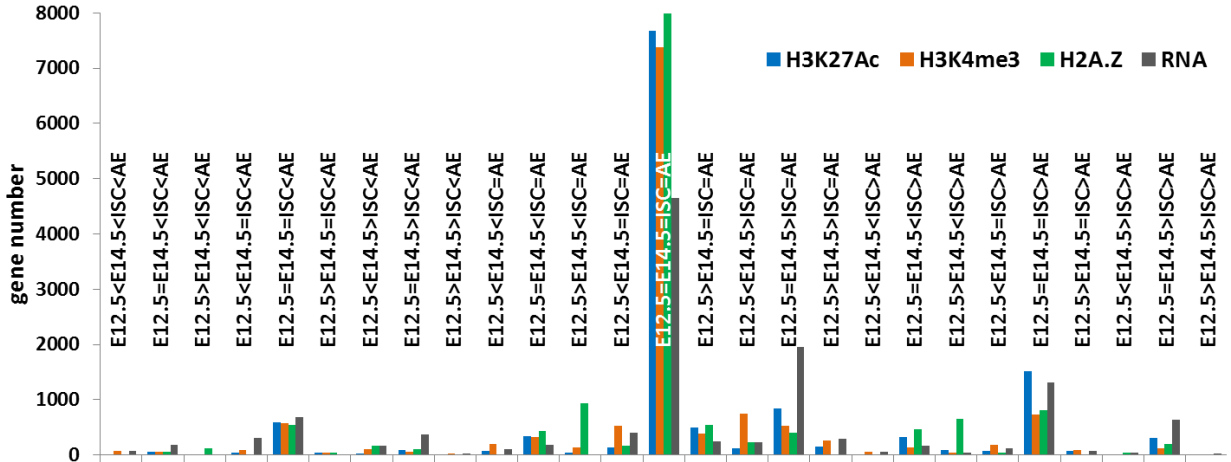


Figure 30: RNA-seq clusters. A: *27 clusters were generated based on at least 2-fold changes of gene expression level between adjacent developmental stages E12.5, E14.5, ISC and AE. Cluster size is shown in total gene number and percentage of all annotated genes, which is additionally indicated by green saturation. Functional cluster groups are indicated by colored frames. The central cluster also contains completely silent genes, later on excluded from analysis.* ***B:*** *Examples of RNA-seq tracks for two cluster annotated genes. The stemness marker Olfm4 is annotated in the cluster E12.5=E14.5<ISC>AE, while Shh, a gene involved in early embryonic gut development, is annotated in cluster E12.5>E14.5>ISC=AE.*

4.3.9.2 Histone ChIP-seq promoter clustering

Clustering of H3K27Ac, H3K4me3 and H2A.Z marked promoters was based on 1.5-fold normalized changes in ChIP-seq signal enrichment against input. As for RNA-seq, the largest cluster contains unchanging promoters, which have a histone mark peak throughout the 4 developmental stages (see ***Fig. 31***). In total, for each ChIP- and RNA-seq clustering, between 12.400 and 13.900 genes were found actively marked (or expressed) at least at one developmental stage. Normalization and clustering were confirmed by ChIP-seq track screening. The threshold of 1.5-fold changes between adjacent stages was chosen for its best specificity and sensitivity. Still, in comparison to gene expression clustering, more genes are annotated in cluster no. 14 (E12.5=E14.5=ISC=AE). This correlates with the higher variation in gene expression in comparsion to histone modifications, but also indicates a lower quantitative sensitivity of ChIP-seq in comparison to RNA-seq.

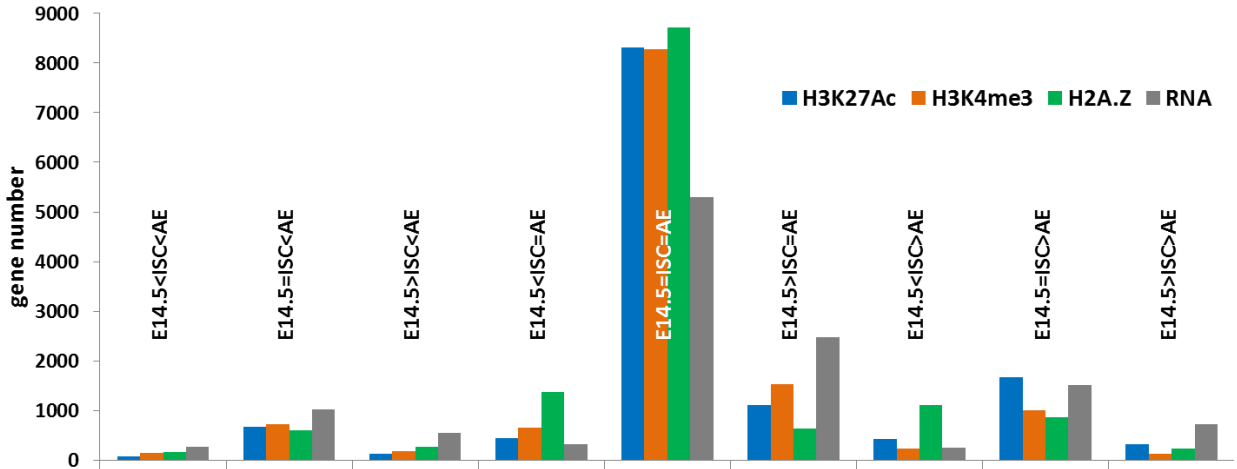


cluster:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
H3K27Ac	6	60	9	35	593	38	24	96	11	66	341	38	140	767	503	116	842	152	17	323	94	75	151	74	6	307	9
H3K4me3	71	62	13	95	577	46	99	55	24	201	327	134	524	737	381	740	530	269	65	132	36	186	731	89	6	120	3
H2A.Z	0	52	113	6	551	47	161	101	0	0	437	928	175	798	551	233	401	0	1	464	647	41	810	6	36	202	0
RNA	74	182	7	309	691	18	166	366	24	110	189	14	399	465	239	231	196	293	57	164	36	123	131	74	47	640	33

Figure 31: Initial 27-fold clustering of epigenetic marks and gene expression. *H3K27Ac, H3K4me3 and H2A.Z: marks are considered increasing/decreasing if the normalized enrichment to input is changing >1.5-fold* between adjacent *developmental stages E12.5, E14.5, ISC and AE. Only genes detected at one or more stages are shown. RNA: gene expression clustering as in* ***Fig. 30*** *except the ~10,000 completely silent genes. Cluster size is shown below as total gene number.*

27-fold clustering yielded for all epigenetic marks as well as for RNA-seq several "empty" clusters, containing no or only very few genes. This occurred for two different reasons: In contrast to the gene expression clusters 1 and 27, histone mark clusters 1 and 27 hardly contain any genes. This effect is due to the comparably low variability of histone mark signal, which is either absent or present, but does not show an as strong intensity range as RNA-seq signal. Hence, only few genes increase/decrease their histone mark signal 3 times in a row. In other words, if the mark is positioned at all histones of a locus, it cannot get any stronger. The second subset of "empty" clusters including numbers 3, 6 and 9 would contain genes, highest expressed or marked at E12.5 and in AE. As generally very few genes are activated or actively marked in early embryos, shut down at later stages but then reactivated on full differentiation, these clusters are not expected to be large. Unlike the first subset, most of the latter clusters are not completely empty due to clustering noise, which is mostly caused by lowly active genes close to the detection level.

Fewer differences in epignetic mark positioning were observed between stage E12.5 and E14.5 in comparison to the other two stage transitions. To reduce clustering noise and to facilitate cluster correlation between different epigenetic marks and gene expression, the 27 clusters were reduced to 9, now only including differential epigenetic state or gene expression between stages E14.5, ISC and AE (see ***Fig. 32***).



cluster:	A	B	C	D	E	F	G	H	I
H3K27Ac	75	666	131	445	8319	1110	434	1664	322
H3K4me3	146	718	178	662	8283	1539	233	1006	129
H2A.Z	165	604	262	1365	8709	634	1112	857	238
RNA	263	1018	556	313	5292	2484	257	1511	720

Figure 32: 9-fold clustering of epigenetic marks and gene expression. *H3K27Ac, H3K4me3 and H2A.Z: marks are considered increasing/decreasing if the normalized enrichment to input is changing >1.5-fold* between adjacent *developmental stages E14.5, ISC and AE. Only genes detected at 1 or more stages are shown. RNA: gene expression clustering based on at least 2-fold changes between adjacent developmental stages E14.5, ISC and AE. Cluster size is shown below as total number of genes.*

4.3.9.3 ChIP and RNA cluster correlations

The correlation analysis of gene expression and histone marks described in section 4.3.5 shows only developmental screenshots with limited information on epigenetic pattern changes during tissue maturation. To see, how far these single time point epigenetic patterns are also reflected in the longterm development, I correlated clusters of different epigenetic marks and gene expression clusters to each other.

Starting from the positive correlation of gene expression with each of the active epigenetic marks H3K27Ac, H3K4me3 and H2A.Z as well as between the histone marks themselves, it was expected that most genes would be annotated to the same clusters. As discussed above, a majority of genes fall into the nonchanging cluster E (E14.5 = ISC = AE) and in this cluster the highest correlation rate was observed (see ***Fig. 33***). At the same time, smaller clusters, as e.g. clusters B and H, showed much fewer parallelly clustered genes, which on the first glance resembled rather negative than positive cluster correlation.

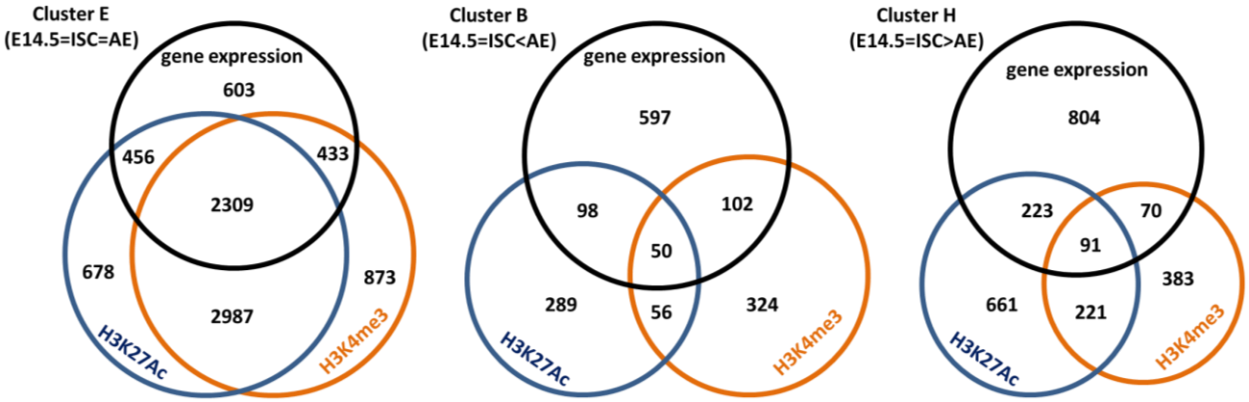
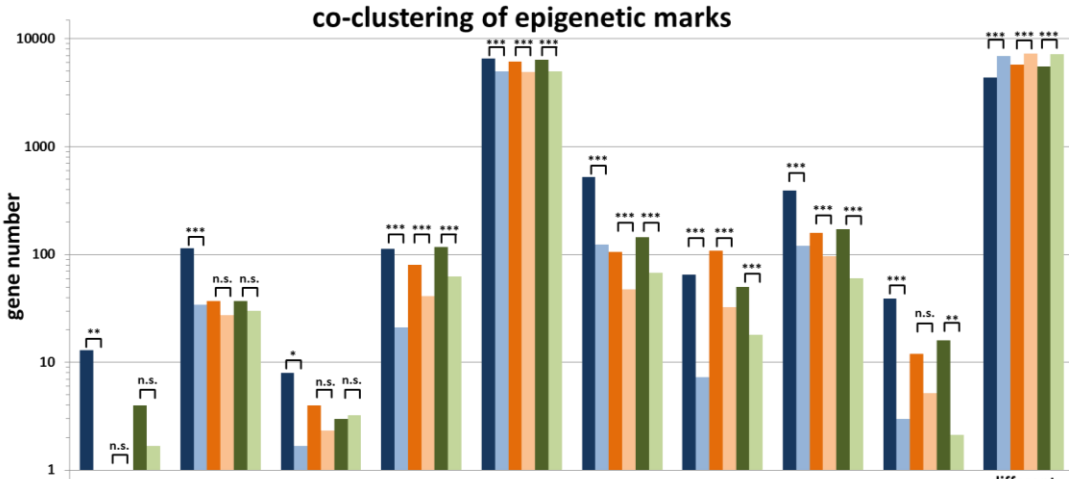


Figure 33: Co-clustering of gene expression and active promoter marks. *Gene expression, H3K27Ac and H3K4me3 clusters and their quantitative overlaps. Only genes were analyzed with an annotated cluster for all 3 traits. 3 out of 9 clusters are shown. Cluster E is the greatest cluster for all analyzed traits, hence most genes in the dataset are annotated to it two or more times. The clusters B and H are lower in single mark annotation frequency, so that double or triple annotation is expected to occur at lower numbers than in cluster E.*

This size bias in co-clustering made the comparison of absolute values or relative fractions in each cluster uninformative. Therefore, a Z-score based approach, comparing pairwise epigenetic mark distribution between the clusters was chosen for further analysis (see method section 7.15.3 for details). Hereby, for each cluster separately, the observed co-clustering gene number of two epigenetic marks was statistically compared to an expected gene number, which was calculated for randomized co-clustering. This theoretical value quantifies the average gene number if co-clustering happened only based on each of the compared marks' cluster frequency. By this method, each combination of active histone marks H3K27Ac, H3K4me3 and H2A.Z was probed for significant parallel clustering, for each cluster separately (see ***Fig. 34***).

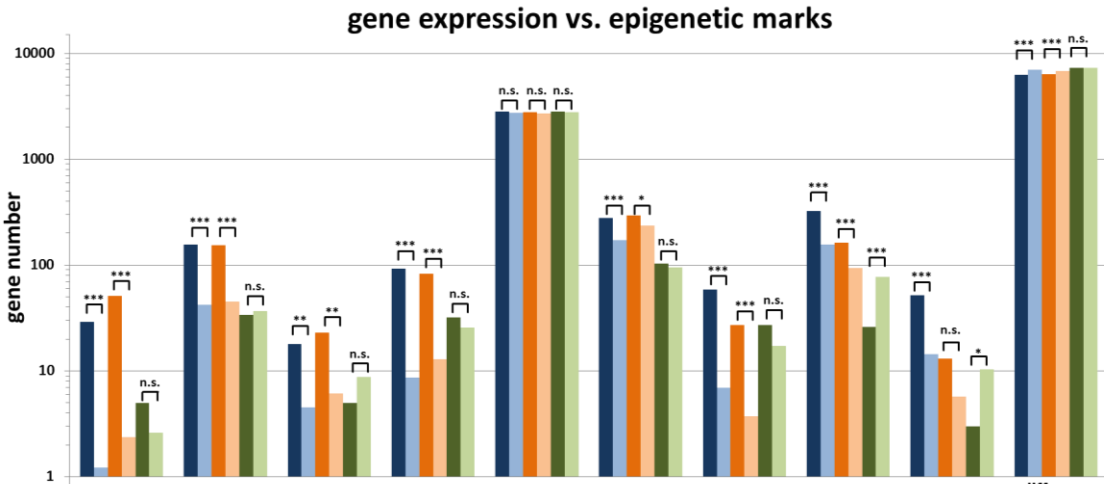


co-cluster: E14.5 – ISC – AE:	A <<	B =<	C ><	D <=	E ==	F >=	G <>	H =>	I >>	different clusters
H3K27Ac = H3K4me3	13	115	8	113	6549	520	65	392	39	4399
H3K27Ac = H3K4me3 random	1	34	2	21	4957	123	7	120	3	6944
H3K27Ac = H2A.Z	0	37	4	80	6183	105	108	159	12	5729
H3K27Ac =H2A.Z random	1	27	2	41	4900	48	33	96	5	7264
H2A.Z = H3K4me3	4	37	3	118	6392	145	50	172	16	5564
H2A.Z = H3K4me3 random	2	30	3	63	5015	68	18	60	2	7240

Figure 34: Pairwise parallel clustering of H3K27Ac, H3K4me3 and H2A.Z. *Clusters A-I: Genes with same cluster annotation for two marks are shown in dark blue, dark orange and dark green. Different clusters: Gene set annotating to two different clusters for the two compared histone marks. Light color bars indicate theoretical co-clustering numbers, calculated from random co-clustering based on single marks' cluster frequencies. Gene numbers are shown in logarithmic scale, absolute values are indicated below. Only genes were analyzed with an annotated cluster for each of the two compared traits. P-values are calculated from pairwise Z-scores between the observed and random co-clustering values. *** P<0.001; ** P<0.01; * P<0.05; n.s. P>0.05. Significant increase in comparison to random co-clustering indicates correlation between two marks positioning and changes during development.*

A significant co-clustering of promoter H3K27Ac and H3K4me3 was observed for all clusters A-I. That means that these two active marks are not only, as shown previously, simultaneously positioning at TSS, but also quantitatively change in parallel during the development from embryo to fully differentiated cells. The same significant co-clustering was observed for each of these two marks paired with H2A.Z in the clusters D-H. The remaining clusters were not significantly elevated in comparison to random clustering. The reason for this might be statistical variation due to the low absolute numbers and single mark cluster frequencies. However, there is also a possible biological explanation: Most of the not significant co-clusters of H2A.Z with any other mark display a ChIP-seq signal increase from ISC to AE. As described in section 4.4, H2A.Z is less expressed and found at lower protein levels in mature AE, indicating a reduced involvement in chromatin regulation which would result in the observed decoupling from other active chromatin marks. This explanation is also supported by the observation, that co-clustering is significant (P<0.0001) in clusters G and H, where H2A.Z decreases from ISC to AE in parallel to compared active marks.

Same as between epigenetic marks, the co-clustering analysis was performed between each single promoter mark and gene expression (see ***Fig. 35***). During development the changes of RNA-levels match each of the active marks H3K27Ac and H3K4me3, mostly with high statistic confidence levels. An exception is cluster E, where the observed co-clustering is only minimally over the random annotation values. In this cluster all the unchanging genes/TSS are annotated. As such genes/TSS are anyway most numerous, the random co-clustering values are already high, while clustering noise by just below/above threshold annotation prevents a substantial co-clustering enrichment. Furthermore, the unchanging histone marks might be uncoupled from gene expression in cluster E, as other, more changeable mechanisms, such as e.g. transcription factors, might play a stronger regulatory role for these genes. Overall, the RNA to H3K4me3/K27Ac correlation during the developmental timecourse is strong, as shown by the significant decrease in co-clustering of non-equally annotated genes in comparison to random association ($P=5*10^{-25}$ for RNA vs. H3K27Ac and $P=8*10^{-13}$ for RNA vs. H3K4me3).



co-cluster: E14.5 – ISC – AE:	A <<	B =<	C ><	D <=	E ==	F >=	G <>	H =>	I >>	different clusters
RNA = H3K27Ac	29	157	18	93	2822	277	59	325	52	6301
RNA = H3K27Ac random	1	42	5	9	2729	171	7	156	14	6999
RNA = H3K4me3	51	153	23	83	2793	292	27	163	13	6312
RNA = H3K4me3 random	2	45	6	13	2714	237	4	94	6	6789
RNA = H2A.Z	5	34	5	32	2800	103	27	26	3	7334
RNA = H2AZ random	3	37	9	26	2760	94	17	78	10	7336

Figure 35: Pairwise parallel clustering of gene expression to H3K27Ac, H3K4me3 and H2A.Z. *Clusters A-I: Genes which same cluster annotation for gene expression and one active marks are shown in dark blue, dark orange and dark green. Different clusters: Gene set annotating to two different clusters for the two compared histone marks. Light color bars indicate theoretical co-clustering numbers, calculated from random co-clustering based on single traits' cluster frequencies. Only genes were analyzed with an annotated cluster for each of the two compared traits. Gene numbers are shown in logarithmic scale, absolute values are indicated below. P-values are calculated from pairwise Z-scores between the observed and random co-clustering values. *** P<0.001; ** P<0.01; * P<0.05; n.s. P>0.05. Significant increase in comparison to random co-clustering indicates correlation between the histone mark and gene expression during development.*

Surprisingly, H2A.Z showed no significant co-clustering with RNA-levels, mostly matching the calculated random association values in each cluster as well as for all clusters taken together. This was especially unexpected, as H2A.Z clusters matched the H3K27Ac and H3K4me3 clusters, which in turn fit to the changes in RNA-levels. These results demonstrate the necessity to review the H2A.Z annotation as an active promoter mark (see section 6.1.4).

4.4 Heterogeneity of tissues

During the analysis of genome-wide data it is often assumed that FACS-purified populations are homogenous. This is generally true regarding the cellular properties and marker expression specifically selected for during cell harvesting and FACS. The dissection procedures and FACS-markers described in section 7.2 are sufficient to purify well specified cell populations which are necessary for developmental comparison of epigenetic states. However, it should be kept in mind that embryonic development and adult homeostasis are continuous processes consisting not only of stable but also transient cell types, which can additionally vary in gene expression and epigenetic states, depending on cell cycle progression, intracellular signaling and stress.

While working with 5 purified cell populations, namely E12.5, E14.5, ISC, AE and Paneth cells, I found numerous evidence for substantial heterogeneity in all of them.

4.4.1 Polar heterogeneity in embryonic intestinal epithelium

The embryonic gut epithelium looks at the first glance as a monolayered tube, consisting of uniform cells which can be FACS purified from surrounding mesenchyme by their even *Epcam* expression. However, already from stage E12.5 some genes are expressed at varying levels only in a subset of cells (see ***Fig. 36***). This effect is especially visible at later embryonic stage E14.5, where e.g. *Lgr5* and *Axin2* are expressed only at polar regions of the gut tube, while being virtually absent from adjacent regions. These polar indentations of the embryonic gut are the nascent intervilli regions, which are strongly expressing the ISC marker *Lgr5* at later stages. Hence the complex architecture of the adult gut epithelium with an array of specified cell types seems to root from earlier, embryonic stages.

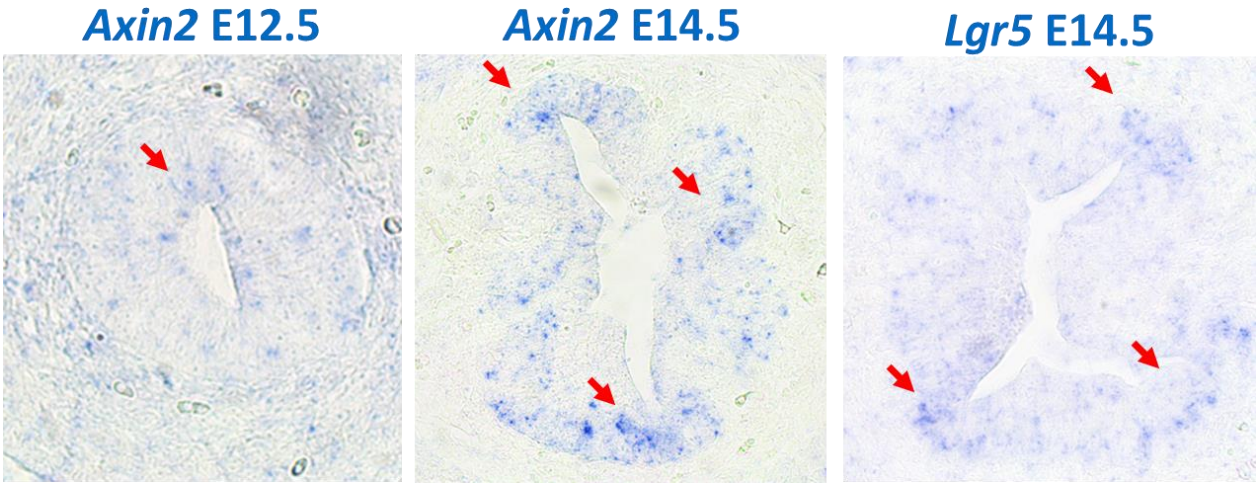


Figure 36: Polar gene expression in embryonic epithelium. *RNA-ISH on E12.5 and E14.5 wt sections. Lgr5 and Axin2 are ISC markers and contribute to the Wnt-signaling pathway. Red arrows indicate heterogeneous signal at polar regions (epithelial indentations which form intervilli structures at later stages).*

4.4.2 *Lgr5*-positive ISC subpopulations

As known from previous studies[126], also adult ISC separate in several subpopulations, including proliferative and senescent ones. RNA-ISH on adult tissue sections confirmed this, as shown in ***Fig. 37***. While *Lgr5* is expressed in all ISC, including both, the proliferative crypt base cells and the quiescent +4 cells, *Axin2* is a specific marker for the +4 to +7 region.

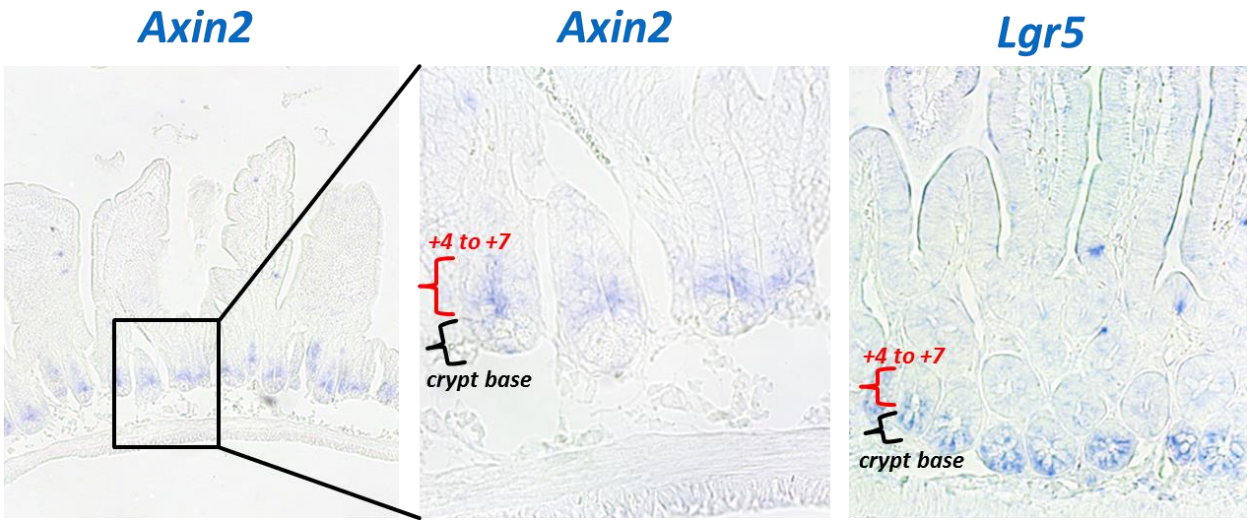


Figure 37: Lgr5-positive ISC subpopulations in adult crypts. *RNA-ISH on wt adult sections, small intestine, central fraction. Lgr5 and Axin2 are both ISC markers, overlapping at the +4 position (quiescent ISC). While Lgr5 is expressed strongest at the crypt bottom, Axin2 shows strongest ISH signal at the +4 to +7 position, corresponding to quiescent ISC and ISC differentiating to TA cells.*

Additionally, I observed varying *Lgr5-GFP* expression levels by FACS as well as clonal differences in GFP signal intensity by fluorescent microscopy (see section 4.3.1), which indicates further heterogeneity in ISC gene regulation between crypts.

4.4.3 AE maturation and anterior/posterior variation

Absorptive cells, leaving the crypt mound, are considered fully differentiated, however they still contain a part of the highly replicative TA cells' mRNA and protein load, as well as epigenetic marks. Those "remains" can only partially be erased during the short lifespan of an adult enterocyte. Therefore, a villus base/tip gradient can occur, which I confirmed on protein and RNA level.

While e.g. H3K27Ac ChIP-seq showed even distribution at all developmental stages, H2A.Z ChIP yielded less AE material. To confirm that this is not a ChIP bias and H2A.Z is less abundant in AE, an IF staining against H2A.Z was performed (see ***Fig. 38***). This staining clearly demonstrates that high H2A.Z levels are only found in the nuclei of the lower villus, while the H2A.Z signal gradually decreases toward the tip, until it is hardly distinguishable from background. The reduced H2A.Z levels are also matching the approx. 2-fold decreased AE expression of *H2afz*, the main H2A.Z coding gene. The mixture of the H2A.Z high, basal villus cells and the H2A.Z low, maturated cells at the villus tip explains, why less material was precipitated in ChIP. It therefore should be kept in mind that the acquired H2A.Z ChIP-seq data mostly represents the basal villus AE population.

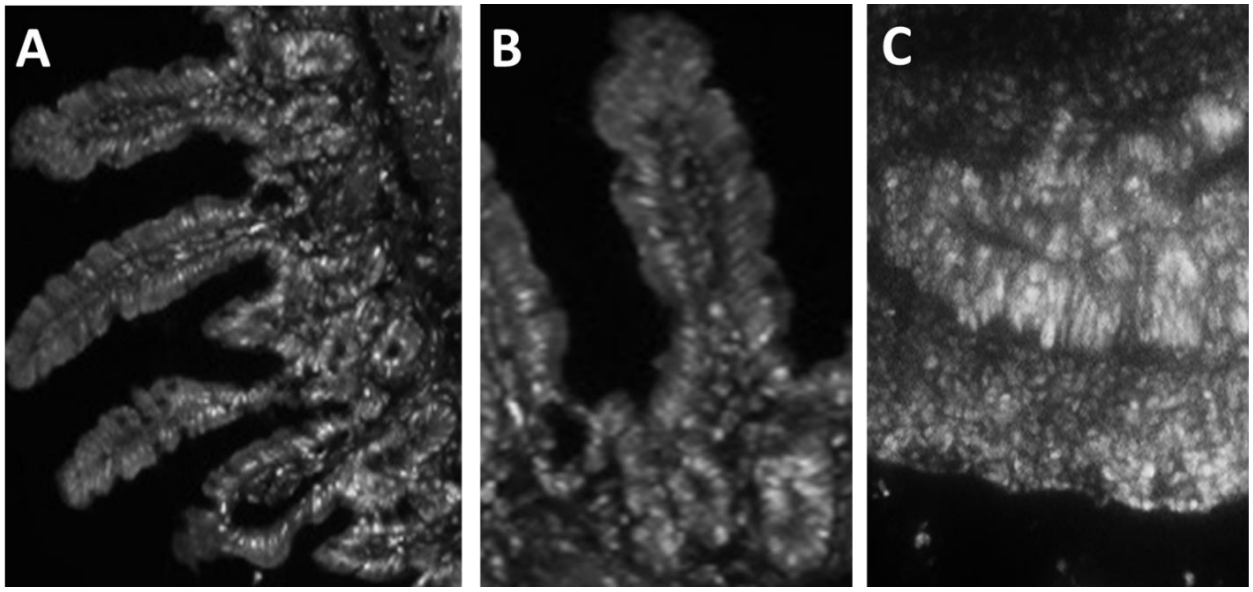


Figure 38: H2A.Z distribution in adult and embryonic gut. *Secondary IF staining on wt cryosections, enhanced contrast.* ***A:*** *Adult epithelium shows strongest signal in crypt nuclei, decreasing towards the villus tip.* ***B:*** *Enlarged upper crypt + villus from panel A.* ***C:*** *E14.5 embryos show even nuclear H2A.Z signal. Negative controls without the primary H2A.Z Ab yielded very low background signal (not shown).*

Gene expression changes were also observed along the maturation axis. Some genes, such as Picalm, are expressed in a gradient, strongest at the villus base, while absent at the tip (see ***Fig. 39***). Other genes show heterogeneous expression in the adult intestinal epithelium, as e.g. the Sfrp5 gene, which is heterogeneously expressed throughout the villus and the crypt base, while depleted in TA cells.

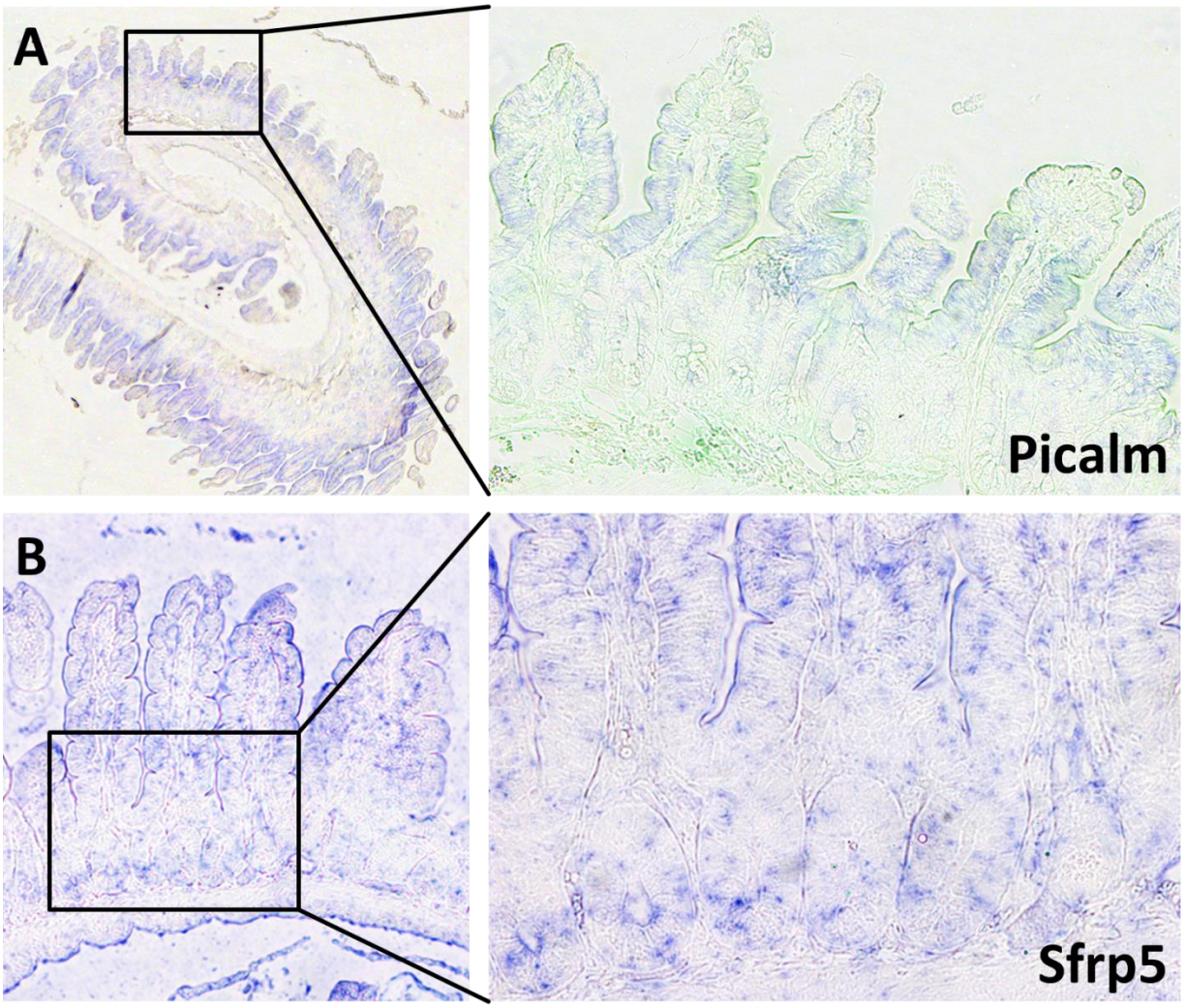


Figure 39: Differential gene expression along the adult villus. *RNA-ISH on wt adult small intestinal sections.* ***A:*** *Picalm is expressed at the villus base and depleted in crypts and villus tips.* ***B:*** *Sfrp5 yields salt and pepper like signal across the whole villus compartment and the crypt base.*

Both observations, the depletion of H2A.Z as well as differential gene expression, indicate transient AE cell populations, determined by their maturation stage.

Differential gene expression was also repeatedly observed along the anterior-posterior axis of the adult gut. The intestinal areas, closest to the stomach or the caecum, to some extent resemble these organs by morphology and gene expression. Although, the approx. 10 % of most anterior and posterior intestinal fractions were not harvested for ChIP, also the central 80 % show a gradient of gene expression (see ***Fig. 40***). In the shown example *Axin2* is only expressed in the crypt compartments of the anterior and central fractions (see also ***Fig. 37***), however in the posterior small intestine additional heterogenic lower villus expression was observed. Similar to the adult intestine, I also observed changes in gene expression along the anterior/posterior axis of the embryonic gut, which is not surprising, as the development of the primitive gut to colon, caecum, small intestine and stomach is based on regulatory gradients[127,128].

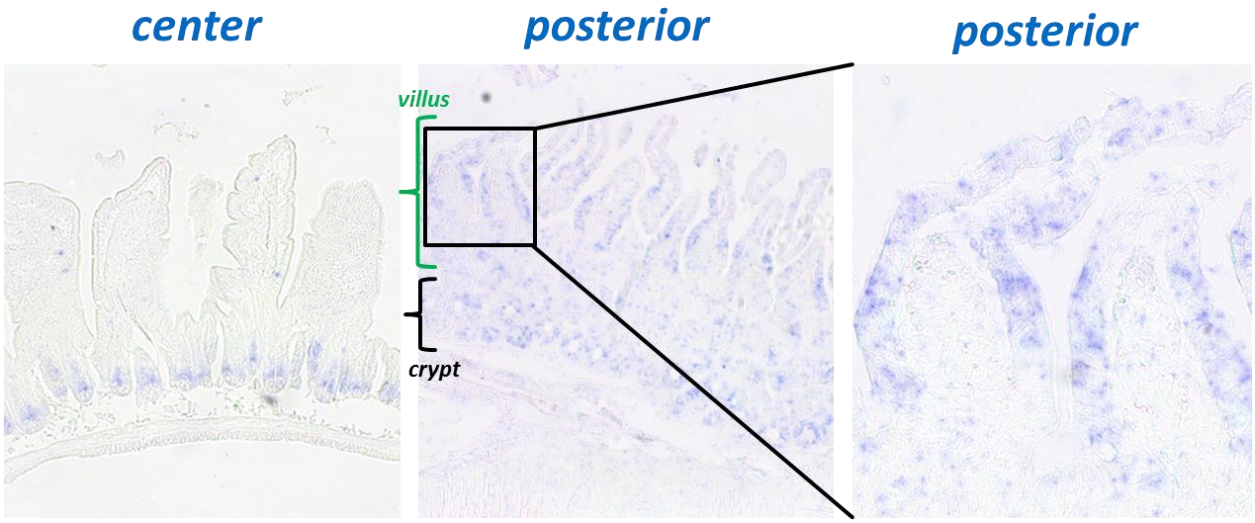


Figure 40: Anterior-posterior variation of gene expression in the adult gut epithelium. *RNA-ISH against Axin2 on wt adult sections.* ***Left:*** *Central fraction of the small intestine, as shown in* ***Fig. 37****. Axin2 expression is limited to the crypts in the anterior and central gut.* ***Center/right:*** *Posterior small intestine. In addition to crypts, Axin2 is also heterogeneously expressed in posterior villi.*

4.4.4 Paneth cell subpopulations

During the optimization of the Paneth cell isolation protocols I observed an additional UEA1 and CD24 positive population on FACS plots. On a closer look, this additional population, which showed a slightly increased CD24 signal, was comprised of PAS-positive Paneth cells mixed with a PAS negative cell type. This population was possibly comprised of freshly differentiating TA/Paneth cells at different maturation stages. Although the final optimized Paneth cell isolation and FACS protocol yields only PAS-positive Paneth cells at a high purity, it should be kept in mind that the comparably long-lived Paneth cells might still undergo further maturation or even show variations in gene expression depending on their exact crypt position, as ISC do.

5. Single locus approach

To complement genome-wide studies, I was investigating epigenetic marks at single genetic loci. While the whole-genome approach provided a good overview of epigenetic patterns, a closer look at active genes and enhancers has the potential of unraveling epigenetic mechanisms and functions.

Therefore, I conducted three projects with a diverse set of methods and biological materials. The first project was a functional *in vivo* study of histone-methylation at the limb-specific enhancer of the Sonic Hedgehog gene (*Shh*). The second project targeted the composition of the RNA-Polymerase II (PolII) complex at different positions of single genes, to determine its components and interactors depending on the exact localization. In the last single locus project, Proximity Ligation Assay (PLA) was applied to shed light on epigenetic heterogeneity of specific genetic loci *in situ*. Hereby the positioning of an epigenetic mark was assessed at alternative TSS throughout the intestinal tissue at single cell resolution.

An additional project was conducted by Berith Messner during her Master studies under my supervision. Here, enhancer-promoter interactions were assessed by PLA on the well-studied example of *Shh* and its tissue specific enhancers.

5.1 TALE-targeted *Shh*-enhancer demethylation

Regulatory elements such as enhancers can be identified and predicted by a variety of methods. So tens of thousands enhancer candidates were predicted by evolutionary conservation and chromatin mark positioning[129]. However, only a small fraction of these was tested functionally by a reporter gene fusion outside of their native locus and an even much smaller number was investigated for enhancer activity in their endogenous genomic context[130,131].

A gene with several functionally well studied enhancers is Sonic Hedgehog (*Shh*). It is encoding a signaling protein involved in a number of embryonic and adult developmental processes including the organogenesis of brain, limb and gut[132-134]. Tissue-specific activity of *Shh* and its enhancers made them to highly interesting loci for the developmental and epigenetic research fields. So the H3K4me1 marked MFCS1 (male fish conserved sequence 1) enhancer is specific for limb development, ensuring correct expression of *Shh* in the embryonic limb buds[135]. Hereby, *Shh* is actively expressed only at one side of the emerging limb bud, the so called zone of polarizing activity, which later correlates to the position of digits no. IV and V[132,136] (see ***Fig. 41***). By the one-sided expression of the *Shh* gene and the diffusion of the secreted protein, a Shh gradient emerges which directs digit development. Highest Shh levels are hereby found at the position of the developing digit V and virtually no Shh at the future digit I[137]. The MFCS1 enhancer, found in the intron of the unrelated *Lmbr1* gene, colocalizes with the *Shh* promoter by looping the 1 Mb distance between them and activates *Shh* expression in the limb bud[138,139]. While not affecting any other Shh-dependent processes,

MFCS1 was shown to be essential for *Shh* expression in the murine limb bud, as the deletion of this enhancer leads to complete loss of *Shh* expression and subsequent limb truncation, including the complete deletion of digits 2-5 as well as overall shorter limbs[135]. On the other hand, mutations in the MFCS1 sequence can lead to ectopic activity of *Shh* resulting in polydactyly phenotypes observed in many vertebrates including humans[138,140].

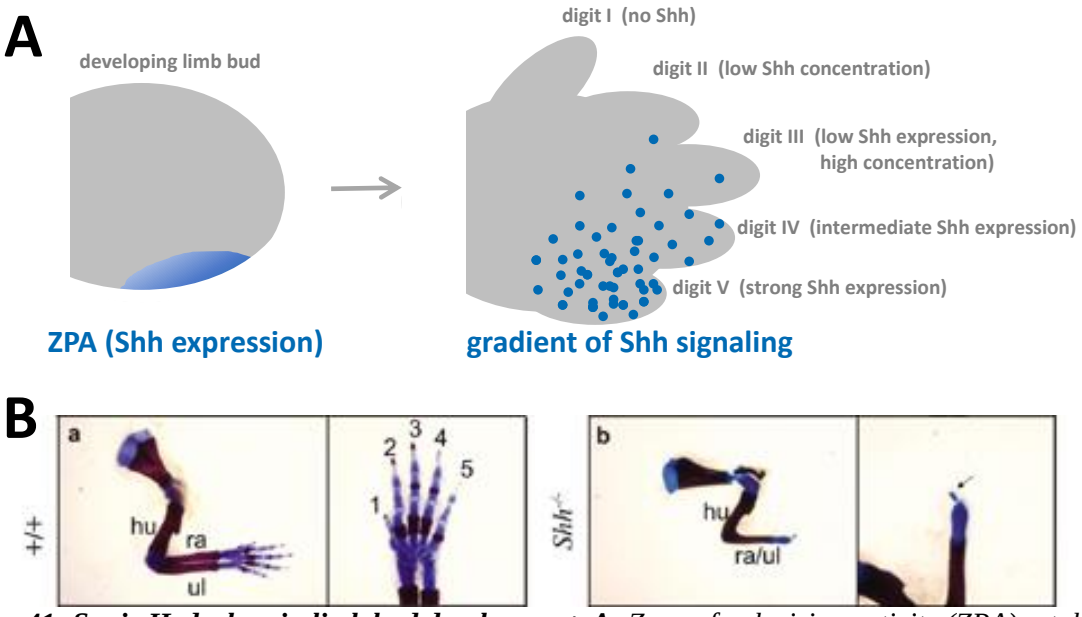


Figure 41: Sonic Hedgehog in limb bud development. A: *Zone of polarizing activity (ZPA) establishes a Shh gradient. In the early developing limb bud, Shh is only expressed and secreted in the ZPA, which correlates to the position of developing digits IV and V. By diffusion through the tissue, decreasing levels of the Shh protein reach also the area of future digits II and III.* ***B:*** *adapted from Litingtung et al.[141] Shh-knockout mice show heavily impaired limb development. The here shown forelimb is shortened and only the Shh-independent digit I developed.*

H3K4me1 is a known marker of poised and active enhancers such as MFCS1 (see section 2.3.4 for details). However, little was known about the function of this histone modification. Two opposing hypothesis were considered: Firstly, H3K4me1 might be an important modification to recruit nuclear factors to the enhancer and hence lead to elevated PolII activity on looping and interaction with the promoter. With regard to the positioning of H3K4me1 at active and poised enhancers, the recruited nuclear factors could be transcription factors, chromatin remodelers or proteins that mediate enhancer-promoter interactions (see ***Fig. 42A***). The opposing hypothesis considers that enhancers can also interact with their target promoters without elevated transcription levels (poised state). Hereby as well as during activating enhancer-promoter interactions, the H3K4me3 mark might be positioned on the enhancer by histone-modifying enzymes bound at the promoter as part of the transcription initiation complex (see ***Fig 42B***), especially as the localization of histone methyl-transferases is very common at promoters which are associated with high levels of H3K4me1/2/3 (see section 2.3.4 for details). Hence, the H3K4me1 mark would be not the cause but a mere consequence of enhancer-promoter interactions.

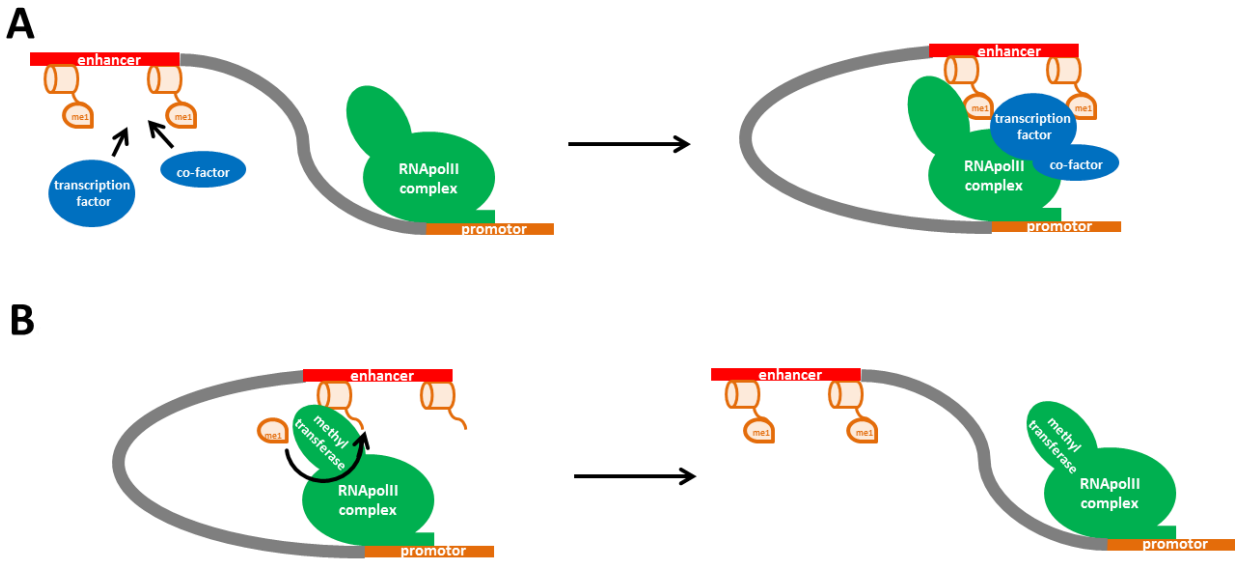


Figure 42: Two models for H3K4me1 action at enhancers. A: *The H3K4me1 histone mark (me1) positioned at the enhancer recruits the activation complex, containing transcription factors, co-factors or other nuclear proteins which are essential to activate target gene expression.* ***B:*** *During interaction between the enhancer and the poised or active target promoter, the nucleosomes of the enhancer are methylated by specific histone methyltransferases, without changing the transcriptional status of the gene.*

A possibility to test these hypotheses is to remove histone methylation specifically at the enhancer locus by the lysine specific histone demethylase Lsd1 (also called Kdm1A) *in vivo*. If the second hypothesis is true, no effect should occur, while if the H3K4me1 mark is responsible for the recruitment of the transcription machinery, *Shh* expression in the limb bud should be impaired.

To pinpoint the impact of H3K4me1 at the *Shh* enhancer locus by demethylation *in vivo*, it was very important to avoid histone methylation changes elsewhere. Genome-wide misregulated levels or positioning of histone modifications usually strongly affect a variety of basic nuclear functions, such as chromatin architecture, TF-binding and ultimately gene expression, which in some cases leads to cell death and for sure makes the assessment of a single locus effect impossible. As the H3K4me1 mark is found at many loci, including enhancers and promoters, a very precise method was needed to remove histone methylation without any off-target effects.

To guide Lsd1 specifically to a single genomic locus, three mechanisms are thinkable: the fusions to Zink fingers, to a transcription activator-like effector (TALE) or an RNA-based guidance based CRISPR/Cas system[142]. At the beginning of this project in 2012, the CRISPR/Cas method was just emerging with still little data on the incidence of off-target effects. In addition, this system would be the most complicated of all, as it would require to introduce permanently two crucial components into the developing embryo: not only the an Lsd1-fusion protein, as for the other two targeting systems but also a guiding RNA. Hence the TALE approach was selected as it is also superior to zinc fingers by having a straightforward correspondence of protein sequence to target sequence binding and only very few restrictions in possible

target sequences[142] (see ***Fig. 43***). Each zinc finger domain binds three nucleotides of the target sequence and the affinity of a zinc finger array to the target sequence might deviate from the theoretical sum of single zinc finger domains. This makes the design of a working single locus targeting system with zinc fingers very laborious and requires a trial and error approach[142]. TALE on the other hand is built of monomers that bind one nucleotide each and are generally unaffected in their binding preference by neighboring monomers. Each monomer consists of 33-35 amino acids with only two variable positions which bind to one nucleotide and determine the binding preference[143]. As the monomers can be arranged in any order, a straightforward design of the TALE amino acid sequence only based on the desired genomic target sequence is possible[144].

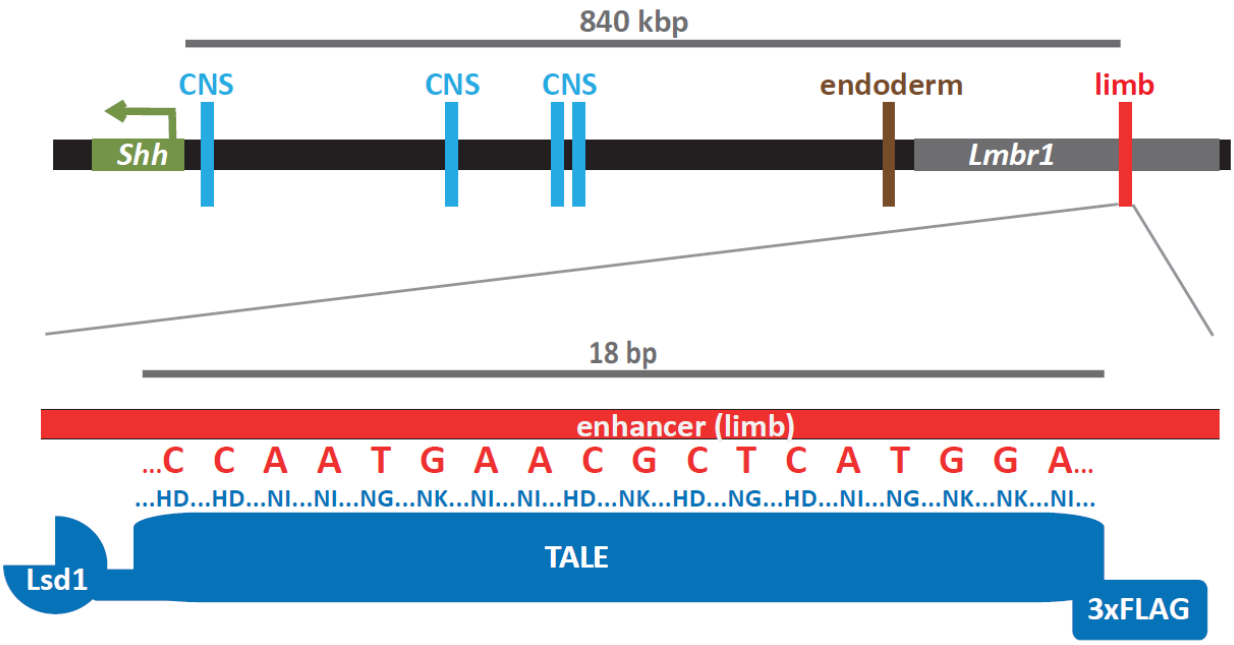


Figure 43: Shh limb specific enhancer targeting by TALE. Top: *Shh has a set of tissue specific enhancers (a subset is indicated by vertical lines), including the MFCS1 limb-specific enhancer, which is located inside the Lmbr1 gene.* ***Bottom:*** *An 18 bp long, genome-widely unique sequence inside the MFCS1 locus was selected for FLAG-TALE-Lsd1 targeting. 18 TALE monomers were designed and cloned to exactly match the target sequence, hereby two variable amino acids (blue letters) of each monomer bind one nucleotide(red letters).*

In addition to the C-terminal TALE-fusion to Lsd1 under ubiquitous CMV promoter and enhancer control, three more features were introduced (see ***Fig. 44***). To monitor expression of the whole construct *in vivo*, an N-terminal eGFP with a self-cleaving P2A domain was added. To allow immunoprecipitations (IP) of Lsd1-TALE for subsequent protein or bound chromatin analysis, a FLAG-tag was fused. As a depletion of *Shh* limb bud expression leads to severe phenotypes in transgenic mice[135], a *loxP*-flanked stop-cassette was introduced after the CMV-promoter. This would ensure no deleterious effects in the transgenic line, while full activation of the *TALE-Lsd1* expression after crossing with *Hprt-Cre* mice (see ***Fig. 44A***) would allow subsequent limb development analysis at embryonic stages. Hereby the Cre-recombinase under the control of the ubiquitous *Hprt*-promoter would excise the stop cassette and lead to the expression of the *eGFP-P2A-*

TALE-Lsd1 construct throughout the embryonic and adult stages of the F1 generation. The ubiquitous expression of *Hprt-Cre* and hence the whole *TALE-Lsd1* construct allowed to test for specific *MFCS1*-histone demethylation by TALE-Lsd1, as it should not affect anything but the developing limb bud.

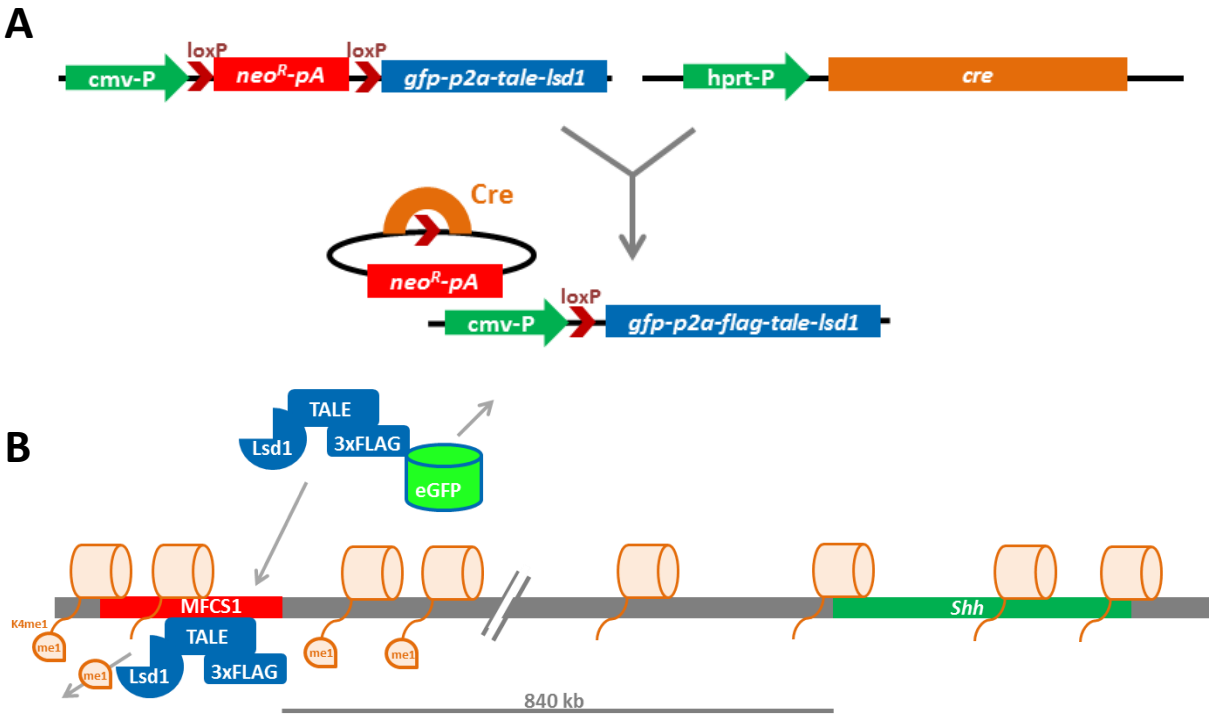


Figure 44: Cre-activation of the eGFP-P2A-TALE-Lsd1 transgene on breeding. ***A: Left:*** *The floxed stop cassette (neo^R-pA) terminates construct (blue box) expression from the CMV-promoter in transgene founders.* ***Right:*** *Hprt-Cre mice ubiquitously express Cre-recombinase.* ***Center:*** *in offspring, positive for both, the transgene and the Hprt-Cre alleles, loxP sites are recombined by Cre-recombinase and the stop cassette removed resulting in a ubiquitously expressed construct.* ***B:*** *Expression can be monitored by eGFP, which is cleaved off by the P2A, a 14 AA long self-hydrolyzing peptide, between 3FLAG and eGFP on translation. The residual protein consists of the FLAG that allows antibody targeting, the TALE, which binds selectively to the Shh long-range enhancer MFCS1, and the Lsd1, a lysine specific histone demethylase removing the H3K4me1 mark.*

To avoid a phenotypic burden of *Shh* depletion on the mice, all analysis was performed at embryonic stage e15.5 where skeletal development can already be fully assessed. Because of the heterozygosity of both, the *TALE-Lsd1* and *Hprt-Cre* mice, ¼ of all embryos was expected to be *TALE-Lsd1*- and *eGFP*-positive. Because of very low eGFP-signal due to low expression, degradation or high autofluorescence, the construct's expression was instead assessed by IP followed by Western blot (WB) against the FLAG-tag. Skeletal phenotypes were analyzed by bone and cartilage stainings, while changes in *Shh*-expression could be determined by qPCR, ISH or WB and changes in histone methylation by ChIP-qPCR.

5.1.1 Results: TALE-targeted *Shh*-enhancer demethylation

5.1.1.1 Cloning of a MFCS1-specific TALE

TALE monomers were assembled with the sequence specificity shown in ***Fig. 43*** and recloned to the eukaryotic vector pcDNA-3.1 (see method section 7.5 for details). Correct assembly was confirmed by restriction analysis and Sanger sequencing.

Prior to the generation of transgenic animals, the construct was checked for correct expression and protein product stability in transiently transfected HEK293T cells (see ***Fig. 45***). Therefore, WB against the flag-tag was performed, confirming the correct size (220 kDa) and showing no degradation products. GFP-levels were checked by FACS and fluorescent microscopy. This, initial part was performed by my colleagues before I joined the research group and took over this project.

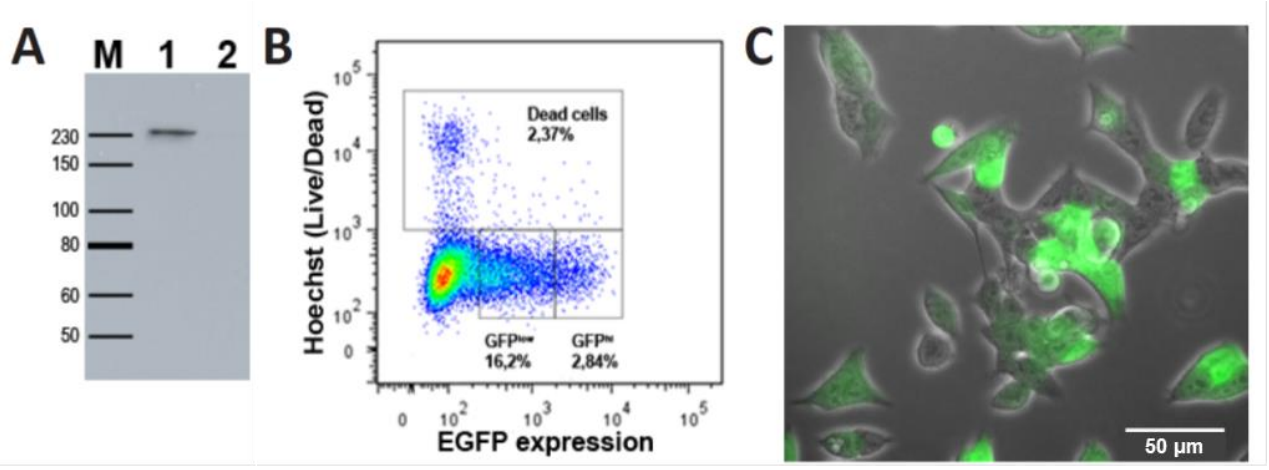


Figure 45: Confirmation of the pcDNA3.1 eGFP-P2A-TALE-Lsd1 construct in HEK-cells. A: *Stability of the 3xFlag-TALE-Lsd1 fusion protein containing 18 binding units (1) was tested on lysate from transiently transfected HEK293T cells by WB using anti-flag Ab. Protein marker (M), non-transfected HEK293T control (2).* ***B:*** *FACS profile of HEK293T cells transiently transfected with pcDNA3.1 eGFP-P2A-TALE-Lsd1.* ***C:*** *Fluorescent and bright-field microscopy overlay of transiently transfected HEK293T cells.*

5.1.1.2 Generation of transgenic mice

The final TALE-Lsd1 construct was sent for the generation of transgenic mice to the Institute Clinique de la Souris in Illkirch, France. 53 biopsies were obtained for genotyping by qPCR, of which five males and four females showed genomic insertions of the construct. However, two males died before the animals were transported to Mainz and one female shortly after, indicating deleterious effect of transgene insertion loci in these animals. As the insertion loci were random, each of the remaining 6 animals was considered a single strain with potentially different expression levels of the construct and hence analyzed separately.

5.1.1.3 Breeding and embryonic analysis of TALE-Lsd1 strains

The two male *TALE-Lsd1+* mice were directly crossed to *Hprt-Cre+* females and the F1 generation was analyzed at embryonic stage E15.5. The 4 *TALE-Lsd1+* females had to be first crossed to CD1 males, to maintain the strain, and only their offspring were crossed to *Hprt-Cre+*, followed by embryonic analysis of the F2 generation. All embryos from the *TALE-Lsd1* x *Hprt-Cre* cross were screened for double positives by PCR (see ***Fig. 46***) and collected for skeletal stainings, WB, flag-ChIP and RT-qPCR, which were performed depending on genotyping results. Fluorescence based screening for GFP-positive embryos was abandoned due to a very low GFP signal.

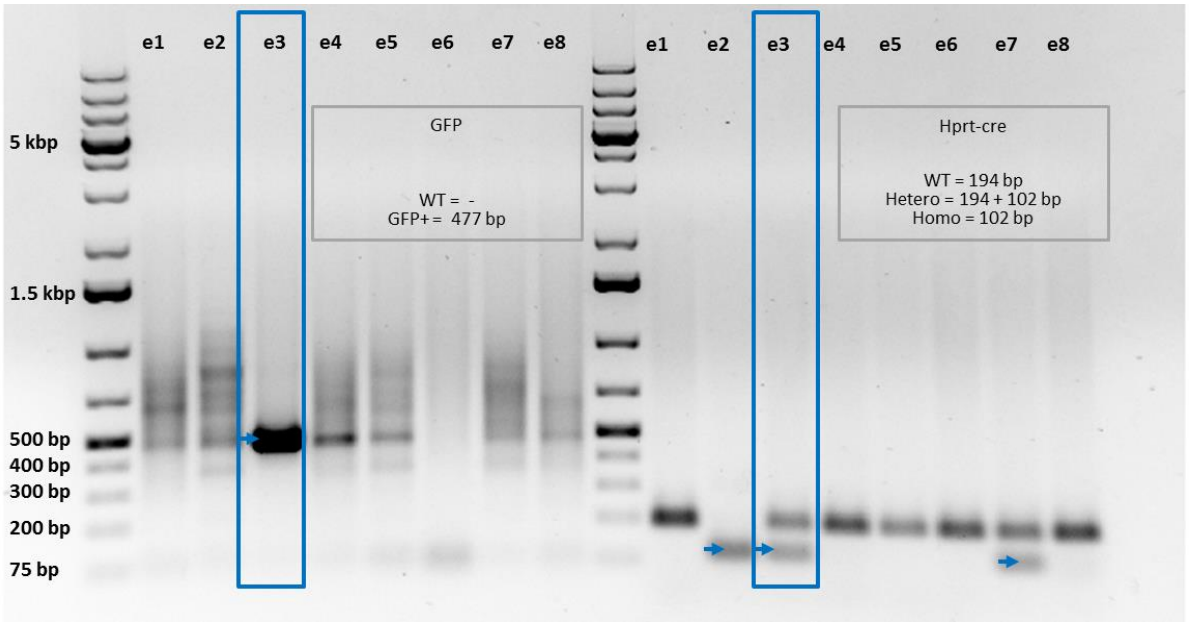


Figure 46: Genotyping example of E15.5 embryos from TALE-Lsd1 x Hprt-Cre cross*. Lanes 1 and 10: Generuler 1 kb plus DNA ladder. Lanes 2-9, 11-18: embryos 1-8 from the cross of a TALE-Lsd1 male with an Hprt-Cre female. Blue arrows indicate mutant alleles, 477 bp for GFP on the left and 102 bp for Hprt-Cre for Hprt-Cre on the right. The only double positive embryo from this litter was no.3.*

In total, 184 embryos were genotyped, with 25 double positive genotypes and 5 inconclusively typed embryos. Of all embryos, 43 % were *Hprt-Cre* positive and 33 % *GFP* positive, which indicates a deleterious effect of the *TALE-Lsd1* construct in at least some of the strains ($P < 0.05$, Chi-square test). However, no further deleterious effect was observed after introduction of Cre-recombinase:

43% * 33% * 184 embryos = 26 double positives expected vs. 25 observed

Out of six founder mice, only two males (F0m1 and F0m2) and two females (F0f3 and F0f5) contributed to the embryonic analysis. Of the remaining two founder females, one was sterile, while the second, although positive by previous genotyping, turned out not to have the construct in its germline and hence only produced wild type offspring after crosses with CD1. The remaining 4 strains contributed each 30-72 embryos from 4-9 litters to the analysis.

5.1.1.4 Skeletal analysis of TALE-Lsd1+ embryos

All dissected e15.5 embryos showed normally developing full-length fore- and hindlimbs with correct digit number. A subset of the embryos, including *TALE-Lsd1* and *Hprt-Cre* double positives, was selected for detailed bone cartilage analysis (see ***Fig. 47***). Overall, no skeletal abnormalities were observed, suggesting either nonfunctionality of the H3K4me1 mark on the MFCS1 enhancer, failed TALE-targeting, or a silenced *TALE-Lsd1* locus.

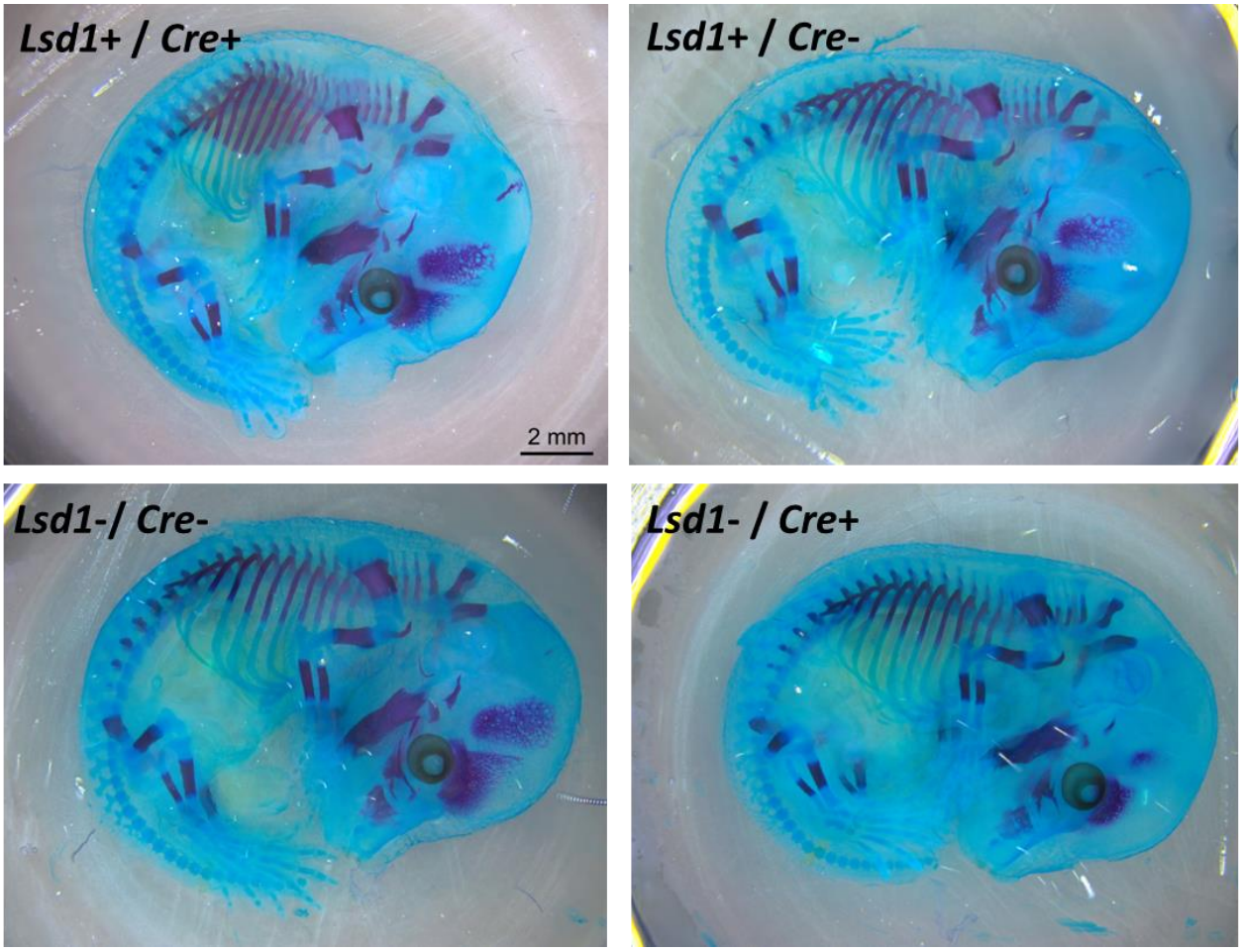


Figure 47: Bone and cartilage staining examples of e15.5 embryos. *All embryos shown are from the heterozygous cross F0m2 TALE-Lsd1+/- x Hprt-Cre +/- and represent the four possible genotypes. Mineralized bones are stained by alizarin red and cartilage tissue is stained by alcian blue. All limbs are equally developed and have 5 digits, independent of genotype. The shown embryos are representative for all analyzed e15.5 embryos. The hunchback of the upper left embryo might be an artefact of sample preparation, as it was not reproduced in any other double positive embryo.*

5.1.1.5 Molecular embryonic analysis

Integrity of the construct in the four remaining strains was checked by PCR, finding that the F0f3 stop cassette is either deleted or at least truncated (see ***Fig. 48***). The same analysis and Sanger sequencing on F1 and F2 embryos from *TALE-Lsd1* x *Hprt-Cre* crosses confirmed Cre activity, which completely excised the floxed stop cassette (data not shown).

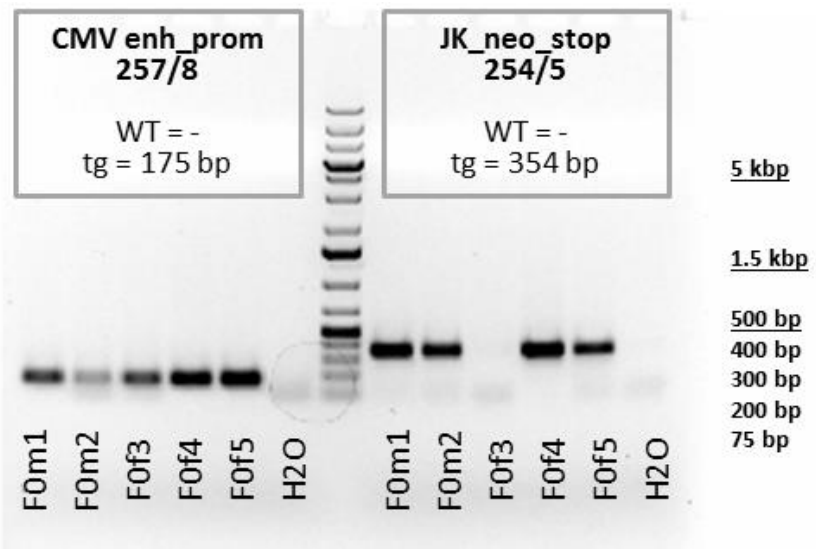


Figure 48:checking construct integrity in TALE-Lsd1 founders. *Specific primers against the CMV promoter-enhancer region (left) and the floxed neomycin stop-cassette (right) were used to confirm presence and integrity of these features in the founder generation. Founder F0f3 failed to yield a PCR-product for the stop cassette, indicating a truncation. H2O: negative control, Ladder: Generuler 1kb plus.*

To investigate if the TALE-Lsd1 construct is expressed after crossing with *Hprt-Cre* mice, RT-qPCR and flag-IP-WB were performed on embryonic material. mRNA levels of *GFP*, *TALE* and the control gene *Msx2* were measured after cDNA synthesis by qPCR (see ***Fig. 49***). Elevated mRNA levels were only observed in embryos from the F0f3 strain.

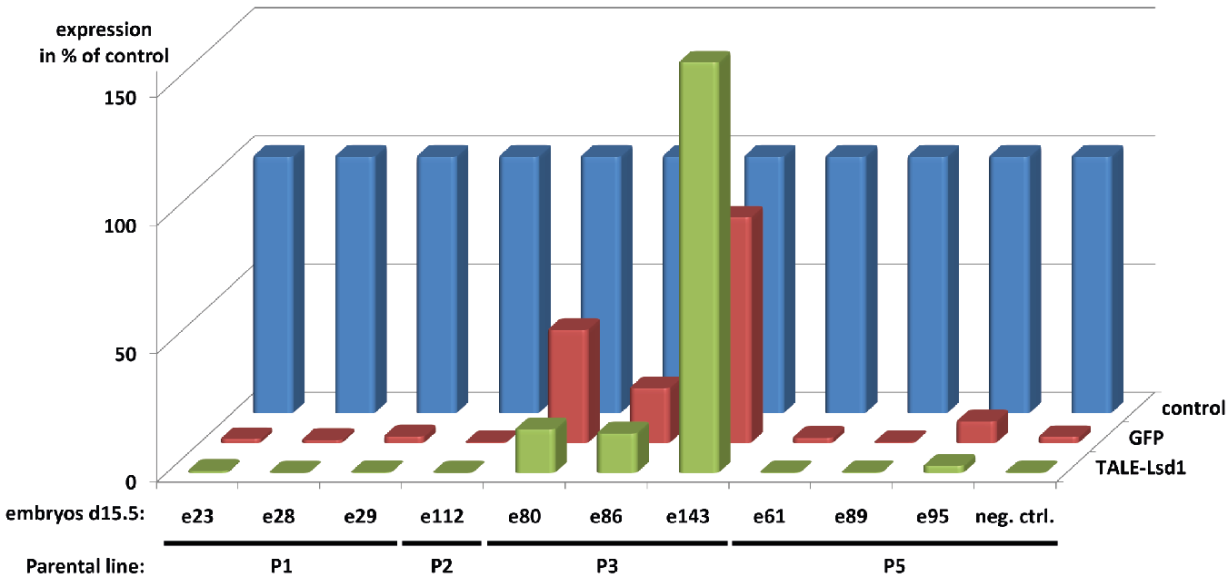


Figure 49: GFP and TALE mRNA levels in TALE-Lsd1 and Hprt-Cre double positive embryos. *Blue: control gene Msx2, red: GFP, green: TALE-Lsd1. Each mRNA-level is shown standardized to the control gene. Embryonic and parental strain numbers are indicated below. Neg. ctrl: TALE-Lsd1 negative, Hprt-Cre positive F2 embryo e90 from strain 5 (Founder F0f5). All double positive embryos from the F0f3 strain (P3) show increased transcription levels of the construct. More double positive embryos were analyzed in separate experiments with similar results (data not shown).*

To detect protein levels, WB with primary FLAG-antibodies was performed on embryonic tissue (see ***Fig. 50***). As no signal was detected, except the *FLAG-TALE-Lsd1* expressing HEK positive control, more *TALE-Lsd1* and *Hprt-Cre* double positive embryos were analyzed with additional flag-IP to detect low protein in WB. However, none of the samples, selected from all four *TALE-Lsd1* positive strains showed any specific FLAG-signal.

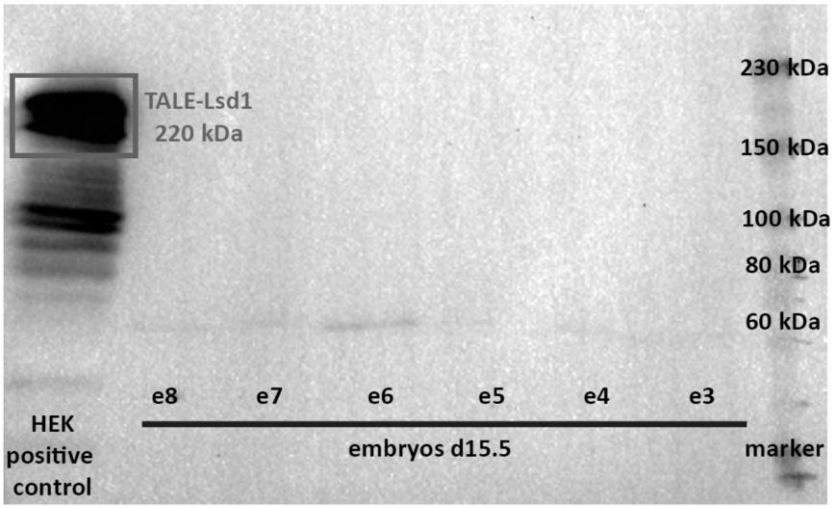


Figure 50: Detection of FLAG-TALE-Lsd1 protein levels by FLAG-WB*. Positive control: HEK 293T cells expressing the complete FLAG-TALE-Lsd1 construct (220 kDa). E15.5 embryos no. 3-8 from the F0m1 strain show no specific FLAG-signal. Embryo no. 3 is double positive for TALE-Lsd1 and Hprt-Cre, the other samples only for TALE-Lsd1. Protein ladder: Colorplus prestained protein marker, (NEB, P7711).*

To summarize, out of 8 generated transgenic animals, only 4 were founders to *TALE-Lsd1* positive strains, with only one strain showing measurable TALE-Lsd1 mRNA levels after crossing with *Hprt-Cre* mice, however, also here no TALE-Lsd1 protein product was detected. Limb development was normal in all animals. Overall, the *TALE-Lsd1* construct might have been integrated in silent, e.g. heterochromatic loci, so that no TALE-Lsd1 protein was produced.

Due to the inconclusive results, requiring the generation of more transgenic mice in combination with the newest published data on enhancerH3K4me1 and Lsd1[131], this project was terminated.

5.2 RNA Polymerase II-complex single locus analysis

The RNA Polymerase II (PolII) acts in a macrocomplex consisting of a variable set of proteins, including scaffold proteins, transcription and elongation factors, chromatin modifying enzymes, enzymes involved in cotranscriptional splicing and many more. Both the exact composition as well as amino acid modifications of the PolII-complex depend on a variety of nuclear pathways, including epigenetic marks, and also substantially change from initiation over elongation to termination of a gene's transcription[145,146]. While the core PolII complex with 12 subunits is well known, many of its transient participants and their modes of action remain unknown. Some of these interactors only bind the complex under specific conditions or at a restricted number of genomic loci[147,148]. As both, the transcription machinery modifies epigenetic marks of the transcribed genes and epigenetic marks themselves are involved in transcriptional regulation from initiation to termination, to investigate these interactions is a key aspect of epigenome research.

This project aimed at a better understanding of PolII-complex composition changes during the transcription of a single gene by chromatin immunoprecipitation in combination with Mass spectrometry (ChIP-MS) (see ***Fig. 51***). To target a single genomic locus at a time, a 3xFLAG tag was fused to specifically designed TALE oligomers (For details on TALE see section 5.1). Hereby single loci at promoters, exonic, intronic and terminal regions of several genes were targeted in separate, parallel ChIP experiments. The crosslinked chromatin with bound PolII complex and the FLAG-tagged TALEs was isolated and sheared to approximately 500-1000 bp which corresponds to a final resolution of 1000-2000 bp. Higher sonication intensity would increase resolution but diminish the input amounts for ChIP and the subsequent MS. In addition, the maximal sonication intensity was limited by the binding area of the PolII complex and the single stranded DNA around it, as TALEs bind only double stranded DNA. Following sonication, the TALE-bound region was enriched against the non-bound chromatin by a FLAG pulldown and proteins/DNA were isolated separately, either for subsequent MS or qPCR analysis.

As the TALE protein is originally a transcription factor from *Xanthomonas spp*., no inhibition of DNA binding by the active PolII complex was expected and at the same time the chromatin binding should be transient enough not to interfere with PolII activity. To assess the efficiency and specificity of TALE-targeting, two methods were available. Either colocalization analysis of fluorescent DNA *in situ* hybridization (DNA-FISH) probes with fluorescently labeled antibodies against the FLAG-tag (see ***Fig. 51C***), or a more elaborate approach by the PLA which would fluorescently label close proximity of the FLAG-tag with a DNA in situ probe (for a detailed description of PLA see section 5.3).

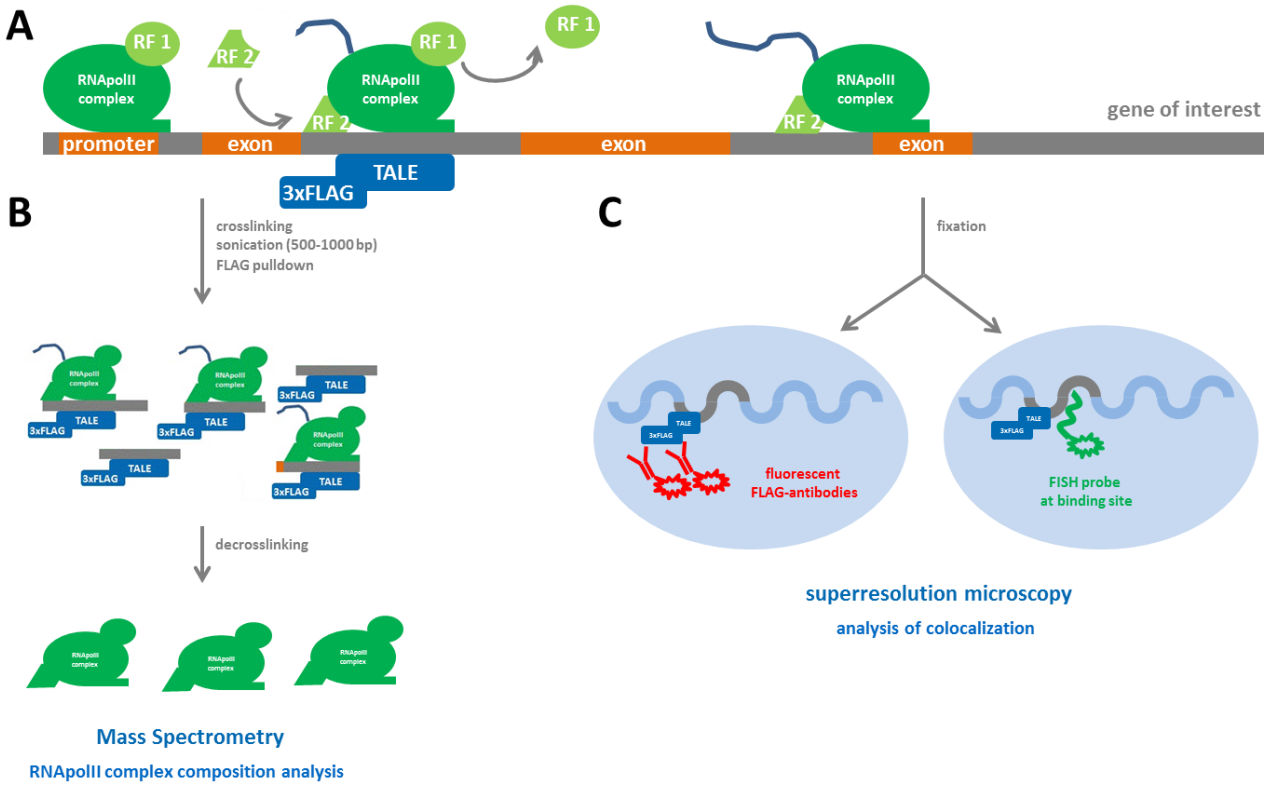


Figure 51: Experimental outline of TALE-targeted single locus analysis. A: *Composition of the RNA polymerase II complex (green) changes during transcription along the gene body (grey). Regulatory factors (RF) are transiently integrated into or released from the transcription machinery depending on positioning along the gene. The RNApolII complex can be captured at a single locus by crosslinking the chromatin to unique binding TALEs with a fused FLAG-purification tag (blue).* ***B:*** *Proteins crosslinked to the TALE-targeted locus are strongly enriched by ChIP using FLAG-antibodies and analyzed by mass spectrometry.* ***C:*** *Colocalization of FISH and fluorescent anti-flag Ab verifies unique TALE binding.*

To keep stress- and tissue-induced variation to a minimum as well as to provide high input material amounts, a cancer cell line was the system of choice. HeLa, the human cervical cancer cell line was selected because of its easy culturing conditions, fast growth and efficient transfectability. On the down side, the hypertriploid chromosome number of the HeLa cell line might lead to allele variation of chromatin- and PolII-bound factors, blurring the final picture[149]. Nevertheless, I expected to find a mostly common PolII-complex composition at specific regions of the same gene during active transcription.

I selected three genes, actin-ß (*ACTB*), lactate dehydrogenase chain B (*LDHB*) and myosin chain 9 (*MYH9*), each of which is well expressed in HeLa cells and have no transcript variants which might obscure the proteomic composition at a single locus (see ***Fig. 52***). These genes differ strongly from 3.5 to 100 kb in length, which might uncover differential PolII complex composition depending on gene length and would also strongly reduce resolution based limitation for the two longer genes, as larger chromatin pieces could be used, increasing material amounts without overlapping with the next targeted site.

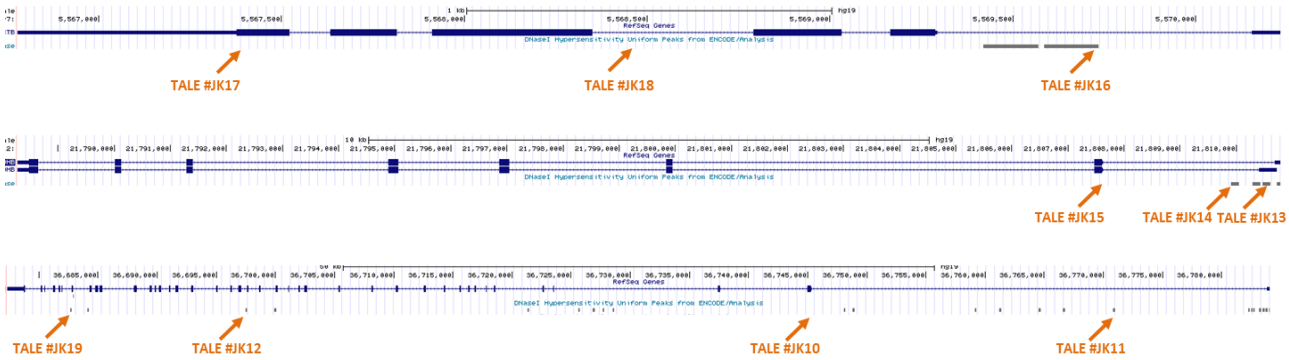


Figure 52: Selection of TALE targets in ACTB (top), LDHB (center) and MYH9 (bottom). *Where possible (TALE #11,12,13,14,16,19), target sites were selected at DNAseI sensitive sites (grey bars) to employ increased chromatin accessibility of these sites. Scale bars: ACTB 1kb, LDHB 10 kb, MYH9 50 kb.*

The 19 bp long TALE binding sites were selected to cover different genetic positions of the gene, including promoter, intronic and exonic regions (see ***Table 3***). Where possible (6 out of 10 binding loci), DNAseI sensitive sites were chosen for an enhanced accessibility of the chromatin for the TALE-FLAG protein. To avoid off-target binding, the sites were also selected for the highest genome-wide dissimilarity by BLAST. DNA-methylation sites were avoided, as the used TALE monomers have a binding preference for nonmethylated DNA.

construct	**gene**	**position**	**Sequence**	**DNAseI - site**
TALE #10	*MYH9*	2nd exon	TGAGTAGTAACGCTCCTTG	no
TALE #11	*MYH9*	1st intron	TGAACGAGGTCCCACCCCT	yes
TALE #12	*MYH9*	19th intron/exon border	TCAGTCCAGGTCGCACGAC	yes
TALE #13	*LDHB*	5' UTR/1st intron	TGCGCCCCAAACTGAGCGG	yes
TALE #14	*LDHB*	1st intron	TACTGCTCGCGGTGCGCAC	yes
TALE #15	*LDHB*	2nd exon	TCCGCAACTGGTGCAATGA	no
TALE #16	*ACTB*	1st intron	TCAAGGCGCTAACTGCGCG	yes
TALE #17	*ACTB*	last exon	TCATAGTCCGCCTAGAAGC	no
TALE #18	*ACTB*	3rd intron	TACACACTCCAAGGCCGCT	no
TALE #19	*MYH9*	6th-last intron/exon border	TCAGCTGAAAGCCCCACGC	yes

Table 3 Selected TALE-binding sites and their properties.

The TALE monomers were assembled by the Golden Gate TALE kit and finally cloned to eukaryotic expression vector including a strong *CMV*-promoter, *eGFP* to monitor construct expression and linked by the autohydrolyzing P2A domain, the N-terminal *3xFLAG*-tag, a nuclear localization sequence (*NLS*) and an eukaryotic antibiotic resistance gene (*HygR*) to establish and maintain a stable cell line (see ***Fig. 53***). The correct plasmid assembly was confirmed by Sanger sequencing and restriction analysis.

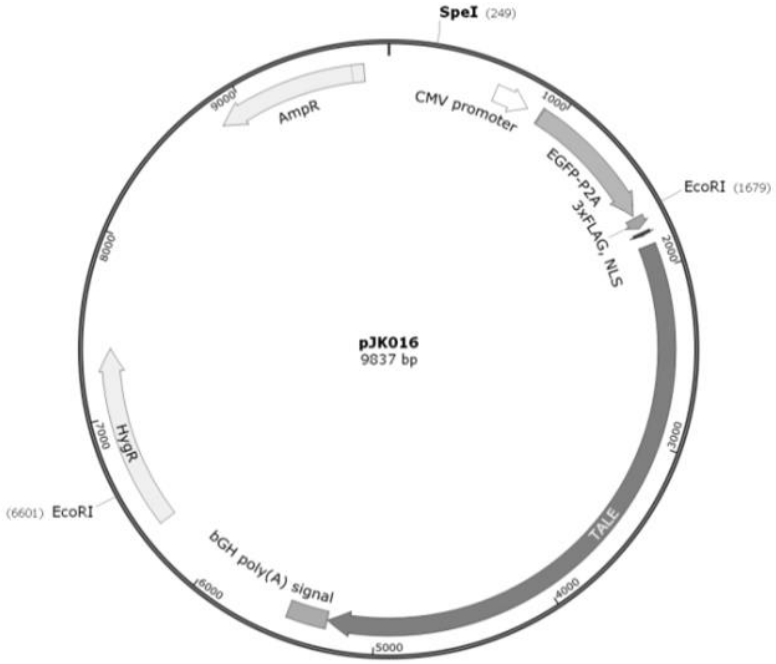


Figure 53: Full assembly of TALE vectors for single locus ChIP-MS. *The eGFP-P2A-FLAG-TALE construct is expressed from the strong eukaryotic CMV-promoter. eGFP is cleaved at the p2a site directly during translation to allow fluorescence based monitoring of expression levels without interfering with TALE-binding. The 3flag-tag and the nuclear localization signal (NLS) are cloned after the P2A site and hence remain at the N-terminal TALE-domain. For selection AmpR (prokaryotic) and HygR (eukaryotic) resistance genes are available.*

TALE-Flag constructs were introduced to separate cell lines and selected for stable expression by Hygromycin. To obtain clonal cell lines with high and homogenous protein levels, single eGFP+ cells were sorted by FACS or handpicked and eGFP levels in growing clones were regularly monitored by fluorescent microscopy. Further, the production of intact FLAG-TALE was checked by WB and best clones were selected for subsequent ChIP.

In preparation of MS analysis, different flag-ChIP protocols were tested and optimized. The assessment of ChIP efficiency and specificity was conducted by qPCR and WB.

5.2.1 Results: RNA Polymerase II-complex single locus analysis

5.2.1.1 TALE-assembly

All 10 TALE constructs were correctly assembled in the *pcDNA-eGFP-P2A-3xFLAG* vector. Constructs #10-18 were cloned into the final hygromycin resistance vector. All constructs were confirmed by restriction analysis and sequencing, stocks were preserved at -80 °C.

5.2.1.2 Generation of stable TALE-HeLa cell lines

HeLa cells were transfected with constructs #10-18 and selected on Hygromycin. To get pure single clones, GFP-positive colonies were handpicked and expanded separately under further selection. The picked clones represented the whole range from bright to very faint GFP-signal, partially uniform across a clone, partially showing fluorescent patches. Most clones showed mixed GFP-positive/negative populations with in some cases as little as 20% positive cells, therefore they were additionally sorted by FACS and further selected on Hygromycin (see ***Fig. 54***). In most cases, the GFP-high population was further cultured. However in some cases, the GFP-high population was not proliferative and GFP-low cells were propagated instead.

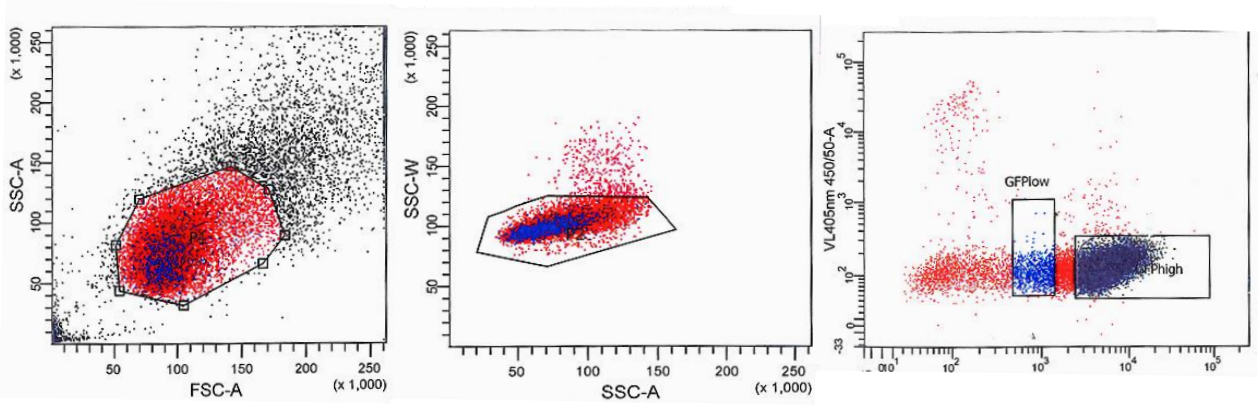


Figure 54: TALE-clone purification by FACS. *As most clones showed mixed, GFP positive/negative populations, they were sorted for signal intensity to obtain uniform, GFP-positive populations. SSC-A / FSC-A: selection for cell sized signal. SSC-W / FSC-A: doublet exclusion. DAPI: dead cells. GFP: GFP-signal corresponds to FLAG-TALE expression. Here a high-GFP example is shown, while some other clonal sorts measured mainly GFP-low cells or as little as 10 % GFP-positive cells in total.*

Sorted clones were further monitored for a constantly high fraction of GFP-positive cells and stable GFP levels. Stable clonal populations were further analyzed for FLAG-TALE protein levels by WB (see ***Fig. 55***). Clones with detectable levels of correctly sized FLAG-TALE proteins and without substantial contamination by truncation or degradation products were preserved at -150 °C and used for further experiments. Clones meeting all these requirements were obtained for TALE-targets #10-15 and #18.

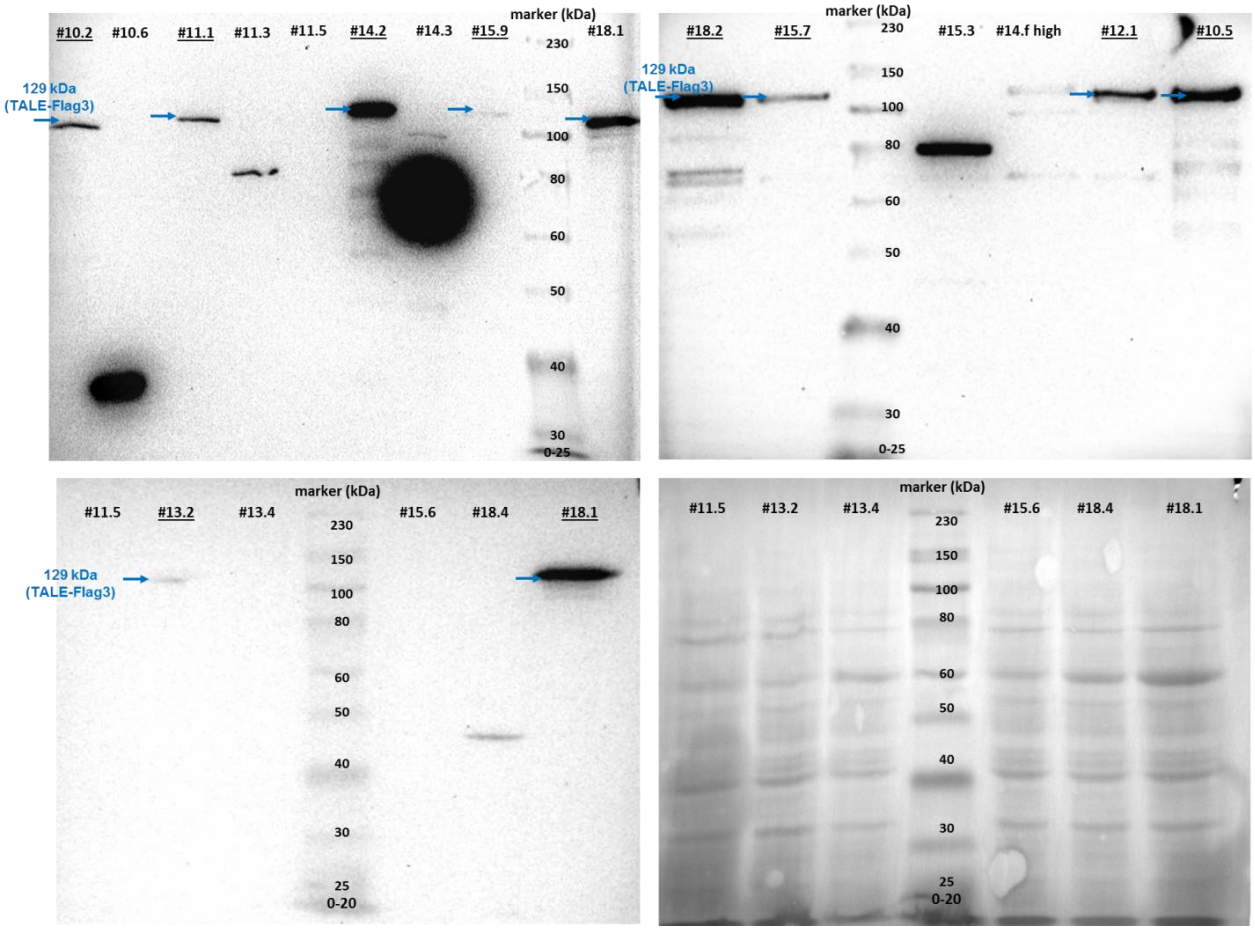


Figure 55: Western blot analysis of GFP-positive clones. *Each clone was checked for FLAG-TALE protein levels and correct size, indicated by blue arrows. Correctly sized clones, with detectable expression and low degradation product levels, are underlined and were used for further experiments. To ensure equal loading, total protein levels were measured by Nanodrop 2000. After 8-9% SDS-PAGE, equal loading was confirmed by Ponceau red staining, as exemplified on the lower right for the WB on the lower left. Protein ladder: Colorplus prestained protein marker, (NEB, P7711).*

5.2.1.3 Cellular GFP and FLAG-TALE localization

While *FLAG-TALE* expression could be monitored directly by GFP protein levels, the nuclear localization of the FLAG-TALE protein, which is necessary to bind to chromatin, was assessed by an IF staining and confocal microscopy (see ***Fig. 56***). Hereby cytoplasmic eGFP was imaged directly, while FLAG-TALE was targeted with an AF594 conjugated secondary Ab. Nuclei were counterstained by DAPI. In the analyzed HeLa clones #14.2 an #18.2, as expected, GFP localized throughout the cell, while FLAG-TALE signal strongly enriched in the nucleus, confirming NLS-functionality. IF imaging additionally demonstrated the clonal heterogeneity of flag-TALE levels: although by selection and clonal screening >90 % of the cells are GFP and FLAG-TALE positive, protein levels vary.

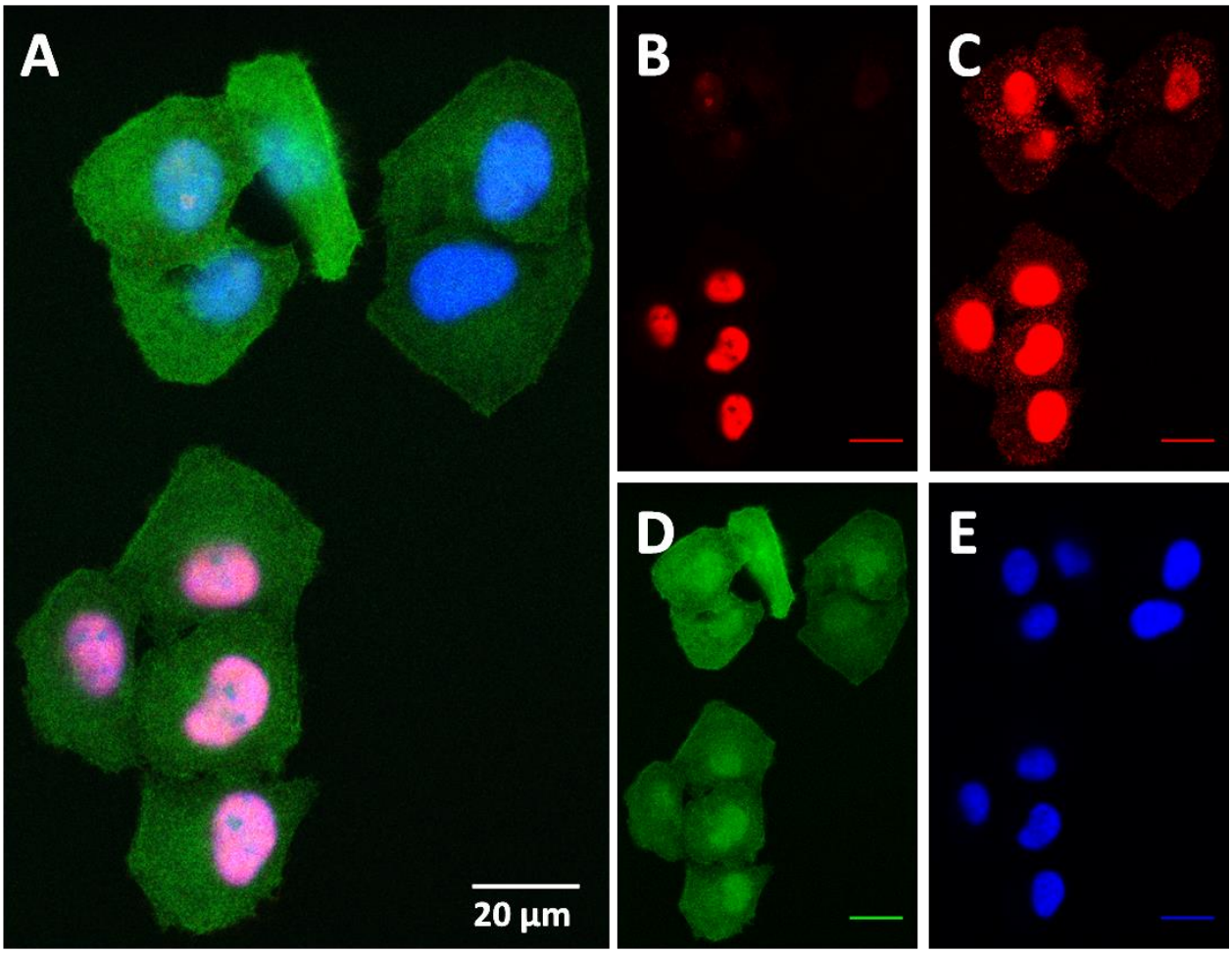


Figure 56: Flag-TALE localization in the HeLa clone # 14.2. *While GFP is cleaved off after translation and hence is detected throughout the cytoplasm, FLAG-TALE is imported into the nucleus.* ***A:*** *overlay of channels B, D and E.* ***B:*** *FLAG-TALE IF signal, mouse anti-FLAG primary Ab., AF594 secondary Ab.* ***C:*** *enhanced contrast on channel B, cells with lower FLAG-TALE levels also show nuclear localization.* ***D:*** *GFP signal.* ***E:*** *Nuclei counterstained by DAPI. Imaged by SP5 confocal microscope*

5.2.1.4 Establishing FLAG-ChIP

FLAG-ChIP was performed on a subset of FLAG-TALE positive clones and ChIP efficiency was determined by WB and qPCR. First, the overall enrichment of FLAG-TALE during ChIP was analyzed by WB (see ***Fig. 57A***). As can be seen by the high intensity ChIP band in comparison to the input, which were loaded at an input to ChIP material ratio of 1:9, the IP of FLAG-TALE was very efficient. Degradation products might have appeared due to prolonged culturing of the clone and were further avoided by regular culture resetting from stock.

Although FLAG-TALE enrichment by ChIP was very efficient, co-precipitated DNA levels were only1.5-2-fold enriched over off-target regions (see ***Fig. 57B***). In combination with the efficient FLAG-TALE precipitation shown by WB, this indicates a transient binding of TALE to chromatin, so that by far not all target loci are crosslinked to a FLAG-tag during ChIP, while high overall FLAG-TALE levels might lead

to an increasing ChIP-background signal. Besides general optimization of the ChIP protocols with buffer and antibody/bead ratio adjustments, the ChIP enrichment could be improved by the usage of clones, producing lower FLAG-TALE amounts.

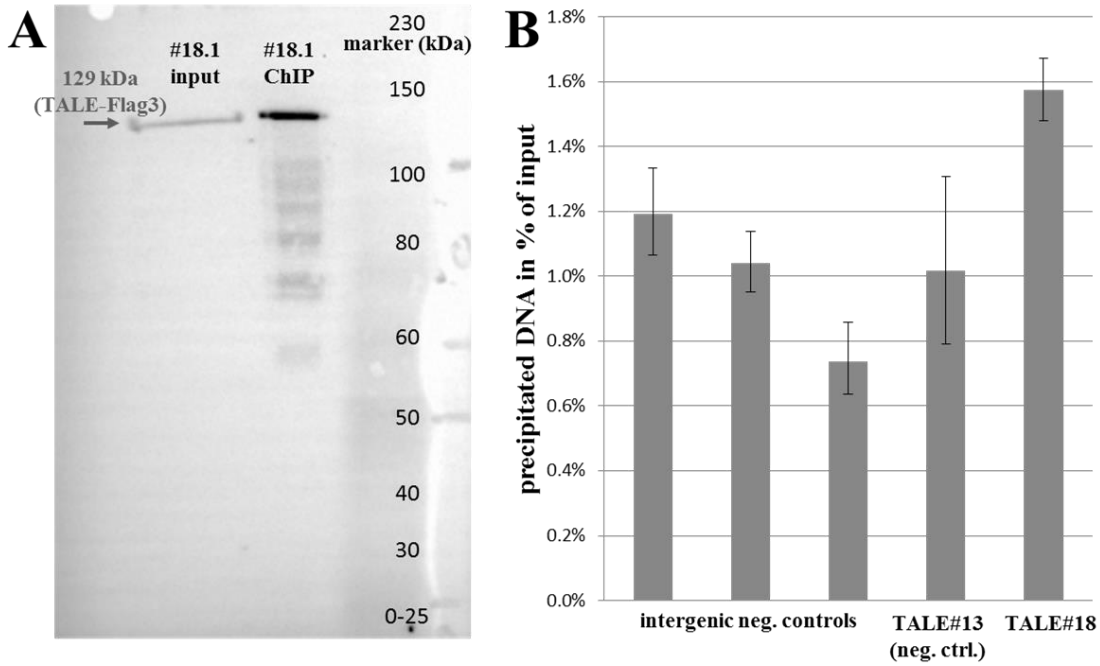


Figure 57: WB and qPCR of FLAG-TALE ChIP and input material of clone #18.1 A: *WB with 9 times more loaded ChIP sample in comparison to the input. Correct sized FLAG-TALE band is indicated by blue arrow. Protein ladder: Colorplus prestained protein marker, (NEB, P7711).* ***B:*** *qPCR-results, precipitated DNA as percentage of input signal. 3 intergenic regions (neg. ctrl.), TALE-target non-corresponding to the analyzed #18.1 clone (neg. ctrl.) and the correlating TALE-target. Error bars show triplicate standard deviation.*

5.2.1.5 Mass spectrometry

MS-sample preparation was tested during a training session with the IMB proteomics core facility by its standard protocols for SILAC. Therefore, FLAG-ChIP samples and inputs were digested, purified and total oligopeptide amounts measured. Although sample preparation went well, total ChIP-sample amounts, prepared from 5*10^6 cells were too low to use them as MS-input. The required amounts for an expectedly successful MS-run, uncovering rare/transient PolII-interactors would be with at least 500 ng after IP 1000-fold higher. To harvest these amounts of ChIP-output, substantially higher amounts of used cells as well as a very strongly improved FLAG-ChIP protocol would be necessary.

Regarding the possibilities of IP protocol optimization, I estimated max. 10-fold enrichment realistic due to the probably transient binding of TALE to a single locus. Due to limited potential of ChIP-protocol improvement as well as the immense costs of performing a SILAC experiment on hundreds of cell culture plates necessary to provide sufficient ChIP-material, this project was terminated.

5.3 PLA DNA-protein

Whole genome analysis by ChIP-seq (see section 4) has one major drawback: it always generates averaged data from a cell population. Although fluorescent labeling of cell type specific markers coupled with FACS allows using pure populations as ChIP-input, complete homogeneity cannot be achieved. Hence, no conclusions can be made on single cell level and potentially meaningful heterogeneities in the given cell population go unnoticed. To take a closer look at this blind spot of genome-wide population analysis, I selected several genomic loci for epigenetic single cell analysis by the PLA. Hereby, the colocalization of an epigenetic mark to a DNA ISH-probe targeting the genomic locus is assessed at single cell resolution by fluorescent microscopy on tissue sections.

The PLA-target site selection was performed based on RNA-seq and ChIP-seq data. Of special interest were genes with alternative transcription start sites (TSS) and differential active epigenetic mark positioning during intestinal development. Alternative TSS provide an additional level of gene regulation, yielding transcripts of different length and potentially different exon composition. Different pathways can sequentially or simultaneously regulate the multi-TSS gene's expression, as alternative TSS have different regulatory sequences and can be associated with different enhancers. Same as for alternative enhancers, the activity of each alternative TSS in a particular cell can be independently regulated by a variety of factors, such as developmental stage, cell type or the tissue specific signaling context[150].

The obtained ChIP-seq data for the active promoter marks H3K4me3 and H3K27ac showed that alternative TSS of many genes change asynchronously during intestinal development from E12.5 to fully differentiated AE. One example for this is the *Hnf4a* gene, which encodes a master transcription regulator (see ***Fig. 58***). While at embryonic stages TSS-1 is preferentially marked and used with an activity peak in E14.5 and ISC, the second TSS, which yields a shorter transcript, is not detectable until E14.5 and has increasing active marks until the AE stage, which also corresponds to RNA-seq data.

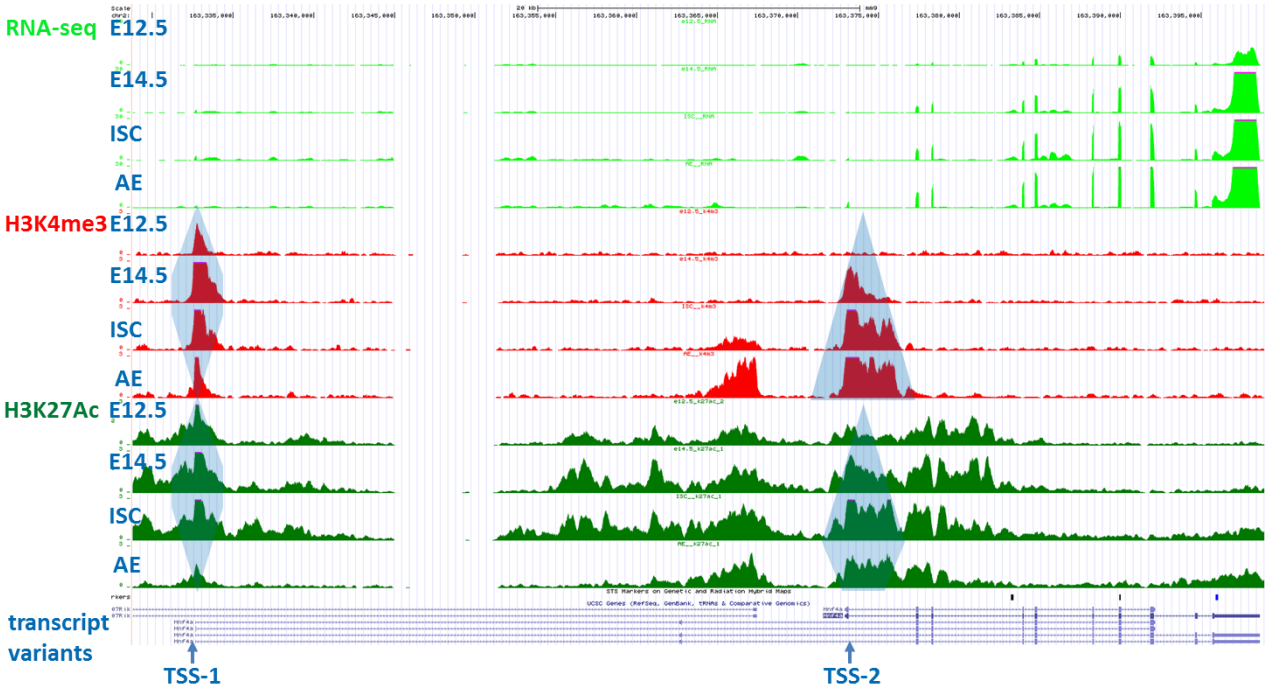


Figure 58: Epigenetic patterns on the Hnf4a locus. *Alternative transcription start sites TSS1/2 show asynchronous activity development during gut differentiation from E12.5 and E14.5, over ISC to the finally differentiated AE. Tracks: RNA-seq, H3K4me3 ChIP-seq and H3K27acetyl ChIP-seq results for 4 developmental stages. Bottom: transcript variants of Hnf4a. Blue shading indicates the differential positioning of epigenetic marks H3K4me3 and H3K27ac on the alternative TSS of Hnf4a. Scale bar 20 kb.*

It is possible, that the asynchronous epigenetic pattern development of alternative TSS represents distinct cellular subpopulations rather than the development in single cells of a uniform population. Although cells appear homogenous by the used FACS cell type markers, some, or even all of the epigenetical marks on a single TSS might be contributed only by a subpopulation. These subpopulations might be associated with lineage commitment and/or defined by their localization in the tissue and hence also by signaling from neighboring cells.

In situ analysis of TSS activity and epigenetic patterns require labeling methods with very high sensitivity and specificity. Although DNA-FISH at the target locus and IF staining against the histone modification of interest are very specific by themselves, histone marks, such as H3K27ac, are found at many loci and would give many signals throughout the nucleus, making colocalization analysis with the ISH-signal impossible. Instead of FISH and IF, PLA was chosen, a method which only yields a strongly enhanced fluorescent signal if both, the ISH probe and the protein of interest are at close proximity, but no signal at all, if only one of the targets is present (see ***Fig. 59***). This method is well established for the visualization of protein-protein interactions and was recently reported to be also suitable for the detection of DNA-protein and RNA-protein interactions[151,152].

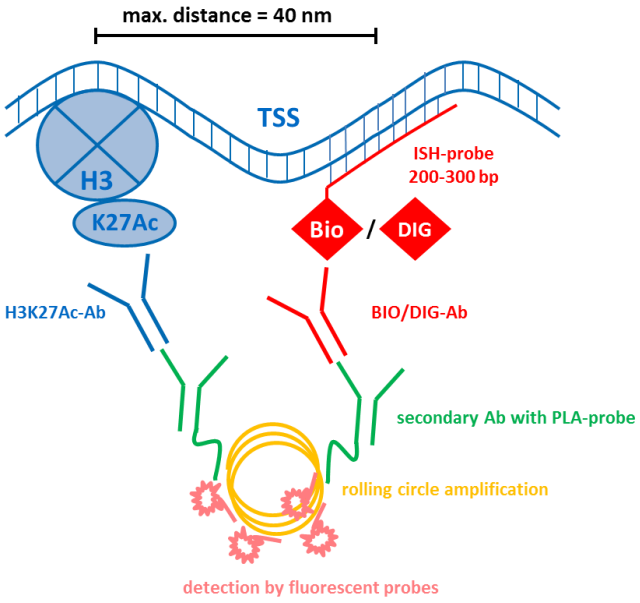


Figure 59: Scheme of the Proximity Ligation Assay (PLA). *Biotinylated (BIO) or digoxigenin-labeled (DIG) ISH-probe (red) and the active promoter mark H3K27Ac are targeted by primary antibodies, which are in turn recognized by secondary antibodies linked to two different PLA-probes (green). If both antibody-linked probes come close together, they hybridize with linker DNA and circularize it, allowing for rolling cycle amplification by a DNA-polymerase (orange). Fluorescent in situ probes (pink) then detect the amplification product. PLA signal only occurs if the ISH probe and the targeted epitope, here a histone mark, are closer than 40 nm, otherwise the linker DNA is too short to reach both PLA-probes and no signal amplification can occur.*

5.3.1 PLA target locus selection

Potential PLA target sites were preselected by clustering H3K4me3 ChIP-seq positive promoters for differential H3K4me3 signal intensity during the four developmental stages from E12.5 to AE (see section 4.3.9.2 for details on clustering). Differential alternative TSS were further screened for asynchronicity, sufficient distance between each other and to other possibly marked elements. Finally, 3 regulatory genes were selected: *Hnf4a*, *Arid5b* and *Smarca2*, each with >30 kb between TSS, which is sufficient to avoid false positive PLA signal by random colocalization (see ***Fig. 58 and 60***).

As positive control, the 200 kb long bacterial artificial chromosome (BAC) RP23-118H24 on chromosome 7 was used, which contains the gene *Kcnq1ot1*. This gene is actively transcribed at E12.5 and E14.5 and hence the BAC should provide a control signal not only for ISH but also for the whole PLA. Additionally, *Shh* and *Epcam*, two control genes with known tissue localization during development, were targeted to establish the PLA protocol prior to alternative TSS analysis.

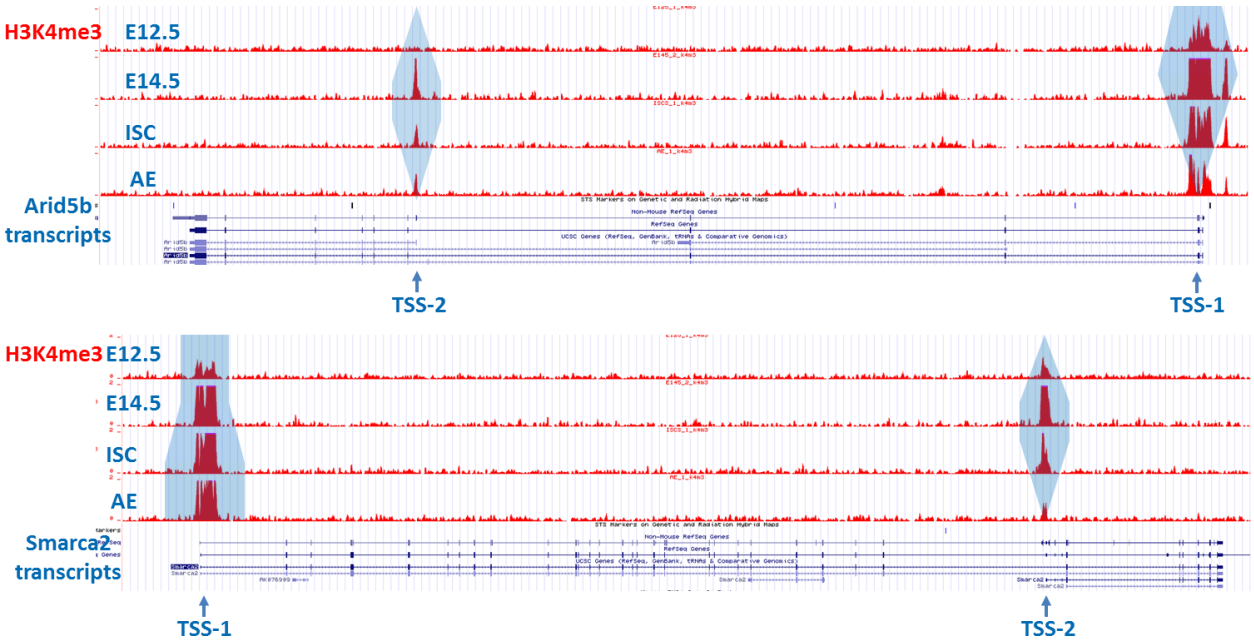


Figure 60: Asynchronous alternative TSS of Smarca2 and Arid5b genes. *Red: H3K4me3 tracks from gut differentiation stages E12.5, E14.5, ISC and AE are shown. Blue: transcript variants of Arid5b (top) and Smarca2 (bottom). Blue shading indicates the differential positioning of the epigenetic mark H3K4me3. Full gene length: Arid5b 183 kb, Smarca2 173 kb.*

Genes selected for their differential and asynchronous H3K4me3 positioning as well as the controls, were confirmed for similar H3K27ac behavior, as it was expected from the high correlation of these marks on promoters (see ***Fig. 58*** and section 4.3.5). Even more, H3K27Ac generally shows a broader promoter peak distribution than H3K4me3, increasing the chance of a locus-specific PLA-signal by proximal looping.

ISH probes, later on to be labeled with biotin or digoxigenin, were designed with the following criteria: localization under the H3K4me3/H3K27ac peak, 200-300 bp length, unique probe and primer sequence in comparison to murine genomic DNA (gDNA).

5.3.2 Results: DNA-Protein PLA

5.3.2.1 ISH probe preparation

DNA-ISH probes were generated for the two alternative TSS of *Smarca2*, *Hnf4a* and *Arid5b* as well as for the control TSS of *Shh* and *Epcam*. Probes were labeled separately with digoxigenin and biotin (sizes and positions are summarized in section 8.3). The positive BAC control, sonicated to 200-1000 bp, was labeled either with biotin for PLA, or with AF594 for FISH.

Correct DIG-11-dUTP labeling was confirmed by gel electrophoresis (see ***Fig. 61A***). Hereby, the band upshifted in correlation to labeling efficiency. Highly DIG-labeled probes (35 % instead of 10 % DIG-dUTP in labeling reaction) ran highest on gels, but exhibited a lower PCR-efficiency and were also shown to perform worse in PLA. The observed lower signal number was probably due to sterical hindrance of ISH or Ab binding at densely labeled loci.

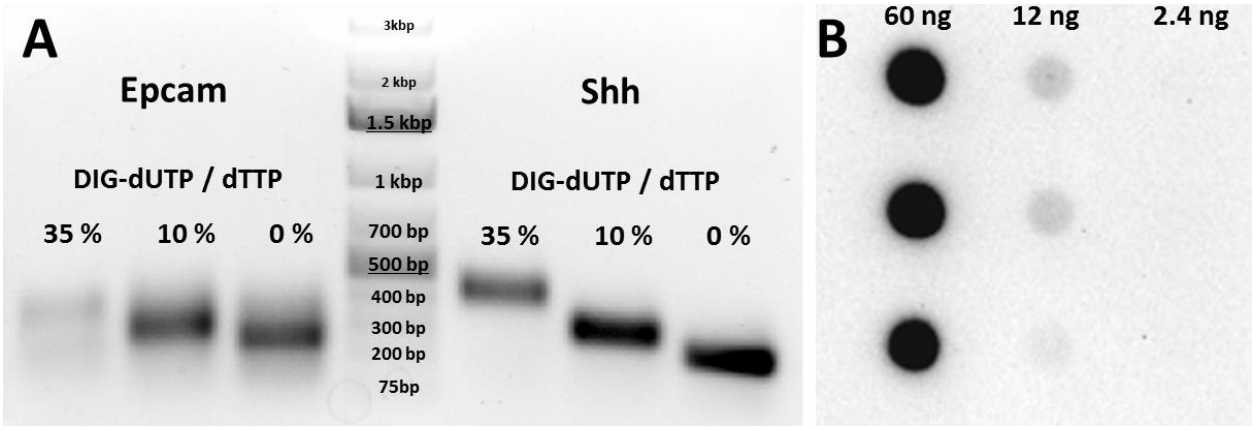


Figure 61: ISH-probe labeling with DIG and BIO. A: *Agarose gel electrophoresis of DIG-labeled probes and unlabeled control reactions. Probes for Epcam and Shh were amplified by PCR with incorporation of 10 % (usual protocol), 35 % DIG-11-dUTP or dTTP only. The band shift correlates to labeling intensity.* ***B:*** *Dot Blot of BIO-labeled BAC. Triplicates of 5-fold dilution series of BIO-BAC were applied to the membrane and imaged for Streptavidin-HRP activity. The amount of BIO-BAC in the initial, undiluted sample was estimated by Nanodrop.*

BIO-labeling was confirmed by Dot Blot with horse radish peroxidase (HRP)-conjugated streptavidin (see ***Fig. 61B***). The BIO-BAC sample was applied to the membrane in a dilution series (first 3 dilutions shown) and corresponding decreasing signal was detected. Non-labeled BAC showed no signal. The positive BIO-labeled DNA control, provided with the labeling kit, also showed no detectable signal, as it was much lower concentrated (2 ng max. per well), hence absolute quantification of labeling efficiency was impossible.

AF594 labeling of BAC was confirmed by spectrophotometric measurements with Nanodrop 2000 and Nanodrop 3300. Based on these measurements, labeling efficiency was estimated with 3000 fluorophores per 200 kb of the BAC locus.

5.3.2.2 DNA-FISH

FISH with AF594-BAC and secondary IF-staining on BIO-BAC ISH confirmed efficient ISH during the PLA protocol (see ***Fig. 62***). In general, FISH and PLA signals were only detectable down to 4 µm into the sectioned, paraffin embedded tissue, hence 5 µm sections were used. The BAC locus was visualized close to the border of nuclei throughout the imaging experiments. AF488 secondary staining on BIO-BAC allowed an estimate on hybridization efficiency: 1100 signals / mm^2 epithelium were detected, which corresponds to 14% BIO-BAC ISH efficiency in 5 µm sections. (For calculations see section 7.12.5.3).

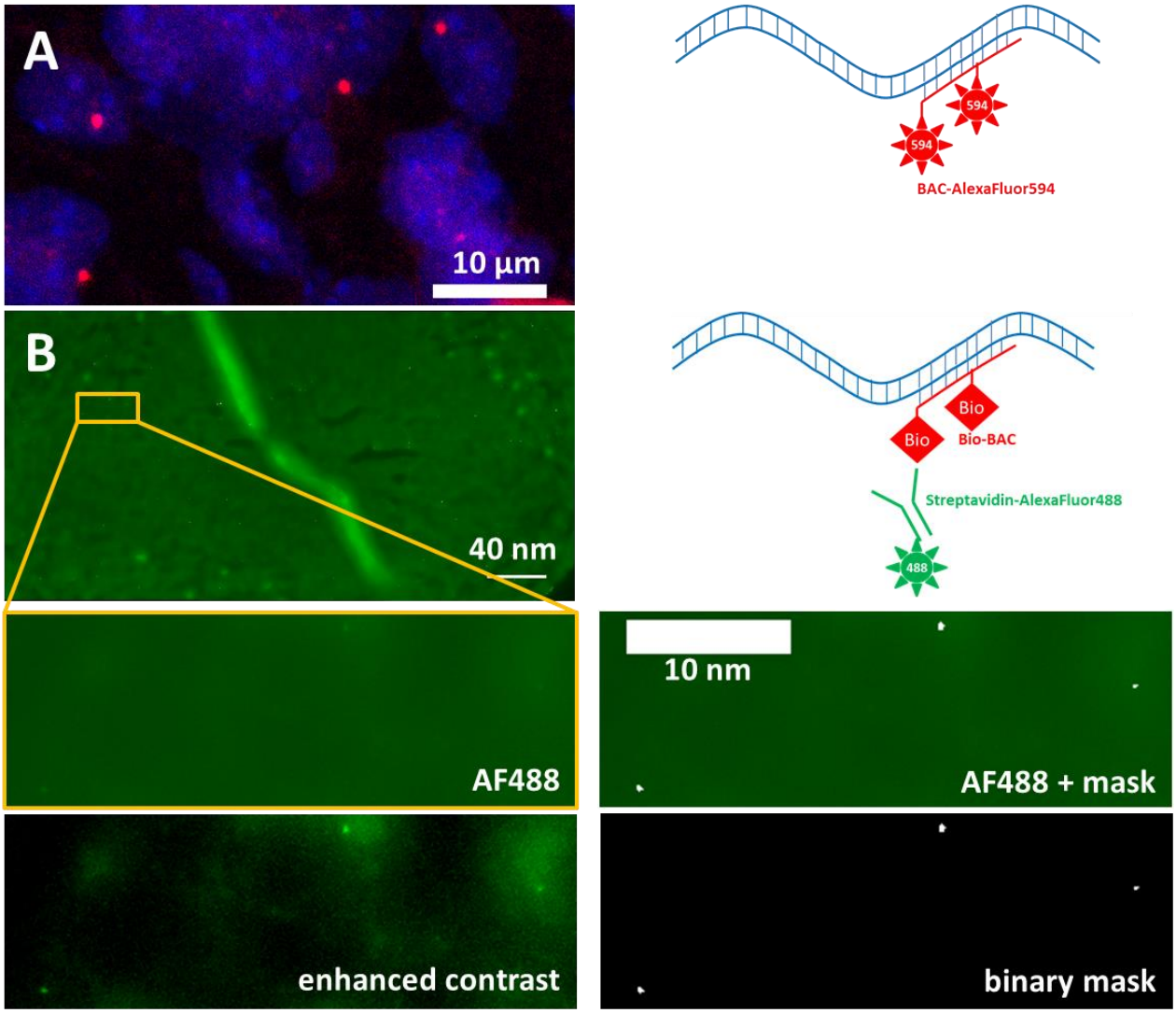


Figure 62: BAC ISH on tissue sections. *The BAC ISH-signal is generally found at borders of nuclei. The maximum tissue section penetration is 4µm, no signal is detected below.* ***A:*** *FISH with AF594-BAC (red dots). Cell nuclei by DAPI counterstaining (blue). Enhanced contrast of a maximum intensity Z-stack overlay imaged by SP5 confocal microscope.* ***B:*** *Streptavidin-AF488 IF staining against BIO-BAC ISH. Maximum intensity Z-stack overlay imaged by AF7000 widefield fluorescent microscope. Weak fluorescent signal (left) was evaluated by Gaussian blur based computational processing yielding a binary signal mask (right). IF on ISH showed 1100 signals per mm^2 of the dissected gut epithelium. Right panel: Scheme of ISH-experiments presented in A (top) and B (bottom).*

5.3.2.3 Primary histone mark antibodies for PLA

Most commercially available antibodies are validated for an array of standard protocols, such as WB, ChIP or IF. However, the specific sample preparation requirements and handling during the PLA protocol strongly deviate from those procedures and make additional validation necessary. I performed IF staining with H3K4me3, H3K4me2 and H3K2Ac antibodies from different manufacturers and batches under PLA conditions, including a mock ISH (full ISH protocol, H_2O instead of a probe), to assess their compatibility with this method. Specificity of nuclear localization was assessed by fluorescent microscopy. None of the assessed H3K4me3 and H3K4me2 antibodies showed signal enrichment in the nuclei throughout the tested concentrations and conditions and hence are not suitable for PLA (see ***Fig. 63***).

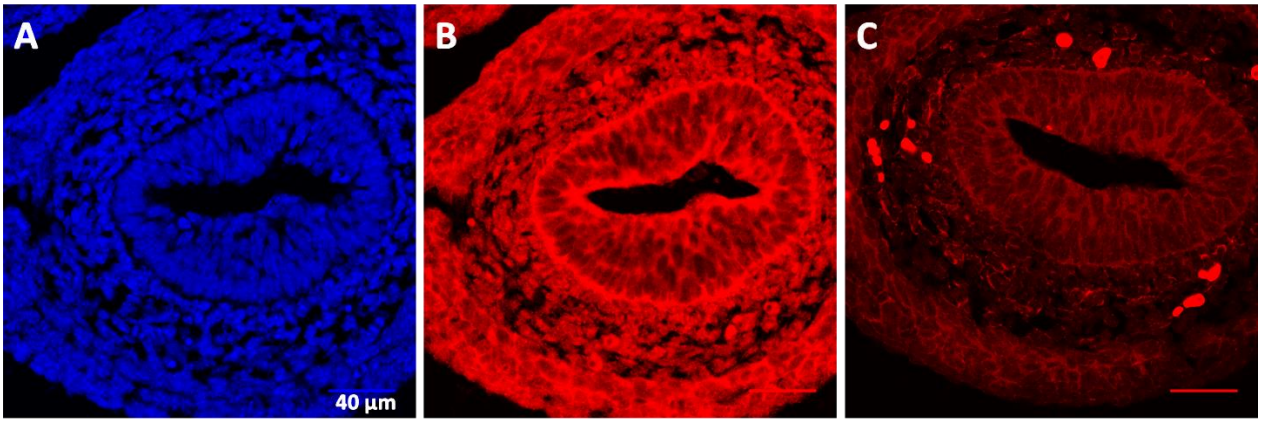


Figure 63: H3K4me3 antibody test under PLA conditions. *Maximum overlay of IF staining on E14.5 paraffin section, imaged by confocal microscopy.* ***A:*** *nuclei counterstained by DAPI.* ***B:*** *Secondary Ab conjugated to AF594, against primary H3K4me3 rabbit Ab (Abcam #8580).* ***C:*** *Background control, only secondary AF594 Ab added. Other H3K4me3 and H3K4me2 antibodies showed similar signal distribution or no fluorescence enrichment above background at all.*

On the other hand, the H3K27Ac Ab showed specific nuclear colocalization with DAPI over a range of concentrations used and with only minor deviations between batches. As H3K27Ac is strongly correlating with H3K4me3 at active promoters, this Ab was further used in this project. H3K27Ac specificity was confirmed by secondary and tertiary immunofluorescent labeling (see ***Fig. 64***).

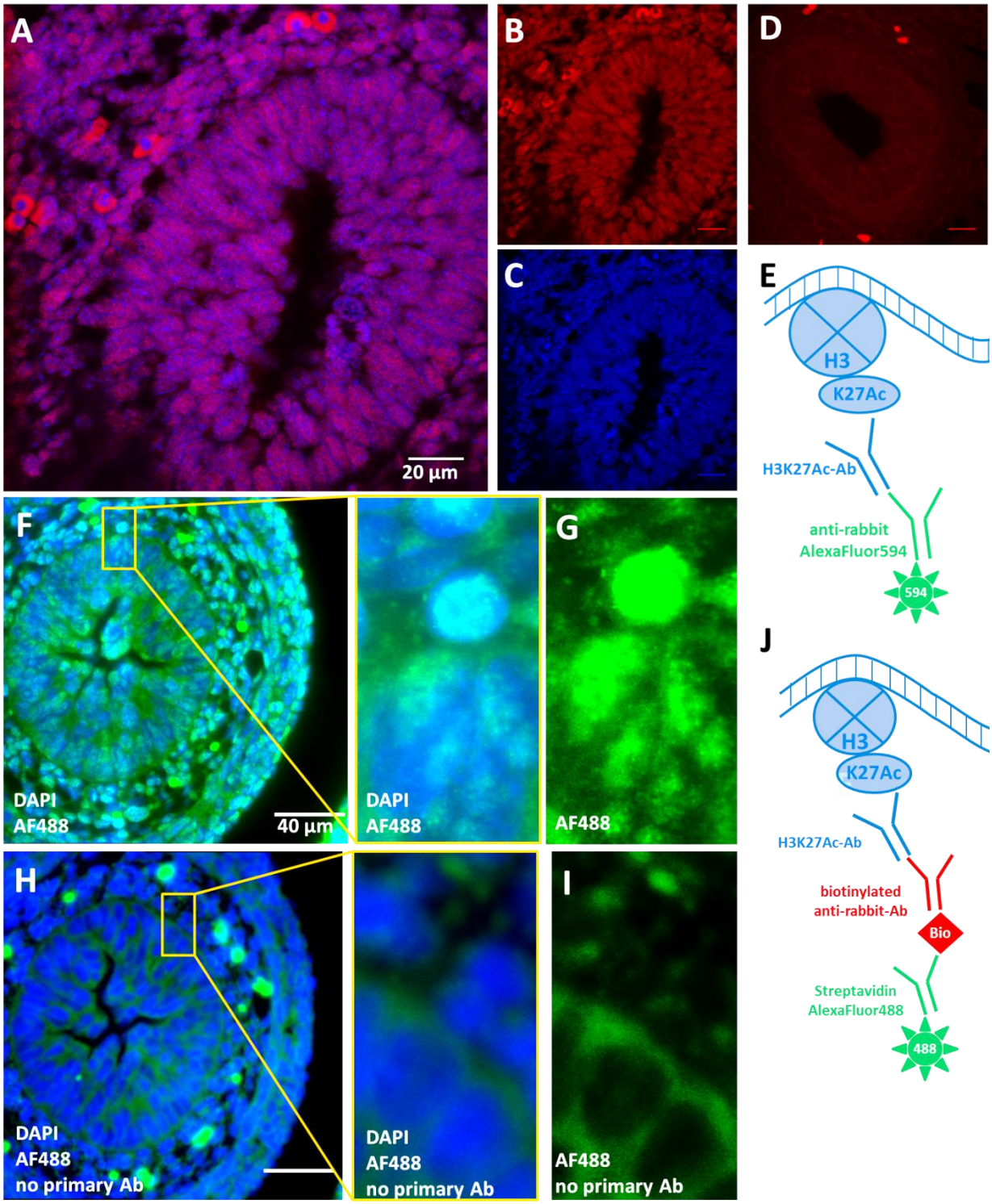


Figure 64: H3K27Ac antibody is specific under PLA conditions. A-D: *Maximum overlay of IF signal on E13.5 paraffin sections, imaged by confocal microscopy.* ***A:*** *overlay of B and C.* ***B:*** *AF594-conjugated Secondary Ab against primary H3K27ac Ab (Abcam #4729).* ***C:*** *nuclei counterstained by DAPI.* ***D:*** *Background control without primary Ab.* ***E:*** *Experimental setup for A-D.* ***F-I:*** *Maximum overlay of IF signal on E14.5 paraffin sections, imaged by widefield fluorescence microscopy.* ***F:*** *AF488- conjugated Streptavidin (Invitrogen S32354), secondary biotinylated Ab (Biozol PK6101) against primary H3K27ac Ab (Abcam #4729). Nuclei are counterstained by DAPI.* ***H:*** *Background control without primary Ab.* ***I,G:*** *Enhanced contrast AF488 signal of the enlarged area in F and H.* ***J:*** *Experimental setup for F-H.*

5.3.2.4 Primary biotin and digoxigenin antibodies for PLA

Same as histone mark antibodies, different antibodies targeting ISH probes marked by biotin were tested for their compatibility with the PLA protocol. After BIO-BAC ISH and incubation with the tested anti-biotin Ab, a fluorescently labeled secondary Ab was applied and imaged (see ***Fig. 65***). The best tested anti-BIO (Biozol SP3000) showed with 3000 signals per mm² of 5µm thick tissue an IF- efficiency of ~36 % (calculations as in section 5.3.2.2). This is substantially higher than the measured ISH efficiency determined above, which can most likely be explained by the use of a secondary Ab, which makes the detection more sensitive. However, detection of rare colocalization events would not be possible because of intrinsic biotin, which is also recognized by this Ab. According to the data presented in ***Fig. 65*** as much as one third of the detected signals also appear without a biotinylated probe. Biotin-labeled probes, especially BIO-BAC in combination with this Ab were further used for the establishment of PLA, but later on replaced by a DIG based system (see section 7.10.3.3). Because DIG is originally derived from plants and not found in mammals, DIG-labeling yielded much fewer false positives PLA signals (see section 5.3.2.8).

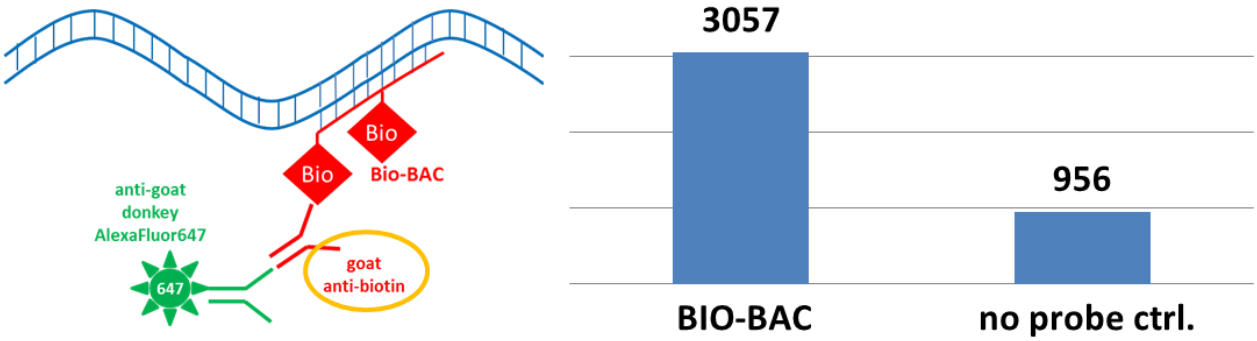


Figure 65: Specificity of the biotin antibody under PLA conditions. Left: *Experimental setup. Biotinylated BAC is hybridized to chromatin according to PLA protocol, targeted by the primary anti-biotin Ab (Biozol SP3000) and detected by an AF647-conjugated secondary Ab (Life technologies A21447).* ***Right:*** *IF-ISH signals per mm² on 5 µm E14.5 paraffin sections, imaged by widefield fluorescence microscopy. After BIO-BAC ISH 3 times more signals were detected than with a mock ISH. Signal detection and evaluation was performed as in* ***Fig. 62B****.*

5.3.2.5 Optimization of PLA imaging

With all antibodies and the ISH established, full PLA was performed on H3K27Ac and BIO-BAC to optimize imaging as well as further PLA protocol details. As previously reported for PLA, circular signals with 400-1000 nm diameter were obtained. Observed signal size and intensity depended on polymerase incubation times, but the signal was binary (either strong or absent) in all cases. I preferred PLA RED over PLA FAR RED detection reagents, as the fluorescent signal was stronger, especially in comparison with background, and hence substantially reduced imaging time. Highest resolution of PLA-signals was achieved in 3D by confocal microscopy. However, as the signals were big in size and virtually never overlapping in 3D, sufficient resolution and sensitivity was also obtained by 3D widefield fluorescent microscopy (see ***Fig. 66***). Without losing biological information this method provided much faster imaging and hence allowed

by tile-stitching the detection of complete embryonic and wide adult intestinal tissue sections instead of small, handpicked and thus possibly biased, frames.

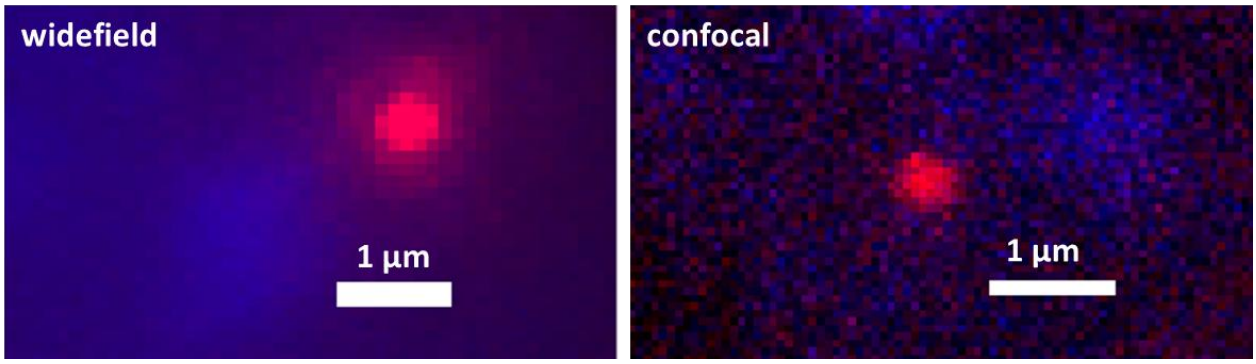


Figure 66: Comparison of PLA imaging techniques. *Maximum intensity Z-stack overlays, DAPI (blue) and PLA-signal (RED detection reagents).* ***Left:*** *Widefield fluorescent microscopy (AF7000 microscope), 63x oil immersion objective, 102x102x200 nm voxel.* ***Right:*** *Confocal imaging (SP5 microscope), 63x oil immersion objective, 60x60x168 nm voxel, 2x line averaged DAPI channel, 3x line averaged PLA channel.*

5.3.2.6 Protein-protein PLA

The predominant application for PLA is the visualization of protein-protein interactions, only few exceptions for DNA-protein and RNA-protein interactions were published so far[151-153]. While determining the best material preparation protocol, I also confirmed the efficient and easy applicable protein-protein proximity detection by PLA on paraffin- and cryosections (see ***Fig. 67***). In both experiments numerous signals were observed: 13.700 histone-PLA signals / mm² on paraffin sections (average data from 3 replicates, including the one shown in ***Fig. 67A***) against 20 signals / mm² on the corresponding negative control and 60,000 vs. 4,000 signals / mm² for CRE-PLA on cryosections. These protocols could probably be even further improved quantitatively. However, cryosections are unsuitable for ISH-based DNA-protein PLA, as it requires PK treatment which completely disintegrated the loosely slide-attached tissue on cryosections. Hence, only paraffin sections were further used.

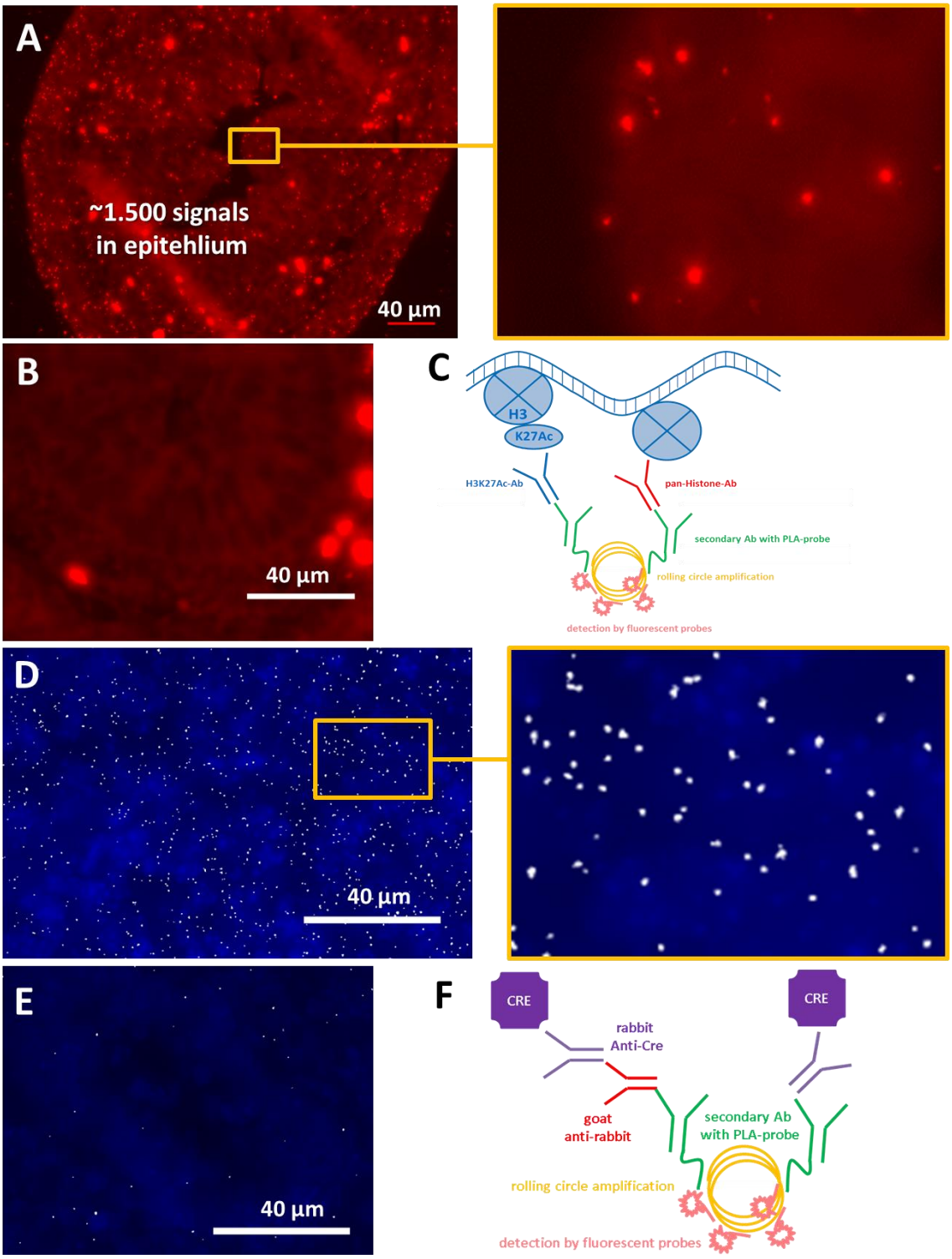


Figure 67: Protein-protein PLA. *PLA-reagents were tested by protein-protein PLA on paraffin- (**A,B**) and cryosections (**D,E**). Imaging was performed with PLA RED detection reagents by widefield fluorescent microscopy. **A:** Histone-histone PLA with H3K27Ac and pan-histone antibodies, as depicted in scheme **C**.1500 PLA-signals were detected in the gut epithelium. **B:** Negative control without primary antibodies. **D:** CRE-CRE PLA as depicted in scheme **F**, PLA signal shown as binary mask (white). PLA MINUS probe targeted the primary CRE-Ab directly, while the PLA-PLUS probe bound to a secondary Ab which in turn targeted the primary CRE-Ab. Cryosections from ubiquitously expressing Hprt-Cre mice were used. **E:** Negative control without primary antibodies.*

5.3.2.7 Semiautomatic PLA-signal evaluation

To evaluate large tissue section areas in a fast and unbiased fashion, I developed a PLA signal detection algorithm. It allows recognizing PLA signals with very high efficiency and specificity on autofluorescent background with the possibility of subsequent automated signal counting. Many widely used signal detection algorithms either just mark high intensity particles, such as the Duolink Image Tool software (Sigma Aldrich, cat no. DUO90806) or, as e.g. the described in section 7.12.5.3 Gaussian blur algorithm, calculate the background level from surrounding pixels, which comes either with a low detection rate of the comparably weak PLA signals or with a high level of false positive pixels in autofluorescent areas. In opposite to this, my algorithm employs a second, background-imaging channel with ~100 nm shorter wavelengths of excitation and detection (see ***Fig. 68***). The wavelength shift is strong enough to substantially reduce fluorescent signal detection from the fluorescent PLA detection probes, and at the same time sufficiently small not to introduce changes in the autofluorescence spectrum of the tissue. Division of both channels eliminates weak and strong autofluorescent background equally well, which is important in imaging of heterogeneous tissues. Only pixels contributing to PLA signals retain high values after background division and are thresholded to a binary (black/white) mask that can in turn be automatically evaluated for signal geometry and number. The developed algorithm detects over 95 % of PLA signals automatically and shows virtually no false positives.

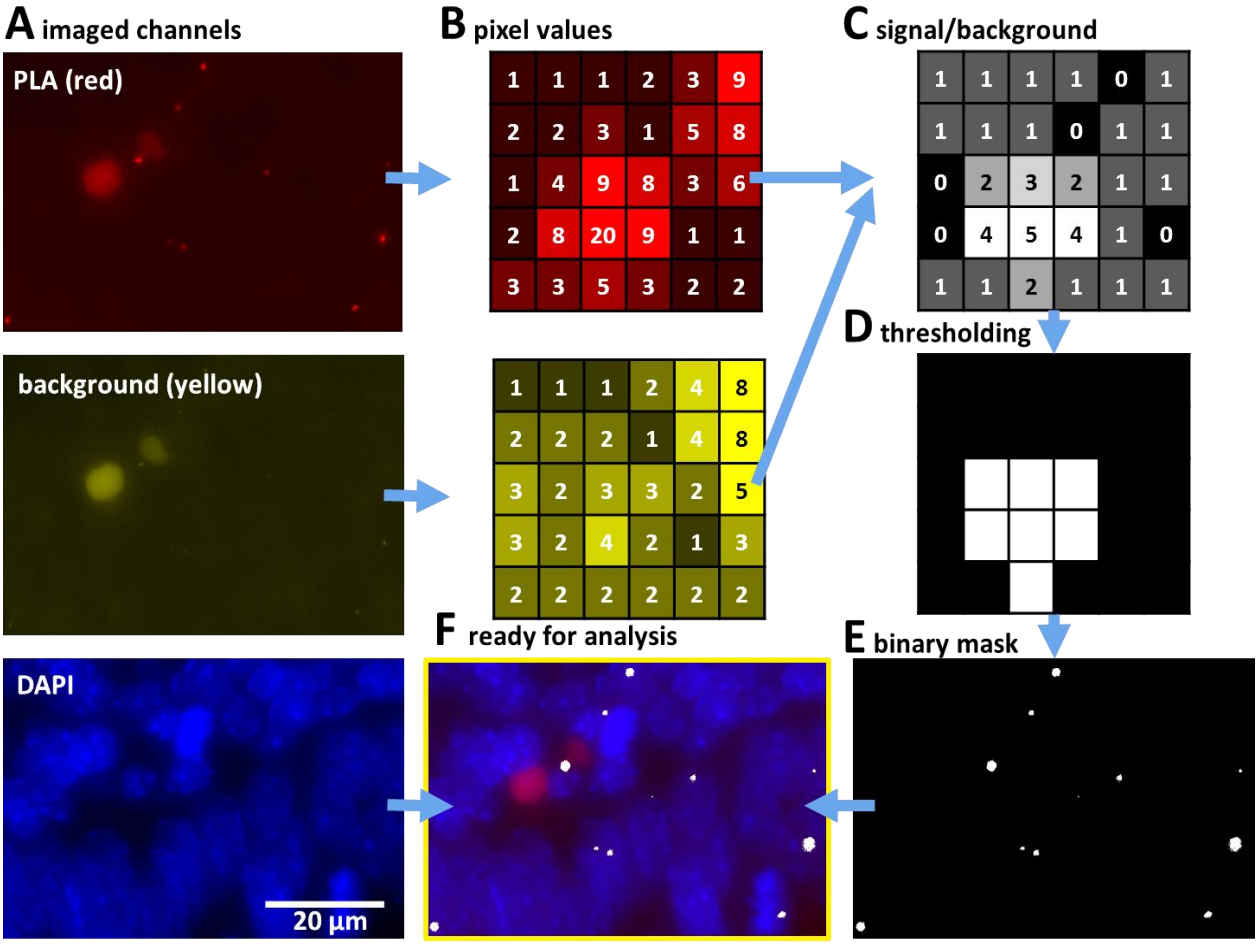


Figure 68: PLA signal calculation. ***A:****Widefield fluorescent imaging of the PLA channel (PLA RED detection reagents, a shorter wavelength background and nuclei, counterstained by DAPI. E14.5 paraffin sections. Yellow/red elevated intensity areas are highly autofluorescent blood cells at the border of mesenchyme (top) and intestinal epithelium (bottom).* ***B:*** *Schematic representation of imaged channels. Numbers represent pixels intensities in each channel. PLA signals only yield strong pixel intensities in the red channel (center), while autofluorescent background is of similar intensity in both channels (top right, edges).* ***C:*** *Each pixel's value from the PLA channel is divided by the background value.* ***D:*** *Thresholding ≥2 (value can be adjusted depending on absolute channel values) sets all background pixels "0" and PLA signal" 1", producing a binary PLA signal mask(****E****).* ***F:*** *overlay of PLA channel (red), binary PLA mask and the DAPI channel allows localization analysis and statistical evaluation of signal distribution.*

5.3.2.8 TSS – H3K27Acetyl PLA

With all single components confirmed for compatibility with the PLA protocol, DIG-labeled TSS ISH probes were tested to assess PLA overall sensitivity and specificity. The DIG-labeled *Shh* TSS probe (see ***Fig. 69A***) showed hereby best results and was further used as a positive control.

On E14.5 tissue, ~2300 signals / mm² were detected with the DIG-*Shh* probe (see ***Fig. 69B***), which corresponds to 30 % PLA efficiency if assuming that all *Shh* loci are marked by H3K27Ac at this developmental stage (compare section 5.3.2.4). Signals were detected up to 3-4 µm into the tissue section, which suggests that ISH is the restrictive factor, as same limitations for tissue penetration apply by this

protocol for FISH. The no probe control showed less than 2 % of this signal density (see ***Fig. 69C***). With a 50:1 signal to noise ratio, specificity of DIG-targeting PLA was clearly superior to the biotinylation approach (compare with ***Fig. 65***). Controls without any one of the used primary or secondary PLA antibodies showed virtually never any PLA signals in the imaged sections.

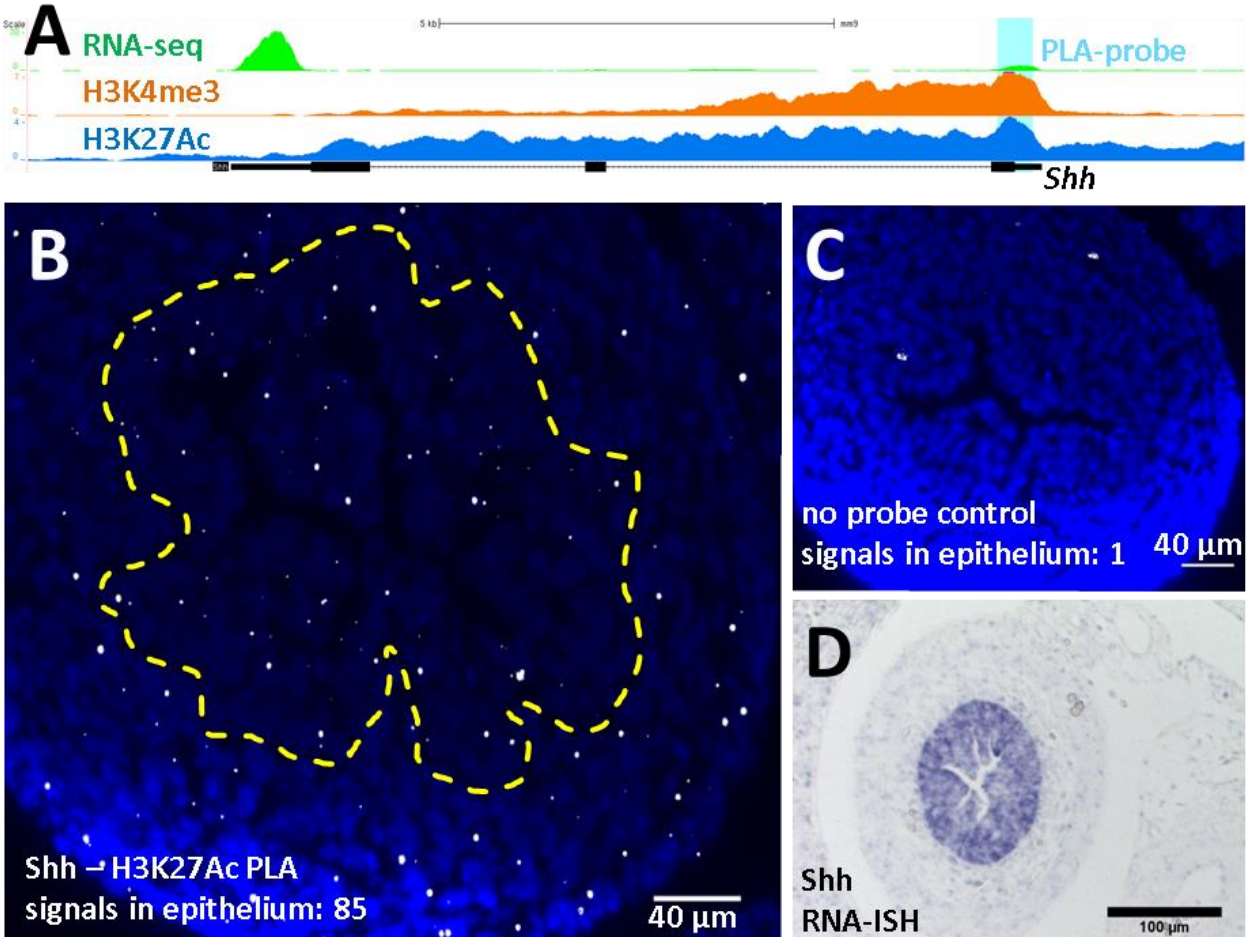


Figure 69: Shh TSS – H3K27Ac PLA. A: *Probe localization of the DIG-Shh probe (light blue) under the active H3K4me3 and H3K27Ac ChIP-seq peaks. Scale bar 5kb.* ***B,C:*** *DAPI Z-stack maximum overlay with binary PLA signal mask. The border between epithelial and mesenchymal tissue is indicated by the dashed line. E14.5 paraffin sections.* ***B:*** *PLA against H3K27Ac and the DIG-Shh probe.* ***C:*** *neg. control without DIG-Shh probe.* ***D:*** *RNA-in situ hybridization staining of d14.5 embryonic gut.*

Surprisingly, also the mesenchyme showed high PLA signal levels for the *Shh*-TSS probe. Although this was unexpected according to RNA-*in situ* hybridization (RNA-ISH) analysis of gene expression (see ***Fig. 69D***), RNA-seq data to detect low level expression or ChIP-seq to quantify histone mark deposition at *Shh* TSS is currently not available for the mesenchyme. Therefore it is possible that the PLA signal is caused by H3K27Ac positioning at mesenchymal promoters, which does not, or only at low levels lead to gene expression. *Epcam*-TSS – H3K27Ac PLA on the other hand showed as expected mostly epithelial signals, although much lower in number, only 260 signals / mm² and hence only a 6:1 signal to noise ratio in comparison to the no-probe control.

In a parallel experiment, DIG-labeled probes showed slightly better signal density and a 10-fold better signal to noise ratio in comparison to biotin-labeled probes. This was expected, as unlike biotin, there is no intrinsic digoxigenin in mammalian cells to give rise to false positive signals.

Alternative TSS showed differential PLA signal distribution as shown for the short and long TSS probes of *Arid5b* in ***Fig. 70***. Interestingly, the less intensely H3K27Ac marked short TSS yielded more PLA signal in the intestinal epithelium in comparison to the long TSS site (*P=0.13*, unpaired t-test, n=3). A possible explanation could be different ISH efficiencies of the probes, e.g. due to their different upstream/downstream orientation to the corresponding TSS and hence interference with chromatin bound factors. Another reason could be the 3D proximity of the short TSS to H3K27Ac rich chromatin domains, such as enhancers.

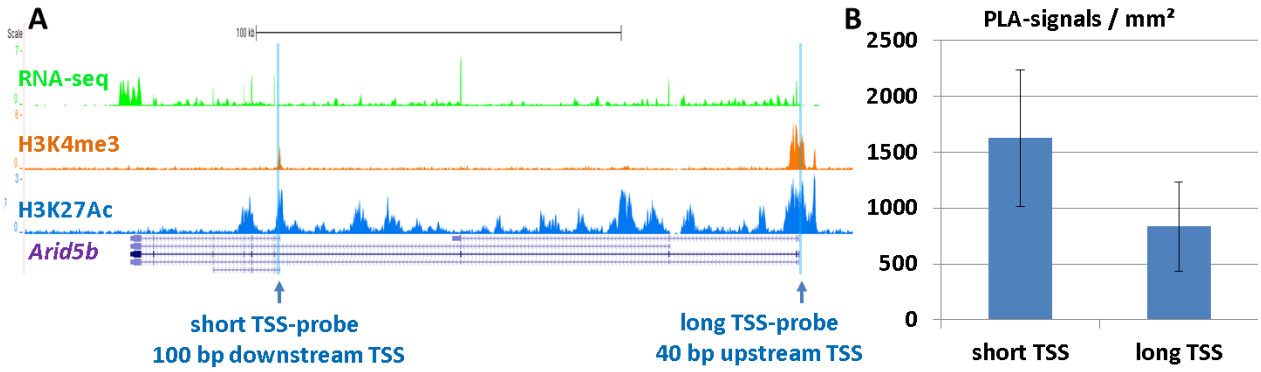


Figure 70: Alternative TSS – H3K27Ac PLA. A: *241-245 bp long biotinylated PLA probes were prepared to target the indicated alternative TSS sites of Arid5b. Scale bar 100 kb.* ***B:*** *PLA signals / mm² of intestinal epithelium. Error bars show standard deviation of 3 imaged intestinal E14.5 paraffin sections.*

While the first results were very promising, the complete protocol only worked on a small subset of embryonic tissue batches and on these with strong quantitative variations. Therefore, each of the PLA components and each of the protocol parts were successfully tested separately on different tissue batches. The performed tests determined tissue sample preparation as the weak spot of DNA-protein PLA. Unfortunately, also extensive troubleshooting of tissue preparation, e.g. variations in fixation intensity and proteinase K treatment did not provide any substantial improvement of reproducibility.

The shown above results underline the potential of PLA in understanding epigenetic mark positioning and TSS activity patterns in complex tissues. Unfortunately, problems with quantitative reproducibility could not be sufficiently resolved. Although positive results could be obtained repeatedly, varying signal densities and signal to noise ratios made statistical analysis impossible, which ultimately led to the termination of this project.

5.4 Promoter-enhancer PLA

A second project employing PLA was the visualization of promoter-enhancer interactions *in situ*. It was performed by Berith Messner as her MSc project under my supervision. Detailed methods and results are described in *"In situ analysis of chromatin mark localization & enhancer activity using proximity ligation assay (PLA)"*, Berith Messner 2015, M.Sc. thesis in anthropology, J.G. University of Mainz.

The limb specific MFCS1 (see section 5.1), the lung/gut specific MACS1[135,154] and the lung/gut specific SLGE[154] enhancers were analyzed for proximal interactions with the promoter region of *Shh* by DNA-DNA PLA. As described above, *Shh* is a crucial player in embryonic development in vertebrates and its enhancers are known key regulators of tissue specification. Understanding the distribution of enhancer-promoter interactions along the anterior-posterior axis and their correlation to cell fate commitment would help to unravel their role in gut development.

For this project, embryonic intestinal section were used, so that the limb bud specific MFCS1 acted as a negative control, while colocalization of the both other enhancers with the *Shh* TSS was expected, analogous to previously published 3C-data[135]. As the 3 tested enhancers are 100 kb – 1 Mb upstream of the *Shh* TSS, very little unspecific colocalization was expected.

Biotinylated ISH probes were prepared for each enhancer and DIG-labeled probes for the promoter region. As the exact interaction loci in the enhancer and promoter sequence are unknown, the whole stretch of each element was covered by several non-overlapping probes. After ISH, PLA was performed according to usual protocol with primary anti-DIG and anti-BIO antibodies.

While DNA-protein PLA produced positive but not quantitatively reproducible results, DNA-DNA PLA lacked positives results completely. As high ISH efficiency was confirmed on the used material and other PLA components were also working properly, the low number of enhancer-promoter PLA signals was probably due to only short-lived enhancer-promoter interactions, insufficient to be captured often enough by PLA.

6. Discussion

6.1 Genome-wide epigenetic patterns during small intestinal development

6.1.1 Novel timecourse approach

Previous epigenetic studies of the small intestine concentrated only on short term developments, restricted to the adult differentiation from ISC to AE[32,87,155]. However, many developmentally important epigenetic processes only take place at embryonic developmental stages or during the embryonic-adult transition and hence cannot be studied in an adult-only model. Moreover, the exciting, early part of intestinal crypts and villi formation as well as embryonic cell fate commitment with the underlying epigenetic mechanisms are completely missed. In this thesis I employed a comprehensive approach, which includes a wide set of active and repressing epigenetic marks followed through 4 stages, from the early embryo to adult tissue.

The combination of MBD-seq, H3K4me3, H3K27Ac, H3K27me3 and H2A.Z ChIP-seq with RNA-seq, acquired at each developmental stage, is a unique dataset, which not only allows to pinpoint the distribution of epigenetic marks at a single time point, but also to elucidate their long-term interactions and their impact on gene expression.

To obtain high quality ChIP-seq data for developmental timecourse analysis and semi-quantitative correlation to gene expression, I performed a set of optimizations on cell isolation, FACS and ChIP protocols. The strongest effect on ChIP-seq quality had the introduction of the HDAC-inhibitor NaButyrate, which prevented *in vitro* deacetylation of H3K27 during single cell isolation, FACS and ChIP. This could strongly improve H3K27Ac ChIP-seq quality in comparison to the HDAC-inhibitor free protocol, used in previous studies[84]. Hereby the sensitivity was increased approx. 10-fold without any aberrant acetylation effects or any detectable change in specificity.

The only acquired ChIP-seq dataset not meeting high quality standards is H3K27me3. This ChIP-seq experiment was previously performed by my colleagues and I additionally repeated it several times after extensive protocol troubleshooting. However, also the improved conditions could not lift the data quality to the level of the other 3 histone mark ChIP-seq experiments. While H3K27me3 peaks are clearly visible in the genome browser on at least two replicates per stage at known positive regions, e.g. Hox-clusters are strongly marked (see ***Fig. 20***), peak calling algorithms failed so far in most replicates. The missing, or where available unreplicated peak annotations strongly impair genome-wide comparison of H3K27me3 to other epigenetic marks and gene expression, which ultimately makes this mark's timecourse assessment impossible. The best acquired replicates, with the exception of the superior E12.5_3 replicate, are comparable to the best, currently published ChIP-seq tracks[156]. This indicates that the main reason for lower quality of H3K27me3 ChIP-seq data are the antibodies available, which are unspecific in comparison to other histone marks.

6.1.2 Genome-wide histone mark distribution and colocalization

H3K27Ac, H3K27me3, H3K4me3 and H2A.Z were annotated to genomic elements for the developmental stages E12.5, E14.5, ISC and AE. The high TSS-occupation levels, observed for all 4 marks, fully match previous reports(see ***Fig. 22***). As expected, with >75% the highest fraction of TSS ChIP-seq peaks, compared to intronic, exonic and intergenic ones, was observed for H3K4me3, which is a known active/poised promoter mark[92,96]. A lower fraction of TSS-associated peaks was called for H3K27Ac (>25%) and H2A.Z (>50%), as these marks are found not only at active promoters but also at active regulatory elements[64,83,87,92]. H3K27me3 showed with ~25% the lowest association with promoter regions, which is based on the very broad H3K27me3 signal distribution, far into the gene body as well as several genes spanning cluster positioning, as e.g. on Hox-clusters (see ***Fig. 20***). No substantial changes in positioning preference were observed throughout the four analyzed stages.

As expected, H3K4me3 and H3K27Ac colocalized on promoter regions and generally correlated to increasing gene expression levels(see ***Fig. 24 and 25***). On the other hand H2A.Z was found also on weakly expressed genes with a peak at moderately expressed genes and completely absent from the top active promoters, where H3K4me3 and H3K27Ac peak annotations were most abundant(see ***Fig. 25***). As expected, all 3 active marks strongly anti-correlated to DNA-methylation, both on active and silenced gene promoters[35,50]. Interestingly, while most active TSS are marked with all 3active histone marks, H3K27Ac seems in many cases to be sufficient for gene activation without H3K4me3 and H2A.Z, while only few genes show activity without H3K27Ac (see section 4.3.5.1). This pattern, observed at all 4 developmental stages, indicates a dominant functional role of promoter H3K27Ac in gene activation.

In addition to genetic element annotation mentioned above, H3K27Ac, H3K4me3 and H2A.Z were annotated to repetitive elements(see ***Fig. 28***). Strikingly, H3K4me3 showed strongly increasing LTR annotation towards the adult stages ISC and AE, which coincided with TE activation. Usually, TE are silenced by several mechanisms, including DNA-methylation at early embryonic stages, however some elements stay active or get derepressed at later stages[157,158]. On visual ChIP-seq track analysis, TE were observed, which completely loose DNA- methylation until ISC stage and gain strong H3K4me3 and H3K27Ac marks instead, matching peaking transcription levels at this stage. Thereby, this novel functional annotation of H3K4me3 in combination with H3K27Ac to active TE could potentially be used to understand the regulation and predict the activity of these repetitive elements. Nevertheless, this observed correlation remains to be tested quantitatively and the here presented ChIP-seq data is not optimal to do so, as many repetitive elements, such as LTRs, cannot be uniquely mapped. Because of this, only moderately old repeats, which already accumulated sufficient mutations to be uniquely mapped, are detected by histone ChIP-seq. Freshly integrated, and presumably most active elements are missed, while increasingly old repetitive elements fail to recruit active marks and are completely silent due to accumulated mutations. The positioning

of H3K4me3 and H3K27Ac on active TE could instead be tested by single locus approaches, such as PLA or by ChIP-seq methods and mapping, which are more suitable for repetitive elements, such as e.g. paired end sequencing of longer fragments.

As expected from previously reported H2A.Z and H2K27Ac positioning on active promoters and enhancers[66,83,87], both marks showed strong peak colocalization with each other and exclusion of DNA-methylation on TSS and intergenic regions (see ***Fig. 26***). In addition to active enhancers, also poised enhancers of not yet active genes already displayed moderate H3K27Ac levels, which were not reported previously[84,85] (see ***Fig. 27***). As discussed above, the crucial improvement to previous studies, where only the strong peaks on active enhancers were detected, is the usage of the HDAC-inhibitor NaButyrate during ChIP, which increased its sensitivity. Premarking with H3K27Ac was not observed at promoter regions, which hence could be used as a specific epigenetic identifier of poised enhancers regions. The combined histone ChIP-seq, MBD-seq and RNA-seq timecourse data clearly shows, that poised enhancer marking precedes not only the correspondent gene expression but also positive TSS marking with H3K4me3 and H3K27Ac as well as the depletion of DNA-methylation.

In addition to this early onset effect of low level enhancer H3K27Ac, the acquired timecourse dataset suggests even stronger impact of epigenetic enhancer states on gene activity during development and differentiation. As described in section 4.3.8, the full set of active promoter marks, including H3K4me3, H3K27Ac, H2A.Z and the absence of DNA-methylation, can be completely negated by a change in epigenetic marks at the *cis*-regulatory region. The shown example *Olfm4* and many more observed throughout the genome underline the necessity not only to focus on promoter deposition of epigenetic marks but their genome-wide distribution. Unlike promoter regions, where detailed functional and genome-wide computational studies were performed in the past years, epigenetic marks' functions at enhancers are still to be elucidated in detail. To achieve this, the annotation of *cis*-regulatory regions has to be strongly improved regarding prediction algorithms as well as functional screens. It remains a priority in epigenetic research to link each potential *cis*-regulatory region to its target gene(s), which gives the only possibility to assess enhancer activity *in vivo*. The described novel poised enhancer pattern might contribute to the identification and annotation of developmentally active enhancers.

6.1.3 Epigenetic changes during intestinal development

Comparison of epigenetic marks and gene expression at single time points has only limited potential to assess long-term dynamics during development. To shed light on these processes, I applied timecourse clustering of RNA-seq, H3K27Ac, H3K4me3 and H2A.Z ChIP-seq. These clusters grouped genes by their decreasing/increasing pattern of gene expression or histone marks along their differentiation from E12.5 to adult stages.

Clustering the 2-fold changes in gene expression correctly annotated well known and revealed novel markers of embryonic stages, stemness, proliferativity and differentiation (see section 4.3.9.1). Gene to RNA-seq cluster annotations were successfully validated by experimental results targeting the expression of single genes, such as RNA-ISH and RT-qPCR.

Differential histone mark positioning at TSS during intestinal development was successfully annotated to the same cluster set as gene expression (see section 4.3.9.2). Hereby 1.5-fold changes in the ChIP to input enrichment were applied as clustering threshold. Histone marks' ChIP-seq signals, in contrast to gene expression, have a limited quantitative range, from absent to fully labeled loci. Because of that and the lower quantitative and spatial resolution of ChIP-seq in comparison to RNA-seq, time course clustering of both differs in two major points: Firstly, a lower fold-change cutoff had to be used for ChIP-seq clustering to increase detection of signal intensity changes. However, to avoid an increase in clustering noise, the cutoff however could not be reduced any further. Secondly, as expected from the greater signal intensity range, gene expression generally clusters more frequently to the changing clusters in comparison to the all-equal clusters "14" or "E" (see ***Fig. 31 and 32***). Moreover, histone marks are underrepresented in clusters with two or more unidirectional changes, as an already completely marked/empty locus cannot gain/loose signal intensity.

The combined dataset of epigenetic and gene expression timecourses allowed narrow screening for genes either involved in developmental regulation or showing specific epigenetic changes, which could be used to elucidate their function. These genes were then assessed in detail by a variety of single locus approaches, a subset of which is described in this thesis.

Most numerous epigenetic and gene expression changes between adjacent developmental stages were observed between E14.5 and ISC, while the least were observed between E12.5 and E14.5. This matches the time distance between the stages, which is the shortest between the latter. Also the transition between ISC and AE happens in only a few days[99]. Here however, major differentiational changes take place, which goes with changed extracellular signaling and stronger variation in epigenetic and gene expression levels compared to E12.5 vs. 14.5. Therefore, the number of changes between ISC and AE is closer to the embryonic-adult transition E14.5 vs. ISC.

6.1.4 Developmental interactions of histone marks and gene expression

To address combinatorial effects of histone marks on differential gene expression, I performed co-clustering analysis. Comparing the gene annotated clusters I looked for epigenetic mark and gene expression interactions that span the whole intestinal development, from embryos at stage E14.5 over ISC to fully differentiated AE, rather than just coinciding at a single time point.

As discussed above, cluster annotation frequencies varied in between histone marks and especially between histone marks and gene expression (see sections 6.1.3, 4.3.9.3). Therefore, co-clustering specificity was determined for each two traits separately and with statistical correction for the expected clustering frequencies.

The active histone marks H3K27Ac, H3K4me3 and H2A.Z generally co-clustered significantly in all 3 pairwise combinations. This means, that all three histone marks not only colocalize at any given time point (see section 6.1.2), but also generally follow each other's differential pattern during small intestinal development. H3K27Ac and H3K4me3 show hereby the strongest clustering correlation: These marks' equal cluster annotation is highly enriched in comparison to different cluster annotation ($P=7*10^{-234}$). Also for each single cluster these two marks correlate significantly ($P<0.05$) (see ***Fig. 34***). For H2A.Z the overall pairwise parallel clustering with other two active marks is also very clear but less pronounced ($P=1*10^{-84}$ vs. H3K27Ac and $P=8*10^{-100}$ vs. H3K4me3). Hereby, several single clusters are not significantly co-clustered ($P>0.05$) due to their very low clustering frequencies, while all larger clusters show significant enrichment. Also, H2A.Z is expressed at lower levels in AE (see section 4.4.3), which indicates reduced involvement of this mark in epigenetic regulatory processes and probably contributes to the H2A.Z cluster decoupling from the other active marks at this, fully differentiated state.

Gene expression also showed strong co-clustering with H3K27Ac ($P=5*10^{-25}$) and H3K4me3 ($P=8*10^{-13}$). The P-values, increased in comparison to the histone mark only co-clustering, are partially due to higher variation of gene expression levels, which reduces the correlation specificity to histone mark clusters. Furthermore, they also clearly indicate the well-studied involvement of non-epigenetic regulatory mechanisms (see section 2.1). Still, as it would be expected from active promoter marks, the significant cluster correlations clearly show, that H3K27Ac and H3K4me3 are not only likely to be found on active promoters at any given time point, but change in parallel to differential gene expression throughout the assessed stages of gut development.

For H2A.Z however, co-clustering analysis revealed a strikingly different connection to promoter activity. At any single time point, H2A.Z was found on active TSS, usually in combination with H3K27Ac and H3K4me3 (see section 6.1.2). However, I observed no significant link to differential gene expression. Although expected from its co-clustering with H3K27Ac and H3K4me3, two marks that clearly follow gene

expression, no single cluster showed significant co-clustering enrichment (*P>0.05*) between H2A.Z and RNA-seq clusters. Moreover, also the combined equal cluster set was not significantly enriched (*P=0.98*). This means, that although H2A.Z tends to localize on active promoters, its increasing/decreasing positioning intensity does not correlate to increasing/decreasing promoter activity.

Taken together, timecourse clustering results show, that although H2A.Z correlates with active TSS at any given time point, it is, in contrast to H3K27Ac and H3K4me3, not a marker of time resolved gene expression. H2A.Z is therefore not an active TSS mark, but rather positions on responsive TSS, allowing both their subsequent activation and repression. Moreover, once the transcription state of the gene is established, e.g. by full repression or stable expression in a finally differentiated cell, H2A.Z can be depleted and no longer correlates to gene activity. ***Fig. 25*** illustrates this depletion of H2A.Z from the most active as well as from the repressed gene subsets. The characterization of H2A.Z as an responsive rather than active mark is supported by its well-studied function in chromatin destabilization and increased accessibility for its remodelers[66-68]. H2A.Z was also specifically linked to responsive genes in plants, i.e. easily activated/repressed, e.g. stress-inducible genes[48]. In this study however, responsive gene H2A.Z was not found on promoters but widely distributed on gene bodies instead, a pattern I did not regularly observe on responsive murine genes, so that this localization could be a plant-specific feature of H2A.Z.

6.1.5 Tissue heterogeneity of the small intestine

Sequencing analysis, such as RNA-seq, ChIP-seq and MBD-seq, still requires high cell numbers, as single cell sequencing does not work for every application and generally has a lower coverage rate in comparison to bulk-sequencing[159-161]. Especially ChIP-seq experiments, which strongly depend on Ab specificity and their precipitation efficiency, fail to produce high coverage data by single cell sequencing[161]. However, when performing experiments on cell populations, here I used up to 500,000 cells per replicate, it should be kept in mind that some observations might not be based on even contribution by all analyzed cells. The observed parameters might in some cases rather represent average values, only contributed by distinct cellular subpopulations.

During my studies of developmental epigenetic patterns on genome-wide and single locus scale I observed numerous heterogeneities in cellular populations that were previously believed to be uniform (see section 4.4 for details). Already at embryonic stages gene expression patterns of developmental genes, such as *Lgr5* and *Axin2*, resemble adult distribution. Corresponding to adult crypt base localization, these genes are expressed in the polar embryonic intestinal tube regions. As these polar region cells contribute only a small fraction to the analyzed embryonic E12.5 and E14.5 populations, their gene expression levels, probably as

much as epigenetic states, are underestimated in RNA-seq and ChIP-seq data. However, especially these cells might have the strongest impact on later gut development, as they acquire ISC properties.

In the adult intestinal epithelium ISC were already previously reported to consist of two cell populations: proliferative crypt bottom and quiescent +4 cells[107,111,162,163]. I also observed these two distinct populations by differential gene expression of *Lgr5* and *Axin2*, both stemness markers. In contrast to *Lgr5*, which localizes, as previously reported, at the crypt bottom, decreasing in expression towards the crypt mound, *Axin2* is only expressed starting from the quiescent stem cell position +4. Interestingly, *Axin2* also showed additional heterogenic expression in the villus but only in the posterior small intestine. This is just one of the observed examples for heterogenic gene expression along the anterior-posterior axis, more of which are shown in section 4.4.3 or were reported previously[123]. As discussed for embryonic polar heterogeneity, also effects observed on sequencing datasets of ISC and AE, harvested from the whole intestinal tissue, might be contributed only by cells from distinct anterior-posterior regions.

While epigenetic pattern and gene expression changes from the proliferative crypts to the fully differentiated villus compartment are well studied, intravillus variation received so far little attention[164,165]. Both on RNA and protein levels I observed substantial changes between AE at the villus base and tip. As shown in section 4.4.3, a subset of genes is only expressed shortly after differentiation in the basal villus and not at all in the fully maturated AE at the villus tip. This is especially striking in the case of H2A.Z: The H2A.Z encoding gene *H2afz* shows reduced expression in the overall AE population by RNA-seq, however, if spatially resolved by IF staining, this histone variant can only be detected at the villus base. Hence, this gene must be only expressed in the lowest villus compartment, while completely shut down and the protein degraded during AE maturation. As discussed in section 6.1.4, this fits to the postulated functionality of H2A.Z, which allows responsive regulation of marked genes, but which becomes no longer required in the terminally maturated and soon discarded AE at the villus tip. Regarding the here presented results, it is therefore clear that, unlike for the H3K27Ac and H3K4me3 ChIP-seq AE replicates, H2A.Z ChIP-seq AE data represents only the Epcam+ basal villus cells.

All of the observed differences along the anterior-posterior crypt-villus axis of the embryonic and adult small intestine make sense in the light of embryonic development and adult differentiation. The embryonic development of specialized tissues of the gastrointestinal tract is based on gene expression gradients and some of those persist and develop until adulthood in accordance to tissue specifications[127,128]. Also adult cells are strongly affected by extracellular signaling, which, as reviewed in section 2.4.2, is dependent on the exact positioning along the crypt-villus axis. Therefore, the differentiation from ISC over TA cells to AE should not be considered as step wise but rather as a continuous process, not finishing with the last cellular division, but ongoing until the maturated AE are shed at the villus tip into the gut lumen.

With this in mind, the used cell populations, although specifically selected for established markers and well suitable for a wide range of NGS-seq analysis, should not be considered homogenous in all aspects but rather additionally examined for possible subpopulations, solely contributing to the detected signal. As long NGS methods do not allow high coverage single cell sequencing, the assessment of rare cell populations has to rely on mostly imaging based techniques, such as RNA-ISH, IF stainings or the discussed below PLA.

6.1.6 Ongoing and future analysis

The now available developmental RNA-seq, MBD-seq and ChIP-seq datasets allow a wide range of additional bioinformatic analysis. Some of this, such as developmental alternative splicing analysis or the comparison of epigenetic development of AE and Paneth cells, is currently performed by our research group and collaborators.

In addition to this, developmental analysis of enhancer activity and the involved epigenetic marks could be conducted on the available dataset, as soon as a better enhancer to gene annotation will be available. A possibility to strongly improve enhancer annotation would be to extend the dataset by 4C or Hi-C data for all developmental stages[166].

Also, as discussed in the next section, the obtained genome-wide developmental datasets laid the ground for a variety of functional, single locus and single cell based projects.

6.2 Epigenetic single locus analysis

Genome-wide analysis by ChIP-seq allows insights in the general distribution and colocalization of epigenetic marks, however it has limited potential to elucidate their functions and physical interactions. For this purpose, I pursued several single locus approaches: A functional *in vivo* study of the H3K4me1 enhancer mark on a limb specific enhancer of the *Shh* gene, an analysis of the RNA-Polymerase II interactome changes along the functional structure of single genes, and a single cell resolved *in situ* analysis of epigenetic marks at alternative TSS by PLA, to assess intestinal tissue heterogeneity.

6.2.1 Functional analysis of the H3K4me1 enhancer mark

I performed this project to determine, whether H3K4me1 is a functional enhancer mark, required for full enhancer activity, or rather a histone-methylation "by-product", due to enhancer colocalization with promoter-bound histone methyltransferases (see section 5.1). To test these hypotheses, the histone demethylase Lsd1 should be specifically targeted by a fused TALE protein to the limb bud specific MFCS1 enhancer of *Shh*. A reduced *Shh* expression in the limb bud, easily observable by the embryonic skeletal

development, would strongly indicate functional impact of the H3K4me1 mark on enhancer activity, while an efficient removal of the mark without any developmental effects would support the contrary hypothesis.

The correctly assembled e*GFP-P2A-TALE-Lsd1* construct was used to generate 9 transgenic mice. However, only 4 of them were viable, fertile and showed germline integration of the *TALE-Lsd1* construct. From these 4 founder lines, a total of 184 embryos was analyzed for developmental effects of H3K4me1 depletion. Without exception, all embryos showed wild type limb development, indicating either non-functionality of the H3K4me1 mark, or the construct integration into silent, heterochromatic loci. Further molecular analysis of eGFP and TALE protein as well as mRNA levels confirmed complete silencing of the construct in 3 strains. In the 4th strain, construct expression was observed on mRNA level, but also here no TALE-Lsd1 protein or eGFP could be detected. As TALE-Lsd1 levels might be just below detection level and still sufficiently functional, or could just as well be too low or completely absent and hence insufficient to reduce MFCS1 histone methylation, the wild type phenotype observation were completely inconclusive in regard to H3K4me1 functionality.

To continue this project, a new, preferably larger set of transgenic animals needed to be generated. However, freshly published data[131] strongly suggested to terminate the project at this time point. Mendenhall et al. employed a transiently *TALE-Lsd1* transfected cancer cell line to demethylate in separate experiments 40 predicted enhancer regions. Although this study's results, due to the used model, are not completely transferable to the more complex mouse or human system, the methodical approach and the targeted questions were very similar to my project. In that study, high specificity of TALE-Lsd1 targeting and the efficiency of locus specific demethylation were confirmed by ChIP-seq. From the 40 putative enhancers, a large subset showed decreased histone methylation levels and 4 neighboring genes were found to have downregulated mRNA-levels on corresponding *TALE-Lsd1* transfection. However, they have also shown by H3K27ac-ChIP that Lsd1 has in addition to H3K4me1/2-demethylase also a strong histone-deacetylase activity, which might as well be the cause for gene downregulation. H3K27ac levels at the targeted enhancers decreased even stronger than H3K4me2. The authors additionally concluded that histone-deacetylation is not caused by histone demethylation, but is a direct effect of Lsd1. All taken together, functional studies involving Lsd1-posititioning on enhancers cannot discriminate a possible negative effect of enhancer histone-demethylation on gene activity from the expected parallel effect of histone deacetylation, as both marks are simultaneously found on active enhancers.

Overall, another, not Lsd1-based approach is needed to sort out the acetylation independent effects of H3K4methylation at enhancers, while TALE-Lsd1 is suitable to assess the overall gene-activating function of histone methylation and acetylation deposition on enhancers. In addition, TALE-Lsd1 could be used to identify target genes of putative enhancers.

6.2.2 Changes of the RNA-polymerase II interactome along the gene body

To compare the PolII interactome at different positions along a transcribed gene, I generated Flag-tagged TALE constructs, specifically binding to promoters, introns and exons of selected genes. After crosslinking and pulldown, these loci would then be analyzed for bound protein composition, including epigenetic marks and the PolII complex, by MS (see section 5.2). Of especial interest were epigenetic marks and transiently bound transcription factors along the gene body.

A set of TALE constructs was successfully generated and stably introduced to HeLa cells. The resulting cells lines, each with a single locus specific TALE, showed high and stable production of intact Flag-TALE proteins, which correctly localized to the cell nucleus.

After initial Flag-ChIP, yielding a specific but not yet optimally efficient enrichment of Flag-TALE proteins, MS was tested with the IMB Proteomics CF. Unfortunately, the Flag-ChIP output amount were by far not sufficient for MS. It was estimated, that >500 ng would be necessary as MS input to detect rare and transient PolII interactors. Taking the optimization of other performed ChIP experiments into account, I estimated that in best case a 10-fold improvement of Flag-ChIP efficiency could be achieved by changes in the ChIP protocol and the usage of clones which produce lower Flag-TALE levels. Reduced Flag-TALE amounts per cell would not only improve ChIP efficiency but also specificity, as the used Ab would not be titrated out by antigens unbound to chromatin. However, even in this best-case scenario of ChIP optimization, the expected precipitated protein amounts would still require hundreds of cell culture dishes per MS run. As numerous parallel experiments would be necessary to detect PolII interactome changes along a single gene, each of which would require the use of relatively expensive isotope labeled SILAC medium to quantitatively compare flag-ChIP output with input, the project was terminated at this stage.

Currently, the Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) method is developed by the research group of J. Krijgsveld, which strongly reduces the minimal input amounts for MS sample preparation[167]. SP3 avoids numerous quantitatively lossy transfer and isolation steps by performing the whole preparation from cell lysis to final purification in one tube. Furthermore, stable isotope labeling of peptides can also be performed *in vitro* during the SP3 protocol, which strongly reduces the costs of ChIP-MS experiments. With the substantial reduction in necessary flag-ChIP material and the *in vitro* isotope labeling this method provides instead of the cell culture based SILAC, the originally planned ChIP-MS experiments would become sufficiently cost-effective to consider resuming this project.

6.2.3 Proximity Ligation Assay: investigating heterogeneous TSS activity

To assess population heterogeneity in embryonic and adult intestinal epithelium, protein-DNA PLA was applied. As discussed in section 6.1.5, many epigenetic effects take place with a strong cell to cell variation and should therefore be investigated on the single-cell level. Here, I focused on asynchronous alternative TSS and their epigenetic states in the intestinal tissue (see section 5.3). While genome-wide ChIP-seq data shows the average distribution of epigenetic marks on alternative TSS in FACS-purified embryonic and adult populations, some of them might only be present in a subpopulation of cells.

The general PLA protocol, usually used to detect protein-protein interactions, was established and could reliably detect protein-protein proximity on paraffin- and cryosections. Also under the modified protein-DNA conditions, which include harsh PK treatment and ISH, protein-protein PLA yielded numerous and reproducible results.

To pinpoint epigenetic mark deposition to specific loci, specific ISH probes were generated to be targeted by PLA in parallel to the histone mark of interest. Hereby, DIG-labeling proved to be more specific than the alternative BIO-labeling of ISH probes, as unlike biotin, digoxigenin is not present in mammalian cells. Therefore, unspecific PLA-signal levels could be strongly reduced, allowing the detection of rare interactions (max. 2 per cell) of labeled probes with histone marks, reaching a signal to noise ratio of 50:1. Hereby, a detection rate of up to 30% of the theoretically possible histone mark to ISH-probe interactions was achieved (see section 5.3.2.8). For the detection of epigenetic marks, H3K27Ac Ab proved suitable, however, no PLA-compatible H3K4me3 or H3K4me2 Ab was found.

To allow unbiased statistical analysis of PLA data, I designed a semiautomatic signal recognition algorithm (see section 5.3.2.7). It allows a fast extraction of the weakly fluorescent PLA signal from strong and/or uneven autofluorescent background with very high sensitivity (>95%) and, most importantly, superior specificity (>99%). This is a strong improvement to other, also commercially available, PLA-signal detection algorithms, which either fail at highly autofluorescent tissues or only detect a low fraction of PLA-signal. Furthermore, it allows a fast and unbiased evaluation of large imaged areas, which are required for the statistical comparison of rare protein to single locus colocalization events. The signal detection algorithms works equally well for protein-protein and protein-DNA PLA.

Differential deposition of the active H3K27Ac mark was observed on TSS of several genes. Unfortunately, these results could not be quantitatively reproduced. Extensive troubleshooting was performed to optimize protein-DNA PLA performance and reproducibility. Taken all tests together, it seems that a very fine-tuned balance has to be achieved between chromatin accessibility for ISH by tissue permeabilization and digestion on the one hand and histone epitope preservation for primary Ab binding on the other. Smallest deviations, that cannot be avoided even under most controlled conditions, shift the balance and compromise the subsequent proximity detection. Incomplete tissue penetration might be another cause of PLA signal density deviations. As described in section 5.3.3.2, ISH and hence also PLA-signals are detected only until 3-4 µm into the tissue. Presumably, the upper layers are also the ones mostly affected by denaturing effects of section treatment for ISH as well as PK-digestion, which would leave less histone epitopes available for PLA. So possibly, there is only a narrow border area that could combine efficient ISH and epitope preservation for proximity detection.

The quantitative irreproducibility of DNA-protein PLA results might explain, why the potential of this approach is hardly reflected in scientific publications. Besides a study on muscle tissue, showing a few single PLA-signals, only cultured cells were assessed for DNA-protein proximity by PLA[151,153]. While Protein-protein and RNA-protein PLA visualize very numerous proximity events, interactions of single genomic loci with specific proteins are rare. That and the described need of combining ISH procedures with epitope-targeting seems to lead to an easily disturbed balance and hence low reproducibility.

While protein-DNA interactions could not be reliably investigated by PLA, DNA-DNA PLA might still work, as it is not relying on epitope preservation during ISH. In addition to genome-wide approaches of chromatin confirmation analysis, as e.g. the 4C and Hi-C techniques[166], DNA-DNA PLA could provide a single cell based view on it. Hereby however, probably only frequent DNA-DNA interactions could be imaged, as indicated by the enhancer-promoter PLA, described in section 5.4.

7. Methods

7.1 Animal experimental procedures

7.1.1 Mouse strains

The outbred CD1 wild type strain was used for most experimental procedures on wild type adult animals, experimental crosses to obtain embryonic tissue and for heterozygous breeding of mutant strains. The CD1 strain is maintained at the Translational Animal Research Center (TARC) and additional animals were ordered to IMB from Charles River Laboratories.

The inbred C57BL/6N wild type strain was used for breeding purposes. The CD1 strain was generally preferred due to its better overall health and a higher offspring. C57BL/6N mice are maintained at the TARC.

Lgr5-GFP and *Hprt-Cre* mice were imported from the Jackson Laboratory to the barrier division of the TARC, where this strain is currently maintained.

The *eGFP-TALE-Lsd1* mice were generated for the described above project by the Institute Clinique de Souris, France. All *TALE-Lsd1+* mice were bred and handled at the quarantine division of the TARC and at the IMB. This strain is no longer available, as after closing the project, all remaining *TALE-Lsd1+* mice were sacrificed.

7.1.2 Handling and crossing procedures

All animals were kept at the TARC and the IMB animal facilities under standardized conditions in accordance with the FELASA recommendations and the current German and European laws.

Adult animals were sacrificed by CO2, followed by cervical dislocation.

Where sufficient mice numbers were available, crosses were performed overnight with one male on two females, otherwise one on one. In case of experimental crosses for embryo extraction at a certain developmental stage, mating was assessed the next morning after crossing by a plug-check and the females were separated from the male until pregnancy could be confirmed or excluded. The day of the plug-check was counted as E0.5 and embryos were extracted at a corresponding later stage, e.g. twelve days later at E12.5.

7.1.3 Genotyping PCR/qPCR

7.1.3.1 Genotyping by PCR

Most mouse strains were genotyped by endpoint-PCR, followed by agarose gel electrophoresis. A biopsy of adult or embryonic mice was taken from the tail tip and digested overnight (ON) at 55 °C in 100 µl tail buffer supplemented with 300 µg/ml PK, followed by 95 °C inactivation. In some cases, an ethanol precipitation was additionally performed to increase sample concentration and purity (see section 7.13.2). The PCR-reaction was generally performed with self-made Taq-polymerase, in some cases with commercial Onetaq-polymerase.

PCR reaction composition (15 µl total):	Reaction Buffer (5x/10 x)	3 µl / 1.5 µl
	dNTP-Mix	0.3 µl
	Forward primer (100 µM)	0.1 µl
	Reverse primer (100 µM)	0.1 µl
	Taq-Polymerase	0.5 µl
	H_2O	9 µl / 10.5 µl
	Sample-DNA	2 µl

PCR-program:

94 °C – 3 min
94 °C – 30 sec
X °C – 30 sec x 35
72 °C – 30 sec
72 °C – 2 min
4 °C - ∞

All used primers are listed in section 8.3. The PCR output was further analyzed by agarose gel electrophoresis (see section 7.13.3).

7.1.3.2 TALE-Lsd1 transgene candidates

Biopsies of F0 transgene candidate animals were obtained from Institute Clinique de la Souris undissolved in Proteinase K buffer. After addition of Proteinase K (PK), the samples were incubated ON at 55 °C, inactivated at 95 °C for 5' and spun down at 21,000 g for 2'. After ethanol precipitation, qPCR was performed on the genomic *loci ndufaf3* (pos. ctrl.), *lsd1/TALE* and *GFP*.

7.2 Sorting epithelial cells by FACS

7.2.1 Sample collection

Embryos at stage E12.5 and E14.5 were opened for sample collection after visual confirmation of developmental stage. In the rare cases of misdeveloped embryos, these samples were not used. Small intestine was excised between stomach and caecum and pooled in 1ml 1x PBS.

Adult small intestine was separated from fat tissue and cut open longitudinally for several washes in 1x PBS. Further, the opened tube was cut in 4-6 mm long pieces and washed several times by gentle inversions before 4' incubation in 5mM EDTA in PBS, gently shaking at room temperature (RT). Villi were separated from the crypt-containing mesenchyme by pipetting the whole volume twice with a 10 ml pipet boy at maximum setting. The supernatant (SN) with villi was saved and the pipetting repeated twice more. The remaining tissue was gently washed and vigorously shaken to extract crypts from the mesenchymal pockets. The obtained SN with crypts was filtered through a 70 µm cell strainer. After 5' at RT in PBS + EDTA, crypts were extracted a second time and pooled. Efficient extraction was confirmed by tissue pieces starting to float on the surface. Crypts and villi were separately spun down at RT, 200 g for 4' and resuspended in 2 ml PBS.

Digestion of embryonic and adult samples was performed with 150 mg/ml collagenase at 37 °C in a shaking heating block at 1300 rpm. Embryonic samples were digested for 7'. For efficient single cell separation of Paneth cells, 50 µg/ml dispase was added. Samples were incubated 3' for villi and 7' for crypt digestion. To remove dead cells' and bacterial DNA from the intestinal lumen, 51 U/ml DNAseI were added to adult samples during digestion. Cells were spun down for 3' at 900 g, 4 °C and vigorously resuspended for single cell separation in RT PBS.

For H3K27Ac ChIP-seq, all PBS buffers were supplemented with 20 mM Na butyrate.

7.2.2 Single cell preparation for FACS

RT formaldehyde was added to single cell preparations at 1 % final concentration. Samples were fixed 8' at RT with shaking and 2' more during centrifugation at RT, 2400 g. Prior to Ab-staining, cells were washed 3x in PBS with centrifugation steps at 2400 g at RT.

Depending on the developmental stage and the targeted cell population, following antibodies were added and incubated 30' in 2% FBS in PBS at RT:

E12.5: 5 µl/ml Epcam (eBioscience, 46-579182)

E14.5: 5 µl/ml Epcam (eBioscience, 46-579182), 3.5 µl/ml CD45 (BD Pharmingen, 553081), 2.5 µl/ml CD31 (BD Pharmingen, 561073)

ISC: none, only expressed GFP is measured

AE: 10 µl/ml Epcam (eBioscience, 46-579182), 3.5 µl/ml CD45 (BD Pharmingen, 553081), 2.5 µl/ml CD31 (BD Pharmingen, 561073)

Paneth cells: 5 µl/ml Epcam (eBioscience, 46-579182), 3.5 µl/ml CD45 (BD Pharmingen, 553081), 2.5 µl/ml CD31 (BD Pharmingen, 561073), 5 µl/ml UEA1 (Sigma, L9006), 5 µl/ml CD24 (Sigma, #560536),

After washing 2x in PBS with 5 mM EDTA, cells were resuspended in PBS with 10 mM EDTA and kept at 4 °C until FACS.

For H3K27Ac ChIP-seq, all PBS buffers were supplemented with 20 mM Na butyrate.

Unfixed cells were prepared as fixed cells with following deviations: no formaldehyde fixation was performed, washing and Ab staining steps were shortened by 50 %.

7.2.4 FACS of epithelial cells

All epithelial cell sorts were performed on the FACS Aria (INST 247/645-1 FUGG) equipment with 85 µm nozzle. Before loading, samples were filtered through 40 µm tube cap filters and vortexed regularly during the sort to prevent clogging. Gating strategies are shown in detail in section 4.3.1.1. Single cells were sorted to precooled Eppendorf tubes with PBS, spun down at 2000 g for 3' and pellets saved at –80 °C.

In case of unfixed cell sorts, as e.g. for RNA-seq experiments, dead cells were excluded by DAPI staining applied shortly before the sort and detected with excitation laser 405 nm, detection filter 450/50 nm. RNA-seq material was directly sorted to RNA-extraction lysis buffer RLT and immediately processed.

7.3 Cell culture procedures

7.3.1 Culturing HeLa cells

The HeLa cell line was obtained from George Reid's research group at the IMB. Culturing was performed at 37 °C, 5 % CO_2, 85 % humidity in DMEM supplemented with 10% heat inactivated (30' at 56 °C) FBS

Gold and 200 mM glutamate. If not otherwise indicated, no antibiotics or nonessential amino acids were added. Harvesting and passaging of HeLa cells was performed by washing twice with PBS, incubation with 2.5 ml trypsin for 2-3' at 37 °C and inactivation of trypsin by 7.5 ml of growth medium.

For long-term preservation, HeLa cells were harvested as described above, pelleted at 200 g for 2', resuspended in 10 % DMSO in FBS Gold, followed by immediate freezing to -80°. After 1 day, the aliquots were transferred to -150 °C.

7.3.2 Transfection

To introduce plasmid DNA into the HeLa cell line stably or transiently, Turbofect reagent was used. Approx. 1.5 x 10^6 cells were plated 24 hours prior transfection on a 10 cm petri dish. 5 µg plasmid DNA and 5 µg carrier DNA (random plasmid without eukaryotic functional domains) were premixed with 120 µl Optimem medium and separately 10 µl Turbofect with 110 µl Optimem. Both mixes were incubated 5' at RT, combined and incubated 20' more before addition to the cells. After 24-36 h under usual cell culture conditions, cells were washed. Next, selection by antibiotics with subsequent screening and FACS were performed to isolate stably transfected clones.

7.3.3 Selection and screening for stable TALE-HeLa clones

TALE transfected HeLa cells were selected for stably introduced construct on Hygromycin 50-100 µg/ml from day two after transfection for 10-12 additional days. Plates were screened under a fluorescent microscope for formed GFP-positive colonies and ones with highest and evenly distributed GFP-signal were picked after mild trypsinization. As trypsinization was short, cells did not detach from plate until collected by a pipette, which minimized contamination risk by neighboring clones. As clones were picked under unsterile conditions, 2x penicillin/streptomycin was supplemented to the growth medium for 2 weeks.

FACS of high and low GFP population was performed on the FACS Aria (INST 247/645-1 FUGG) equipment. Hereby gates were set to restrict SSC-A, SSC-W and FSC-A to sort only single cell sized and shaped events. Dead cells were excluded by DAPI staining, excitation laser 405 nm, detection filter 450/50 nm. GFP signal was measured by excitation laser 488 nm and detection filter 530/30 nm. GFP-positive cells were sorted to 6-well plates.

7.4 Bacterial procedures

7.4.1 Preparation of chemocompetent *E. coli*

DH5alpha strain was obtained from previous laboratory stock, streaked on an LB-Agar plate for single colonies, inoculated to LB-Lennox medium and cultured until visual density at 37 °C shaking. The cells were harvested in 50 ml Falcon tubes by 2,000 g, 10' centrifugation and resuspended in 30 ml of 4 °C 0.1 M $CaCl_2$. Cells were again pelleted by centrifugation and resuspended in 15 ml precooled $CaCl_2$. After 30-

45’ incubation on ice, cells were pelleted and resuspended in 1.5 ml precooled $CaCl_2$. Glycerol was added to a final concentration of 20% and 50 µl aliquots were stored at -80 °C.

7.4.2 Transformation and strain preservation

50 µl chemocompetent DH5alpha were thawed, incubated 10’ on ice and plasmid DNA added. The exact amount of DNA varied, usually 1 µl of a plasmid miniprep sample or 5 µl of a TALE-assembly reaction. After 30’ incubation on ice, the tube was heat shocked at a 42 °C shaker for 90’’ and then immediately transferred on ice for 1-2’. 700 µl 37 °C warm LB-Lennox medium was added to the sample followed by 1 h incubation on a shaker at 37 °C. A fraction of the culture was finally plated on LB-agar with an appropriate antibiotic (see section 8.2.1 for concentrations) and, in case of blue-white selection, with additional 40 µl/plate Xgal (20 mg/ml) and 4 µl IPTG (1 M). Single colonies were picked after 1 day at 37 °C or 3 days at RT, plasmid isolated, further analyzed by restriction or Sanger sequencing and the correct clones preserved at -80 °C in 20% glycerol, 80% LB-Lennox.

7.4.3 Plasmid isolation by alkaline lysis (Miniprep)

1 ml of clonal overnight culture was centrifuged 3’ at 6,000 g, RT. The pellet was resuspended in 100 µl miniprep P1 buffer. 100 µl P2 buffer were added and mixed by gentle tube inversions. After 5’ at RT, 150 µl of P3 was added. The lysate was centrifuged at RT 10’, 21,000 g. The SN was added to 0.7 volumes of Isopropanol, mixed and centrifuged at RT, 25’ 21,000 g. The pellet was washed with 100 µl 70% EtOH and dried at 37 °C, followed by resuspension in 30µl mQ-H_2O with 300 µg/ml RNAse A.

7.4.4 Midiprep

Midiprep purification of plasmids was performed by the peqGOLD Xchange Plasmid Midi Kit from Peqlab according to the manufacturer’s manual using the filter-based approach.

7.4.5 BAC-plasmid isolation by alkaline lysis (Maxiprep)

500 ml DH10B RP23-118H24 ON culture on chloramphenicol selection were used as Maxiprep input. The Maxiprep was performed according to Sigma Aldrich protocol: (http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/na0100bul.pdf) Hereby the referred kit’s components were substituted by self-prepared buffers: Resuspension solution by Miniprep P1 buffer, Lysis solution by Miniprep P2 buffer and Neutralization solution by Miniprep P3 buffer. mQ-H_2O was used for elution, RT 100% isopropanol for the precipitation step. Endotoxin removal was not performed.

7.5 TALE-cloning

A fast cloning procedure with standardized 2-3 step cloning protocols for up to 31 monomers was performed with the Voytas lab Golden Gate TALEN and TAL Effector Kit 2.0 (Addgene kit, #1000000024) according

to manual. The kit contains plasmids for all single monomers, the cloning intermediate stages and the final C- and N-terminal TALE-domains[168]. Hereby, in the first reaction up to 10 monomer sequences are ligated in the correct order. In the second step, the preformed monomer arrays are combined and recloned to the final eukaryotic expression vector. The 4 monomers containing binding site amino acids NI, NG, HD and NK corresponding to nucleotides A, T, C and G respectively were used in the assembly. Both C- and N-terminally a variety of fluorescent markers, purification tags or functional domains as demethylases, phosphotransferases, transcription activators or restriction enzymes can be fused[144]. Here, N-terminally a 3xFlag-tag and eGFP were fused, latter of which is release on translation by the autohydrolyzing P2A peptide. First, the pTal5 vector was restricted with HindIII-HF and EcoRV and ligated into the HindIII/EcoRV restricted pcDNA3.1 vector. The 3xFLAG-tag sequence was introduced by a phosphorylated 3xflag oligonucleotide, ligated into the HindIII-restricted pcDNA3-TALE plasmid. *eGFP-P2A* sequence was inserted at the NheI restriction site to the pcDNA3-3xFLAG-TALE plasmid from the NcoI/EcoRI-restricted pSlax eGFP-P2A vector.

Correct assembly of TALE-plasmids was assessed by restriction analysis and additionally by Sanger sequencing to exclude possible point mutations.

7.5.1 Flag-TALE

7.5.1.1 Target selection

TALE-binding site selection was performed with the following criteria: Unique 19 bp sequence was checked by BLAST software, highest off-target hit was 16 out of 19 bp, corresponding to a 40-fold lower binding[131]; First base in a sequence was a T due to general TALE architecture; Targets on the same gene were at least 1000 bp apart to provide sufficient resolution. Where possible, DNAseI sensitive sites were selected; DNA-methylation was avoided.

7.5.1.2 Recloning to expression vector

The full TALE-assemblies #10-19 in the pcDNA-eGFP-P2A-3xFLAG-TALE vector were restricted with XhoI, SpeI and ApaLI, the 5 kb band excised from gel and purified, followed by ligation to the final pcDNA 3.1 vector with a hygromycin resistance gene.

7.5.2 TALE-Lsd1 in vivo

Lsd1 sequence was cut from the pGemT-easy-Lsd1 vector by PmlI and inserted to the EcoRV-restricted pcDNA-eGFP-P2A-3xFLAG-TALE vector. The lox-Neo-stop-lox cassette was introduced via the NheI/NotI restriction sites. The fully assembled TALE-construct was further recloned to the hygromycin expression vector as described in section 7.5.1.2.

Transgenic animals were generated at the Institute Clinique de la Souris in Illkirch, France.

7.6 Chromatin immunoprecipitation

7.6.1 H3K27Acetyl-ChIP

300,000 FACSed formaldehyde fixed cells were used for each ChIP replicate, 1/16 of the total volume was taken as input after sonication.

Saturated antibody-bead complexes were prepared prior to chromatin addition. 10µl Protein A and 10 µl Protein G agarose beads (EzView RED A/G affinity gel, Sigma-Aldrich P6486, E3403) slurry were resuspended and washed 3x in 500 µl RIPA buffer at 4 °C with 600 g centrifugation steps. 0.5 µl H3K27Acetyl Ab (Abcam, #4729) was added in 200 µl RIPA buffer and incubated 3h at 4 °C on an overhead wheel at 11 rpm. Beads were washed 3x with RIPA buffer to eliminate unbound Ab, and SN removed.

Cells were thawed on ice, spun down for 5' at 4°C and 10,000 g to a residual volume of ~20 µl and lysed by vortexing 2x 5'' after the addition of 18 µl RT lysis buffer. After 5' on ice, cells were vortexed again. Sonication with 56 cycles 10'' ON / 90'' OFF at high setting yielded 150-700 bp chromatin fragments. Next, the sample was diluted with 600 µl RIPA buffer and centrifuged at 16.200 g 10', 4 °C. SN was saved and pellet resuspended in 200 µl fresh RIPA buffer, centrifuged, and SNs pooled. 1/16 was taken as input and kept at 4 °C.

The remaining material (~ 1 ml) was added to the saturated antibody-bead complexes for ON incubation on an overhead wheel at 11 rpm, 4 °C. Next, tubes were centrifuged 2' at 600 g, 4 °C and the SN replaced by 1 ml cold RIPA buffer. Beads were resuspended on an overhead wheel for 4'. The washing steps were repeated 3 times, plus and additional wash with TE buffer.

After removal of TE buffer, bead bound ChIP-material and the previously taken input were further processed simultaneously. 150 µl elution buffer were added and incubated 20' at 65 °C, shaking at 1300 rpm. SDS (1% final) and RNAse solution, 20 µg for input and 10 µg for ChIP material, were added, followed by 30' incubation at 37 °C, shaking. Proteinase K, 20 µg for input and 10 µg for ChIP material, were added, followed by 2h incubation at 55 °C, shaking. SN was removed after 1' centrifugation at 11.600 g, pellet resuspended in 150 µl elution buffer, incubated 5' at 65 °C, shaking, spun down 2' and SN pooled. Elution buffer was added to a total volume of 550 µl and DNA purified by Phenol/Chloroform extraction and ethanol precipitation (see section 7.13).

Initial QC was performed by Qubit and Bioanalyzer measurements for quantity (ChIP and input samples) and chromatin fragment size distribution (input only) respectively.

Library preparation with NuGEN Ovation Ultralow Library System V2 and sequencing on Illumina HiSeq with 50 bp single reads was performed by IMB CF Genomics.

7.6.2 H3K27me3

7.6.2.1 NP-buffer based protocol

This is a slightly modified version of the protocol used previously by my colleagues to generate the first replicate of H3K27me3 ChIP-seq. Main differences to the H3K27Acetyl ChIP protocol is the usage of NP buffer instead of RIPA buffer and a chromatin preclearing step with empty beads. Preclearing beads were prepared in parallel with antibody-loaded ones, each 10 µl Protein A and 10 µl Protein G conjugated. Preclearing was performed directly before addition of Ab-loaded beads by incubation the chromatin sample (input set aside) 45' with unloaded beads at 4 °C, 11 rpm on an overhead wheel. After preclearing bead removal, beads loaded with 2 µg / sample H3K27me3 Ab (Millipore #17-622) were applied.

7.6.2.2 final H3K27me3 protocol

The final protocol, used for the 3rd replicate of H3K27me3 ChIP-seq, is a modified H3K27Acetyl ChIP protocol with a strongly reduced bead amount. 1 µg H3K27me3 Ab (Millipore #17-622) was applied on 2.5 µl Protein A mixed with 2.5 µl Protein G beads per sample. Additionally, TE buffer was used with TRIS-Cl concentration increased to 20 mM.

7.6.3 H2A.Z

500,000 FACSed formaldehyde fixed cells were used for each ChIP replicate. ChIP was performed according to the H3K27Acetyl protocol described in section 7.6.1 with following modifications: 0.25 µl H2A.Z Ab (Active Motif, #39113), no NaButyrate was used.

7.6.4 Flag-ChIP

Although several other protocols were tested, best results were achieved by the Fujii lab protocol[169-171] with the following modifications: 10^7 HeLa cells input, 50 cycles (10'' ON, 80'' OFF) sonication, no IgG preclearing step.

7.7 RNA isolation

7.7.1 RNA isolation from tissue

Total RNA was isolated from embryonic intestinal tissue by the EN-RNeasy-Micro kit according to the manufacturers manual with the following specifications: Material and RNA were generally stored at –80 °C, while all protocols were performed at RT. The tissue was harvested in RLT buffer and immediately

frozen. In addition to the lysis buffer and thawing procedure, the tissue was disrupted by the Polytron homogenizer.

7.7.2 cDNA-synthesis

RNA was reversely transcribed by Superscript II reverse transcriptase by the following protocol: 150 ng total RNA were premixed with 1 µl of 4x 10 mM dNTPs mix, 1µl random primers and filled up to 10 µl with water followed by 5' incubation at 70 °C and pausing on ice. In the second premix, 4 µl 5x First Strand buffer were combined with 1µl 0.1 M DTT, 1 µl RNAsin, 1 µl Superscript II reverse transcriptase and 2 µl water. The second premix was added to the first, incubated 10' at 25 °C, 50' at 42 °C and finally the reaction was inactivated at 70 °C for 15'.

7.8 DNA/RNA analysis

7.8.1 Qubit

Qubit measurements of DNA (e.g. ChIP input and output), DNA-ISH probes (labeled/unlabeled) and RNA-probes was performed with DNA high sensitivity and RNA high sensitivity reagents respectively according to manufacturer's protocols.

7.8.2 Bioanalyzer

DNA analysis, such as ChIP-input and library size assessment, was performed on Bioanalyzer with DNA high sensitivity chips, fully according to Agilent manuals.

7.8.3 Sanger sequencing

Sanger sequencing of plasmids and transgenic embryonal DNA was performed with StarSEQ GmbH. Sample preparation was performed with own sequencing primers according to the provided manual. Analysis of sequencing results was performed with FinchTV and CLC Sequence Viewer 6 software.

7.8.4 Dot Blot

Sample and control DNA were boiled at 99 °C in up to 500 µl/well 2x SSC buffer for 10' and cooled on ice for 5'. The Dot Blot gasket (Bio-Dot, BioRad) was assembled with nitrocellulose membrane according to BioRad manual. Samples were applied and washed with 2x SSC. The 2x SSC rinsed membrane was air dried and the DNA immobilized by UV. The membrane was washed in 1x PBS + 0.1 % Tween20 and blocked in Blocking buffer with 0.1 % Tween20 for 1h. Streptavidin-HRP was added at 1:500 dilution ON at 4 °C. After washing 4x in Blocking buffer with 0.1 % Tween20, the membrane was imaged as in Section 7.9.4.

7.9 Protein analysis

7.9.1 protein isolation

7.9.1.1 Protein isolation from HeLa cells

Approx. 5 x 10^6 HeLa cells were harvested per sample, pelleted and resuspended in 300 µl HeLa lysis buffer. Cells and chromatin were disrupted by sonicating at high setting 5-12x (10'' ON + 90'' OFF) and SN further used after centrifugation at 16,000 g, 10'.

7.9.1.2 Protein isolation from murine tissue

Embryonic heart and lung material was extracted for subsequent WB and IP-WB and frozen to -80 °C if not immediately processed. For WB, the tissue was taken up in 100 µl SDS-lysis buffer, mechanically lysed by a pestle in an Eppendorf tube and sonicated 5 cycles 30'' ON / 90'' OFF at high setting. For Immunoprecipitation, the tissue was taken up in 1 ml HEPES-buffer, mechanically lysed and sonicated 3 cycles 20'' ON / 90'' OFF at high setting.

7.9.2 Immunoprecipitation

The non-fixed, sonicated material from embryonic tissue or cultured cells was centrifuged at 21,000 g, 4 °C for 5'. The SN was split: 50 µl were stored as input at 4 °C, while 900 µl were incubated ON at 4 °C with 4 µl FLAG-Ab. A/G agarose beads were washed in cold HEPES-buffer and centrifuged at 400g, 2' twice before being added to the IP material and incubated at 4 °C for 2h at 11 rpm in the an overhead wheel. The beads were washed with cold HEPES buffer, the SN removed and protein SDS-PAGE loading buffer added to beads and the input material for subsequent SDS-PAGE.

7.9.3 Polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE setup was handled according to manufacturer's protocols. Separating gels were used with 6-12% acrylamide, depending on the protein of interest size. E.g. the 9% separating gel was prepared by mixing 1.5 ml lower SDS-gel buffer with 1.2 ml glycerol, 1.8 ml acrylamide 30%, 1.5 ml H_2O and polymerizing with 80 µl 10 % APS and 6 µl TEMED at RT for 20'. The 4% stacking gel was prepared accordingly with upper SDS-gel buffer. The samples in 1x SDS-PAGE loading buffer as well as the protein standard ladder were boiled 5' at 95 °C prior to loading on the gel. The gels were usually run in 1x Laemmli buffer at 80 V until the separating gel was reached, 120-160 V thereafter.

7.9.4 Western blot

Proteins were transferred from the gel to a PVDF membrane according to the manufacturers protocol for the Semi-dry blotting apparatus. Equal loading of input material was assessed by Ponceau red staining of the membrane before blocking. Ponceau red staining solution was added to the membrane and washed several times, imaged and blocked with blocking buffer 1h at RT, which also removed the remaining Ponceau staining. Primary antibodies were added to blocking buffer and incubated either 1h at RT or ON at

4 °C shaking. After washing 3x 5’ with 1x PBS, a secondary, peroxidase-conjugated Ab was added in blocking solution and incubate 1h at RT shaking. After 3 more washing steps, chemiluminescent substrate from the SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Fisher Scientific, #34087) was added and imaged by Chemidoc XRS+ according to manual.

7.10 *In situ* hybridization

7.10.1 PCR-based PLA-probe design and preparation

ISH probes and the primer pairs used for their generation were selected for unique genome/transcriptome binding by NCBI Primer-BLAST software. Probe sequence was amplified by PCR from wild type template gDNA with homemade Taq-polymerase as in section 7.1.3.1 with the following program:

PCR-program:
94 °C – 3 min
94 °C – 30 sec
67 °C – 30 sec x 33
72 °C – 30 sec
72 °C – 2 min
10 °C - ∞

The PCR product was purified by gel electrophoresis and ligated into the pGEM-T vector by manufacturer’s protocol with blue/white clone selection after transformation. Plasmids were isolated by the miniprep method and correct insertion of the probe sequence was confirmed by PCR, restriction analysis and Sanger-sequencing. Correctly assembled plasmids were isolated for further use by Midiprep and clones were preserved at -80 °C. In case of PLA-probes for *Epcam* and *Shh*, the pGEM-T cloning step was skipped and the following labeling protocol was performed on the purified PCR-product instead of a plasmid.

7.10.2 BAC ISH-probe preparation

A Maxiprep of the BAC-plasmid was sonicated 10 cycles 30’’ ON / 90’’ OFF at high setting and purified by the Illustra MicroSpin G-50 columns according to manufacturer’s protocol with centrifugation steps at 1500 g.

7.10.3 DNA ISH-probe labeling

7.10.3.1 Biotin probe labeling

Purified probe plasmids were amplified by PCR as in section 7.10.1 with 36 cycles and purified by the Illustra MicroSpin G-50 columns according to manufacturer’s protocol with centrifugation steps at 700 g.

10 µg sonicated BAC or 1 µg probe PCR-product were biotinylated by thermal coupling in accordance to the Longarm Photoprobe Biotin kit manual and purified by the Illustra MicroSpin G-50 columns with centrifugation steps at 1000 g.

Biotinylated probes were relatively quantified by Nanodrop to ensure equal loading amounts during ISH. Absolute quantification with Nanodrop or Qubit was not possible due to the interfering fluorescent spectrum of biotin. BIO-probes were stored at +4 °C.

7.10.3.2 Fluorescent probe labeling

BAC was labeled with AF 594 according to ULS Alexa Fluor 594 labeling kit protocol with following deviations: 1 µg sonicated BAC input in 2 µl H_2O, no DNAseI treatment. Labeled BAC was purified by Illustra MicroSpin G-50 columns with centrifugation steps at 1000 g and stored at +4 °C.

7.10.3.3 Digoxigenin probe labeling

DIG-labeling was performed by an PCR-based approach. DIG-11-dUTP was incorporated together with unlabeled dNTPs (1 mM dATP/dCTP/cGTP, 0.9 mM dTTP, 0.1 mM DIG-dUTP).

PCR reaction composition (15 µl total):	Reaction Buffer 10 x	1.5 µl
	dNTP/DIG-dUTP Mix	3 µl
	Forward/Reverse primer (10 µM)	1 µl
	Taq-Polymerase	0.25 µl
	H_2O	8.25 µl
	Template-DNA (PCR-product)	1 µl

PCR-program:
94 °C – 3 min
94 °C – 30 sec
67 °C – 30 sec x 25
72 °C – 30 sec
10 °C - ∞

The degree of DIG-incorporation was checked by agarose gel electrophoresis in comparison to unlabeled PCR-product and concentrations measured by Nanodrop after purification by Illustra MicroSpin G-50 columns with centrifugation steps at 700 g. DIG-probes were stored at -20 °C or up to a month at +4 °C.

7.10.4 RNA ISH

7.10.4.1 RNA probe design and cloning

RNA ISH probes and the primer pairs used for their generation were selected for unique genome/transcriptome binding by NCBI Primer-BLAST software. Probe sequence was optimally selected

to be 300-700 bp long, contain no repetitive elements and to coincide with the targeted gene's RNA-seq peak at the 3'UTR prior the polyA signal sequence. Probe sequence was amplified by PCR and cloned into the pGEM-T vector as in section 7.10.1. Plasmid Midipreps were analyzed for correct insertion and its direction by Sanger-sequencing. Correct clones were preserved at -80 °C.

7.10.4.2 RNA probe labeling with digoxigenin

20 µg of the probe containing plasmids were linearized 3 h at 37 °C. Forward inserted probes were linearized with NcoI enzyme for later transcription by the SP6 RNA-polymerase and the reverse inserted ones with NdeI for the T7 polymerase respectively. Both restriction enzymes were non-cutters in the used probe sequences. Linearized plasmids were isolated by Phenol/Chloroform extraction and subsequent ethanol precipitation as described in section 7.13.

DIG-labeled RNA-probes were synthesized from 1 µg linearized plasmid for 2 h at 37 °C in a 20 µl reaction containing: 2 µl DIG RNA Labeling Mix, 4 µl 5 x Transcription-Buffer (PROMEGA), 2 µl T7 or SP6 RNA-Polymerase, 2 µl 100 mM DTT and 0.5 µl RNAsin. Template DNA was then digested by 1 µl DNAseI for 30' at 37 °C.

Synthesized DIG-RNA probes were purified by the Illustra MicroSpin G-50 columns according to manufacturer's protocol with centrifugation steps at 600 g. Quality was assessed spectrophotometrically and by gel electrophoresis. Until use, probes were aliquoted and stored at -80 °C.

7.10.4.3 DIG-RNA ISH

10 µm paraffin sections were melted at 60 °C for ≥2 h and deparaffinized in xylol 2x 5'. Rehydration was performed with the following series of EtOH/PBS dilutions: 2x 5' 100 % EtOH, 2' 70 %, 2' 50 %, 2' 30 % and finally 5' in PBS only. After fixation in 4 % freshly prepared FA in PBS for 15', slides were washed 3x 2' in PBS followed by 15' bleaching in 6 % H_2O_2/PBS solution and 3 washing steps. 10 µg/ml Proteinase K was added for 10' and sections fixed a second time for 15' in 4% FA with two subsequent 5' washes in PBS. Samples were transferred for 2' into 100 mM Tris-HCl pH 7.5 buffer and acetylated with additional 0.25 % acetic anhydride. After two washes in 2x SSC pH 5.0, slides were dehydrated with the following series of EtOH/PBS dilutions: 2' 30 %, 2' 50 %, 2' 70 % and finally 5' in EtOH only. Samples were air-dried and BioRad Frame-Seal chambers (9x9 mm) applied. Alternatively, grease pen delimitations of the ISH area covered with a glass coverslip was used instead of the chambers.

RNA-hybridization buffer was prewarmed to 70 °C and DIG-RNA probes were thawed on ice before 3-5' denaturation at 80 °C. 24 volumes of hybridization buffer were immediately added to the denatured DIG-RNA, mixed and quickly applied to the sample, sealed and incubated ON at 65 °C in a dark and humid chamber. Chambers were carefully removed and slides were washed at 60 °C with gentle shaking: 15' in 5x

SSC + 50 % formamide, 2x 15’ in 2x SSC + 50 % formamide, 2x 15’ in 1x SSC, 30’ in 0.2x SSC and 5’ followed by 5’ at RT in TBS. 700 µl/slide blocking solution was applied in a humid chamber for 120’. Alkaline phosphatase (AP)-conjugated anti-DIG Ab was added ON at 4 °C at a 1:1000 dilution in the blocking solution. Slides were washed 6x 30’ in TBSX with gentle shaking and 2x 10’ in NTMT. 700 µl/slide staining solution (NTMT + 0.1% NBT + 0.35 % BCIP) were applied and kept in a dark humid chamber until blue precipitate developed. After sufficient development, staining was terminated by 2x 5’ washes in PBS.

Samples were mounted in Roti-Histokitt II after following EtOH dehydration series: 2’ H_2O, 2’ 30 % EtOH, 2’ 50 %, 2’ 70 %, 2x 2’ 100 % EtOH and finally 2x 2’ 100 % Xylol. Slides were dried ON and stored at 4 °C until imaging with the DM2500 microscope.

Where not otherwise indicated, the protocols was performed at RT.

7.10.5 FISH

FISH was performed according to the first part of the PLA protocol (Section 7.11). 150 ng of fluorescent probe were applied. After ISH and all washing steps including PBS, the samples were air dried and mounted in Duolink In Situ Mounting Medium with DAPI. Imaging was performed with the SP5 Confocal fluorescent microscope.

7.10.6 IF on DNA-ISH

BIO-BAC ISH was performed according to the first part of the PLA protocol (Section 7.11). 200 ng biotinylated probe were applied. After ISH and initial blocking in BSA/PBS, AF488-conjugated streptavidin was added for 1h at RT in 1xPBS + 3% BSA. After washing in PBS, the slides were mounted in Immumount and imaged with AF7000 microscope.

In case of secondary IF on BIO-BAC ISH, same protocol was followed with an additional washing and antibody incubation round. Primary Ab: anti-biotin, secondary Ab: AF647-conjugated anti-goat.

7.11 Proximity ligation assay (PLA)

7.11.1 Sample preparation

Adult or embryonic (E12.5 or E14.5) small intestine was extracted from wild type animals and washed in 1x PBS, supplemented with 20 mM NaButyrate to preserve acetylation. Adult gut was cut open

longitudinally and additionally washed. Fixation was performed at 4 °C, ON in 4% formaldehyde in 1x PBS with NaButyrate, followed by 2x 10' washes with cold 1xPBS. The tissue was incubated 2x 2h in 70 % Methanol / 30 % PBS and 2x 2h (or longer if stored at -20 °C) in 100 % Methanol. After 2x 10' in RT xylol (1x 30' for adult tissue), the samples were transferred for at least 3h to liquid Paraplast Plus and embedded in fresh Paraplast with slow solidifying at RT for 30'. Afterwards, the block was transferred ON to +4 °C. Solidified Paraplast blocks were cut to 5 µm section with the Leica RM2255 microtome.

7.11.2 DNA-protein PLA

This PLA protocol was developed from Gomez et al.[151] and the Duolink PLA manual.

5 µm paraffin sections were melted at 60 °C for 2h or ON and deparaffinized in xylol 2x 5'. Xylol was removed by two 1-2' washes in 100 % EtOH and preheated 1mM EDTA pH 8.0 was added for 20' at 37 °C. After washing in H_2O, 0.5 % TritonX in PBS was added for 20' at RT with gentle shaking. After brief washing in H_2O, preheated PK buffer was added and slides were incubated 10' at 37 °C. Dehydration was performed by short (~1') steps in an ethanol/H_2O series: 30 %, 50 %, 70 %, 90 %, 100 %, followed by air drying at RT for 10'. 100 ng BIO- or DIG-labeled ISH-probe (or H_2O as negative control) were denatured in PLA-hybridization buffer at 80 °C, 10' in 25 µl total volume and briefly iced before application to the sample in an BioRad Frame-Seal chamber (9x9 mm). The slide with the closed chamber was heated 3-5' at 80 °C in a PCR block and transferred for 3 days to 37 °C.

The chamber was then removed and samples washed in 1ml/slide: 15' and twice 5' at 42 °C in PLA-WashB-L1, 3x 10' at 60°C in 0.1 SSC, once at 42 °C in PLA-WashB-L2 and 3x at RT in PBS. The slides were blocked in 6 % BSA / PBS for 30' at RT. Before application, primary antibodies were spun down at 16,000 g for 3' to avoid precipitates. In 3 % BSA / PBS a combination of a rabbit (anti-H3K27Ac Abcam #4729, anti-pan-histone Abcam #1791) and a goat (anti-DIG Biozol MB7000, anti-BIO Biozol SP3000) primary Ab was added in an 200 µl droplet for 2h at RT or ON at 4 °C. Drying of the samples was necessarily avoided throughout the following steps. With gentle shaking, samples were washed with 200-250 ml PBS at RT 3x 5'. PLUS and MINUS PLA-probes (anti-goat/rabbit antibodies with attached oligonucleotides) were diluted 1:5 (20% of each PLA-probe) in 3 % BSA / PBS and incubated 20' at RT before applied as an 40 µl open droplet to the sample for 1h incubation at 37 °C in a humidity chamber. As previously with PBS, the sample was washed twice with PLA-WashB A. 20% Ligation stock and 2.5% Ligase from the Duolink detection reagents kit RED or FAR RED were mixed in H_2O and added as a 40 µl droplet to the sample for 30' at 37 °C at high humidity. Washing with PLA-WashB A was repeated. 20% Amplification stock and 1.25% polymerase from the Duolink detection reagents kit RED or FAR RED were mixed in H_2O and added as a 40 µl droplet to the sample for 2h at 37 °C at high humidity. During the incubation light was excluded and evaporation checked: in case of drying, the remaining droplet was diluted by 10-20 µl H_2O. Under

continuing light exclusion, the sample was washed twice with PLA-WashB-B 1x, 10’ at RT and once more with PLA-WashB-B 0.01x, 1’ at RT, each with gentle shaking. Slides were dried at RT and mounted in the aqueous Duolink In Situ Mounting Medium with DAPI for widefield fluorescent imaging with AF7000. Slides were stored in the dark at 4 °C for several days or at -20 °C long term. Fluorescent PLA signal remained unchanged for at least 1 month at -20 °C.

7.11.3 Protein-protein PLA

Protein-protein PLA on paraffin sections was performed with E14.5 wt type tissue according to DNA-protein PLA protocol with a mock ISH and following deviations: Both H3K27Ac (Abcam #4729 and pan-histone (Abcam #1791) primary antibodies were used raised in rabbit. PLA probes Rabbit PLUS and Rabbit MINUS were used.

Protein-protein PLA on cryosections was performed with tissue from a constitutively active *Hprt-Cre* mouse at stage e11.75. Cryosections were treated with 0.1 % Triton X in PBS, 20’ at RT, washed in H_2O and proceeded with primary Ab incubation as in section 7.11.2. A single primary Ab, rabbit raised anti-Cre (Novagen #69050-3) was used and, after washing, partially bound by a diluted secondary goat anti-rabbit Ab (from Biozol kit PK6101). PLA-probes than were applied to bind either the primary rabbit or the secondary goat Ab by the usual protocol.

7.12 Microscopy

7.12.1 Sample preparation

7.12.1.1 Tissue sections

Paraffin sections for PLA and ISH were prepared as described in section 7.11.1. For ISH, 10 µm thick sections were prepared.

Cryosections were prepared from embryonic and adult tissue, fixed 20’ at 4°C with 1% formaldehyde. After 2x 5’ washes with 1x PBS, samples were incubated in 30 % sucrose ON and embedded in O.C.T. medium. Frozen blocks were cut with the Leica cryotome system to 10 µm sections. Sections were stored at -20 °C or -80 °C for longer periods. Before imaging, sections were brought to RT, equilibrated in PBS and fixated 10’ in 1 % formaldehyde, followed by 2x 3’ washes in PBS.

7.12.1.2 Whole crypts

Intact crypts were collected from *Lgr5-GFP* mice according to FACS sample preparation protocol in section 7.2.1 until and including formaldehyde fixation but without any of the single cell digestion and resuspension procedures. Fixed intact villi can be collected by the same protocol if avoiding harsh pipetting. Unfixed

crypts were processed the same way, however they disintegrated quickly and had to be processed within 1h at 4 °C. Native processing of villi was not possible.

7.12.2 Immunofluorescent staining

7.12.2.1 FLAG-TALE localization by confocal microscopy

Cells were grown in IBIDI chambers, washed with 4 °C PBS, fixed in freshly prepared 4 % paraformaldehyde/PBS for 20' at RT and washed twice in cold PBS prior to permeabilization in 0.5% TritonX / PBS for 60'' at RT. After washing twice in cold PBS, cells were blocked 20' at RT in 5% BSA / 5% Goat serum / 1x PBS. FLAG-Ab was added for 2h, RT at 1:500 dilution in 1.5% BSA / 1.5% Goat serum / 1x PBS. After 3x 5' washing in PBS, DAPI and AF594-anti-mouse Ab were added at 1:1000 dilution in 1.5% BSA / 1.5% Goat serum / 1x PBS for 1h at RT under light exclusion. After 3x 5' washing in PBS, cells were imaged in PBS with the SP5 confocal microscope for DAPI-, AF594- and eGFP-signal.

7.12.2.2 IF test of primary antibodies for PLA

Paraffin sections were incubated according to the first part of the PLA protocol (Section 7.11). A mock ISH was performed without a probe. After ISH and initial blocking in BSA/PBS, primary Ab was applied: H3K4me3 (Millipore CS200580 or Abcam #8580), H3K4me3 (Millipore #05-1338) or H3K27Ac (Abcam #4729) for 1h at RT in 3% BSA / 1x PBS (see section 8.2.5 for details on antibodies). After 3x washing with PBS at RT, secondary Ab was incubated for 1h at RT in 3% BSA / 1x PBS: anti-rabbit AF594 (Invitrogen A11072), anti-mouse AF488 (Invitrogen A10684) or anti-rabbit biotinylated Ab from the Biozol Vectastain ABC kit (PK6101). In case the biotinylated secondary Ab, a tertiary fluorescently labeled Ab was applied after washing for 1h at RT in 3% BSA / 1x PBS: Streptavidin-AF488 (Invitrogen S32354). Samples were washed in PBS and either counterstained with DAPI and mounted in Immumount for confocal imaging with SP5 or directly mounted in DAPI containing Duolink In Situ Mounting Medium for widefield fluorescent imaging with AF7000.

7.12.2.3 IF staining with FACS- and ChIP-antibodies on cryosections

Cryosections were blocked in 5% FBS (for UEA1, CD24), goat (for Epcam) or donkey (for H2A.Z) serum, depending on the fluorescently labeled antibody's host, in PBS supplemented with 0.1 % Tween20. Hereby, the sample area was delimited with a grease pen, 120 µl blocking solution were applied and closed with a cover slip for 1h RT incubation in a humid chamber. After removal of the blocking solution, primary antibodies were added and incubated in 1 % serum solution as above. Primary antibodies: Epcam (eBioscience, 46-579182 and 17-579183), UEA1 (Sigma, L4889), CD24 (Sigma, #560536), H2A.Z (Actif Motif, #39113). In case of Epcam and H2A.Z stainings, samples were washed 3x 5' in PBS and stained by secondary fluorescent Ab under same conditions as above: goat anti-rat AF568 (Invitrogen, A11077) for

Epcam or donkey anti-rabbit AF568 (LifeTechnologies A11070) for H2A.Z. Final step antibodies were incubated with 1µg/ml DAPI to counterstain nuclei. Samples were washed 2x 10' in PBS and mounted with Immumount. Slides were stored until imaging with AF7000 for up to 3 days at 4 °C.

7.12.2.4 Lgr5-GFP and IF staining with FACS-antibodies on whole crypts

Staining of intact crypts was performed in Eppendorf tubes and imaging in IBIDI chambers.

Native Lgr5-GFP imaging on fixed/unfixed crypts was performed directly in PBS, without any additional processing.

For Paneth IF staining, crypts were resuspended in 2 % FBS and stained 15' at RT with 1:200 dilution of primary antibodies: UEA1 (Sigma, L4889), CD24 (Sigma, #560536).

For secondary IF Lgr5-GFP staining, crypts were blocked in 5 % goat serum in PBS + 0.1 % Tween20 for 30' at RT and incubated 30' at RT with primary anti-GFP Ab (LifeTechnologies, A11122) at 1:200 dilution in 1 % goat serum/PBS + 0.1 % Tween20. After 2x washes in PBS at 4 °C with 200 g centrifugation steps, secondary anti-rabbit Ab. conjugated to AF488 (LifeTechnologies, A11070) was added for 15' at RT.

Fixed crypts were incubated with 1µg/ml DAPI to counterstain nuclei. Crypts were washed 2x in PBS at 4 °C with 200 g centrifugation steps and resuspended in 200 µl cold PBS in IBIDI chambers for immediate imaging with AF7000.

7.12.3 Single cell PAS-staining

Formaldehyde fixed cells were FACSed to Eppendorf tubes and pelleted at 2600 g for 3'. Cells were rinsed twice with H_2O, covered with 0.1 % periodic acid solution for 5' and rinsed twice more with H_2O. After 5' incubation in Schiff's solution, samples were washed 3x 3' in tap water and transferred in a 50 µl droplet to a glass slide for drying at 60 °C, before mounting in 5 µl H_2O. Imaging was performed with the DM2500 microscope.

7.12.4 Microscopes and imaging parameters

7.12.4.1 Widefield fluorescent microscope AF7000

AF488 IF and IF-ISH samples were imaged with filter cube L5 (480/40 nm excitation, FT505 nm, 527/30 nm emission), exposure: 100 ms, gain: 200, intensity: 5.

AF568, PE-Cy7 and TRITC IF samples were imaged with filter cube (546/12 nm excitation, FT565 nm, 600/40 nm emission)

DAPI channel was imaged with filter cube A4 (360/40 nm excitation, FT400 nm, 470/40 nm emission), exposure: 10-20 ms, gain: 90-100, intensity: 2-5.

GFP was imaged with filter cube L5 (480/40 nm excitation, FT505 nm, 527/30 nm emission).

PLA RED signal was imaged with the filter set TRI-red (570/20 nm excitation, 640/40 nm emission), exposure: 130 ms, gain: 255, intensity: 5.

PLA FAR RED signal was imaged with the filter Y5(620/60 nm excitation, FT660 nm, 700/75 nm emission), exposure: 500 ms, gain: 255, intensity: 5.

Yellow channel for PLA RED background correction was measured with the filter N3 (546/12 nm excitation, FT565 nm, 600/40 nm emission), gain 255, intensity 5 and the exposure manually adjusted to yield signal slightly below the PLA channel.

PLA, IF-stainings by PLA protocol and FISH stainings on 5 µm sections were imaged with 0.2 µm resolved Z-stack. Usual voxel size: 102x102x200 nm. In most cases, a tile-scan was performed with automatic tile stitching by the Leica imaging software.

IF stainings on 10 µm cryosections were imaged as Z-stacks followed by best focal plane selection for each channel.

Whole crypts IF-stainings were imaged as Z-stacks followed by best focal plane selection for each channel.

7.12.4.2 Confocal microscope SP5

AF594 FISH and IF samples were imaged with 561 nm excitation, 605/28 nm emission for AF594 and 405 nm emission, 450/30 nm excitation for the DAPI counterstaining. DAPI channel was 2x line averaged, AF594 channel 3x line averaged. 6 µm imaging depth (1.3 µm for IF), 37x (9x for IF) Z-stack. 60x60x168 nm voxel (2048x2048, 1.4x zoom).

AF488 IF samples were imaged with 496 nm excitation, 530/22 emission for AF488. AF488 channel was 2x line averaged. 3.4 µm imaging depth, 28x Z-stack. 86x86x126 nm voxel (2048x2048, 1.4x zoom). DAPI settings as above.

7.12.4.3 Other microscopes

DM2500 microscope was used for bright-field imaging of RNA-ISH and PAS-staining samples.

7.12.5 Image processing with FijI/ImageJ software

7.12.5.1 Initial sample processing

Most IF, FISH and PLA experiments were imaged as Z-stacks and tile-scans with the AF7000 microscope. Tile stitching was generally performed by the Leica microscope operating software, all further processing by FijI. Z-stacks were generally used as maximum-projections, as PLA and ISH signal were sufficiently few and far between to be resolved in 2D. Where no further computational evaluation followed, fluorescent signals were contrast adjusted avoiding saturation. Where contrast enhancement led to black/white saturation, this is indicated in the shown figures. Although imaging was performed in the 16-bit format, all data was reduced to 8-bit for further processing and analysis.

7.12.5.2 PLA-signal calculation

The rationale of automated PLA-signal calculation is described in section 5.3.2.7. It was successfully performed with PLA RED and PLA FAR RED detection reagents. For PLA RED, the yellow and for PLA FAR RED, the PLA RED channel was imaged as shorter wavelength background channels (see section 7.12.4.1 for parameters).

PLA signal and background channels were combined to a binary mask by the following FijI macro:

```
imageCalculator("Divide create 32-bit", "PLA channel.tif","11-13 Background channel.tif");
// divides signal by background, background only pixels yield values around 1, PLA signal higher

setAutoThreshold("Default dark");      setThreshold(2, 255);
// assigns "0" to all pixels <2 (background), "1" to all pixels ≥2 (PLA signal).
// the threshold value can be adjusted depending on absolute values of both channels

setOption("BlackBackground", false);  run("Convert to Mask");
// generation of a binary black/white mask
```

For quality control, a subset of masked signals was compared to the original channels and vice versa. Counting of signals was either performed manually or by the following macro:

```
run("Analyze Particles...", "size=4-Infinity pixel circularity=0.40-1.00 show=Masks exclude include");
// only masked signals with 4 or more pixels and circular geometry are counted
```

7.12.5.3 IF-signal calculation

Weak IF signal on BIO-BAC ISH was transformed for quantitative analysis to a binary mask by the following FijI macro:

```
run("Duplicate...", "title=orig");  run("Duplicate...", "title=blur");
// creates two copies of the fluorescent channel

run("Gaussian Blur...", "sigma=4");
// the Gaussian blur algorithm generates a background noise image from the imaged signal

imageCalculator("Subtract create", "orig","blur");  setAutoThreshold("Default dark");
// the calculated background is subtracted from the original signal

setThreshold(7, 255); setOption("BlackBackground", false);
// minor signal/noise deviations are thresholded

run("Convert to Mask");
// binary mask of the fluorescent signal is generated
```

The binary mask was then analyzed either by hand or automatically for positive fluorescent signals. The used geometric properties were confirmed by visual screening of a subset of original fluorescent signals:

```
run("Analyze Particles...", "size=4-Infinity pixel circularity=0.40-1.00 show=Masks exclude include add in_situ");
// circular signals with at least 4 pixels
```

The imaged epithelial area was measured by FijI and signal density calculated. Efficiency of ISH was approximated from signal density, section thickness (5 µm) and average nuclear size (10 µm diameter) and distribution. With two ISH-probe loci per nucleus and ~4000 nuclear volumes per mm^2 of section (empirically determined), 100% efficiency would correspond to 8000 signals per mm^2.

7.12.5.4 RNA-ISH processing

RNA-ISH images were processed for enhanced contrast and background removal by ImageJ. Hereby, where necessary, the ImageJ background subtraction tool was used with following settings:

```
run("Subtract Background...", "rolling=500 light separate sliding disable");
// each RGB color was background corrected separately, which removed unwanted yellow or red imaging bias
```

7.13 General molecular methods

7.13.1 Phenol/Chloroform DNA extraction

From a non-/decrosslinked DNA-Protein mixture, e.g. after ChIP, DNA was extracted in the following way: An equal volume of cold phenol/chloroform/isoamylalcohol mix (25:24:1 pH 8.2) was added, vigorously vortexed and centrifuged at 15,000 g, 3'. The aqueous SN was extracted a second time with pure chloroform and taken for ethanol precipitation.

7.13.2 Ethanol precipitation of DNA

To the purified DNA sample 3 volumes of 100% EtOH, 0.1 vol. 3 M KOAc and 6 µl of 10 mg/ml glycogen were added, vortexed and incubated at -80 °C for 1h. The sample was centrifuged at 21,000 g for 30' at 4 °C and the SN replaced by 70% EtOH. After 10' centrifugation the SN was discarded, the pellet air dried and dissolved in 20µl H_2O.

For linearized plasmid precipitation in the RNA-probe preparation protocol, no glycogen was used and KOAc was replaced by 5 M NaOAc.

7.13.3 Agarose gel electrophoresis

A Biorad equipment set, including power supply, electrophoresis chambers, gel gasket, pocket combs and a gel pouring frame, was used according to the manufacturers guidelines. Agarose gels were prepared with 1-2 % agarose, depending on the size of the expected product. Therefore the agarose was dissolved with microwave heating in 1x TAE buffer, 3 drops of Ethidium bromide added, mixed and poured into shape. After solidification, the gel was placed in the electrophoresis chamber and loaded with a ladder and the samples, which were premixed with DNA-loading dye. After electrophoresis, the gel was imaged by a Biorad UV photodocumentation system.

In case of preparative agarose gels, the correct band was excised and DNA/RNA isolated according to manual by the peqGOLD gel extraction kit.

7.13.4 qPCR

qPCR was generally performed in triplicates with qPCR SYBR Green Master mix on the ViiA7 qPCR-cycler by the default settings with melt curve analysis for quality control. Standard reaction volume was 10 µl with varying input gDNA and cDNA concentrations. Primers were generally designed to span 70-200 bp and BLASTed for genome/transcriptome wide target specificity.

7.14 Skeletal staining

Adapted from "Manipulating the mouse embryo : a laboratory manual", Behringer et al., 2014, CSH Press[172].

This protocol was optimized for skeletal staining of E15.5 embryos. Embryos were extracted and washed in PBS. The tails were taken for genotyping, otherwise the skeletal tissue remain intact. The embryo was eviscerated and stomach, heart, lung and gut were processed for further use in WB, ChIP, RNA-analysis or microscopy. The remaining embryo was fixed for 3 days in 95 % EtOH, incubated ON in acetone and briefly washed in 95 % EtOH. For 3 days, cartilage was stained with alcian blue staining solution and briefly washed in 95 % EtOH. Rehydration was performed 3h-ON per step, twice with 70 % EtOH, and once with 40 % and 15 % each. Bones were stained with 150 mg/L alizarin red in 1 % KOH for at least 3 h, followed by clearing of the embryo in a 1 % KOH / Glycerol gradient: 80 / 20, 60 / 40, 40 / 60, 20 / 80 and twice 0 / 100, with at least 12 h incubation time per step.

7.15 Bioinformatic analysis

Initial QC of ChIP-seq data as well as mapping and annotation to the mm9 assembly genome was performed by the IMB bioinformatic core facilities of the IMB. Peaks were called with Galaxy software, using the MACS algorithm[173].

7.15.1 Gene expression clustering

RNA-seq data was previously annotated to 27 clusters based on 2-fold or higher changes in triplicate average gene expression values between each two adjacent developmental stages from E12,5 to AE. Correct clustering was confirmed by visual track inspection as well as RT-qPCR and RNA-ISH experiments.

7.15.2 ChIP-seq clustering

To assess epigenetic mark variation throughout the embryonic development and adult homeostasis, I clustered ChIP-seq data with MS Excel and Galaxy[174] software to 27 clusters based on 1.5-fold or higher changes ChIP to input enrichment between each two adjacent developmental stages from E12,5 to AE.

ChIP-seq data was restricted to Promoter ±1 kb annotated peaks to allow direct comparability with RNA-seq clusters by gene annotation. Duplicate entries, mostly TSS with two flanking called peaks, were reduced to the strongest peak value. Empty entries, where a peak was not called at a certain stage, were marked by value 1, which is clearly below the ChIP peak enrichment to input cutoff (>2) and hence allows detection of present absent changes during clustering with a 1.5-fold change threshold. Genes without any one detected promoter peak at any of the analyzed stages were not used in clustering.

As ChIP-seq replicates varied in enrichment efficiency between stages, especially in the H3K4me3 dataset, each developmental stages duplicate datasets were averaged and then normalized to the average ChIP to input enrichment.

To reduce cluster set complexity, 27-clustered ChIP-seq and RNA-seq data were reduced to the 9-cluster system, annotated A-I, disregarding the E12.5 to E14.5 change in gene expression or epigenetic TSS states.

7.15.3 Co-clustering analysis

To see if different epigenetic marks as well as gene expression, both further referred as trait, change simultaneously during development, clusters A-I of each were compared. As cluster frequencies have a strong effect on the chance that a gene will be found in the same cluster for two or more traits, a pairwise, cluster frequency- adjusted approach was used. Genes, cluster-annotated in both compared traits were counted separately for each cluster equally annotated (e.g. *Lgr5* annotated with cluster H for H3K4me3/K27Ac) as well as for different annotation (e.g. *Smarca2*, annotated with D/E for H3K4me3/K27Ac). The observed co-cluster size S_{obs} was than compared to theoretical size S_{rnd}, calculated from random cluster combinations with the observed single mark cluster frequencies ***F***:

cluster**:* A-I (see ***Fig. 32) or "different" for co-clustering

***trait**:* gene expression, H3K4me3, H3K27Ac or H2A.Z

***n**(cluster; trait):* number of annotated genes in a cluster for a single trait

***N**(trait):* number of all cluster annotated genes for a single trait

N_2*(trait1, trait2):* intersection of ***N**(trait1)* and ***N**(trait2)*

$$F(cluster;\ trait) = \frac{n(cluster;\ trait)}{N(trait)}$$

S_{obs}*(cluster; trait1; trait2):* annotated gene number in a co-cluster or "different" in two traits

$$S_{rnd}(cluster;\ trait1;\ trait2) = F(cluster;\ trait1) * F(cluster;\ trait2) * N_2(trait1; trait2)$$

$$S_{rnd}(different;\ trait1;\ trait2) = N_2(trait1; trait2) - \sum_{cluster=A}^{I} S_{rnd}(cluster;\ trait1; trait\ 2)$$

An S_{obs} greater than S_{rnd} indicates parallel clustering of two traits, while a lower number means anti-correlation of the two traits in the same cluster. $S_{obs} = S_{rnd}$ indicate that there is no correlation or anti-correlation between the compared traits' timecourses. To test, whether the calculated differences for each co-cluster and trait combination are significant (P<0.05), Z-score based P-values were calculated:

$$Z(cluster;\ trait1;\ trait2) = \frac{S_{obs} - S_{rnd}}{\sqrt{(S_{obs} + S_{rnd}) * (1 - \frac{S_{obs} + S_{rnd}}{2 * N_2})}}$$

$$P(cluster;\ trait1;\ trait2) = (NORM.S.VERT(-ABS(Z); 1)) * 2$$

//MS EXCEL formula, two sided test, normal distribution assumed

8 Material

Abbreviations: EP: Eppendorf, TF: Thermo Fisher Scientific, NEB: New England Biolabs

8.1 Equipment

+4° cold room (Wiessmann), +4° fridge (Liebherr Mediline), 10 cm Petri dishes for cell culture (Falcon), 10/20/200/1000 µl Pipetman (Gilson), -150° freezer (Sanyo), -20° freezer (Liebherr Mediline), 6/24-well plates (Falcon), -80° freezer (Sanyo), agarose gel electrophoresis system (Biorad), AriaII-SORP Cellsorter (BD), Bioanalyzer 2100 (Agilent), Bio-Dot Dot Blot vacuum gasket (Biorad), Bioruptor Plus (Diagenode), cell culture fluorescence microscope DM IL LED (Leica), cell culture incubators (TF), centrifuge MultifugeX3 < 3.700 g (TF), centrifuge tubes 15/50 ml (Nerbe plus), Chemidoc XRS+ UV photodocumentation system (Biorad), colorless nail polish (Essence), confocal microscope SP5 (Leica), counting chamber (Neubauer), coverslips 24x60 mm (Menzel), cryotome CS 3050 (Leica), DM 2500 microscope (Leica), EASYstrainer cell strainer 70 µm (Greiner Bio-One), Eppendorf tubes(EP), FACS tube with cell strainer 40 µm (Falcon), flattening table HI1220 (Leica), Frame-Seal incubation chambers 9x9 mm SLF0201 (Biorad), glassware (Schott Duran), grease pen (Dako), HiSeq 2500 sequencing system (Illumina), hybridization oven OV3 (Biometra), IBIDI chambers (IBIDI), laminar flow hoods BSC-SG403 EN (Dometic), lo-bind Eppendorf tubes (EP), LRS Fortessa flow cytometer (BD), M80 binoculars (Leica), magnetic rack for Dynabeads (self-made), magnetic stirrer/heating plate (Heidolph), microtome RM2255 (Leica), microwave oven (AEG), mini-centrifuge Sprout (NeoLab), Multipette M4 (EP), Nanodrop 2000 spectrophotometer (TF), Nanodrop 3300 spectrophotometer (TF), overhead wheel (Kisker Biotech), PCR thermocycler (Biometra), pH-meter seven compact (Mettler-Toledo), pipet boy (Integra), pipet tips 10/200/1000 µL (Nerbe plus), Polytron homogenizer (Kinematica), qPCR-cycler ViiA7 (Life Technologies), qPCR-plates 384-well (Applied Biosystems), Qubit 2.0 (Life Technologies), scale ED822-OCE (Sartorius), SDS-PAGE chambers (Biorad), semidry blotting apparatus TransBlot SD (Biorad), serological pipets 5/10/25 ml (Falcon), shaker for bacterial culture (Infors Multitron), ST5 Cat Shaker (NeoLab), Superfrost Ultra Plus object slides (Menzel), tabletop centrifuge Fresco < 21,000 g (Heraeus), tabletop centrifuge Pico < 21,000 g (Heraeus), Thermomixer comfort (EP), ultrapure water purification system (Millipore), Vortex-Genie 2 (Scientific Industries), water bath HI 1210 (Leica), water bath (GFL), Whatman blotting paper (Biorad), widefield fluorescent microscope AF7000 (Leica).

8.2 Reagents

8.2.1 Chemicals

$(NH_4)_2SO_4$ (SA, # A4418), acetic acid (SA, # A6283), acetic anhydride (SA, # 45830), acetone (Roth, # 7328), acrylamide 30% ProtoGel (National Diagnostics, # EC-890), agarose (AC, # A2114), alcian Blue 8GX (SA, # 5500), alizarin red S (SA, # A5533), ampicillin sodium salt (100 µg/ml final) (AC, # A0839), APS (ammonium persulfate) (SA, # A3678), blocking grade milk (Biorad, # 170-6404), Boehringer blocking reagent (Roche, # 1096176), BSA (SA, # A7906), $CaCl_2$ (SA, # C5670), chloroform (Roth, # 3313.2), citric acid (SA, # 251275), DAPI (laboratory stock), Denhardt solution (SA, # D2532), dextran sulfate (SA, # D8906), DMEM 4.5 g/l glucose, w/o glutamine, with phenol red (PAA, # 21969-035), DMSO (SA, # 41640), donkey serum (laboratory stock), Duolink In Situ Mounting Medium (SA, # DUO82040), EDTA (AC, # A1103), EGTA (AC, # A0878), ethanol (Roth, # K928.1), ethanol pure (SA, # 32205), ethidium bromide (AC, # A2273), FBS Gold (PAA, # A15-151), formaldehyde 37% stabilized with methanol (SA, # 252549), formamide (Roth, # 6749.2), glutamate (PAA, # 25030-024), glycerol (SA, # G5516), glycine (SA, # G8898), glycogen from bovine liver, Type IX (SA, # G0885), goat serum (SA, # G9023), H_2O_2 30 % (SA, # 95302), HCl (SA, # 4625.1), heparin (SA, # H4784), HEPES (SA, # H3375), hygromycinB 50-100 µg/ml (Invitrogen, # 10687-010), IGEPAL-CA630 (SA, # I8896), Immumount (Thermo Scientific, # 9990402), IPTG (AC, # A1008), isopropanol (Roth, # 6752.2), KCl (SA, # P9541), KH_2PO_4 (SA, # P9791), KOAc (SA, # P1190), LB-Agar (SA, # L2897), LiCl (laboratory stock), maleic acid (SA, # 63180), methanol (Roth, # 4627.2), $MgCl_2$ (SA, # M2670), Na deoxycholate

(SA, # D6750), Na_2HPO_4 x $2H_2O$ (SA, # S7907), Na_3Citrate (SA, # C8532), NaButyrate (SA, # 303410), NaCl (SA, # 31434), NaOH (SA, # S8045), Nonidet P40 Substitute (NP-40) (SA, # 74385), O.C.T. cryosection medium (Tissue-Tek, # 4583), paraformaldehyde (SA, # 158127), Paraplast (Surgipath, # 39601006), penicillin/streptomycin supplement (PAA, # 15140-122), periodic acid solution (SA, # 395132), phenol/chloroform/isoamylalcohol for DNA, pH 8.2 (Invitrogen, # 15593-031), PIC (protease inhibitor cocktail) (Roche, # 4693132001), PMSF (SA, # 78830), Ponceau red (SA, # P3504), Roti-Histokitt II (Roth, # T160.1), Schiff's solution (SA, # 3952016), SDS (SA, # L4390), spectinomycin (50 µg/ml final) (AC, # A3834), sucrose (SA, # 84097), TEMED (SA, # T7024), TRIS-base (TRIS ultrapure) (AC, # 1086), Triton X-100 (SA, # T8787), tRNA (laboratory stock), trypsin (PAA, # 15400-054), Tween20 (SA, # P7949), X-gal (AC, # A1007), xylol (Roth, # 4436.2).

AC: AppliChem, SA: Sigma-Aldrich

8.2.2 Enzymes and supplements

2x Rapid Ligation Buffer, T4 DNA Ligase (Promega, #C671A), 5x Transcription buffer (for RNA-pol) (Promega, #P11813), Antarctic phosphatase (NEB, #M0289), ApaLI (NEB, #R0507), BCIP (Roche, #11383221001), BSAI-HF (NEB, #R3535), collagenase (SA, #C7657), Colorplus prestained protein marker (NEB, #P7711), Control Insert DNA (pGEM-T cloning) (Promega, #A363A), DIG RNA Labeling Mix (Roche, #11277073910), DIG-11-dUTP 1 mM (Roche, #11093088910), Dispase 100 mg/ml (SA, #D4693), DNA-loading dye 6x (NEB, #B7024), DNAseI (Qiagen, #148037682), dNTP-Mix (TFS, #R0181), DTT (Promega, #P117B), EcoRI HF (NEB, #R3101), Esp3I (TFS, #ER0451), Gene ruler 1 kb plus DNA ladder (TFS, #SM1331), HindIII-HF (NEB, #R3104), NBT (Roche, #11383213001), NcoI-HF (NEB, #R3193), NheI-HF (NEB, #R3131), NotI-HF (NEB, #R3189), Onetaq-polymerase (NEB, #M0480), Plasmid-Safe nuclease (Epicentre, #E3101K), PmlI (NEB, #R0532), Proteinase K (SA, #P6556), qPCR SYBR Green Master mix (TFS, #4368706), RNAse A (Qiagen, #1007885), RNAsin (Promega, #N251B), SP6 RNA-polymerase (NEB, #M0207), SpeI-HF (NEB, #R3133), T4 DNA ligase (NEB, #M0202), T4 DNA ligase (Promega, #M180A), T7 RNA-polymerase (Promega, #P207B), Taq-polymerase (self-made laboratory stock), Thermopol reaction buffer 10x (NEB, #B9004), XhoI (NEB, #R0146)

8.2.3 Buffers

Alcian blue staining solution (1 vol. 2.5 mg/ml alcian blue in H_2O, 20 vol. 100% EtOH, 5 vol. 100% AcOH), blocking buffer (5 % Blocking grade Milk in 1x PBS), blotting buffer (192 mM glycine, 25 mM Tris-base, 20 % Methanol), Boehringer blocking solution 10x (10 % Boehringer blocking reagent in 100 mM maleic acid, 150 mM NaCl, pH 7.5 by NaOH), cell lysis buffer (FLAG-ChIP) (0.5% NP40, 10 mM Tris-Cl pH 8.1, 1 mM EDTA, freshly supplemented with 1x PIC and 1mM PMSF), elution buffer (histone-ChIP) (20 mM TRIS-base pH 7.5, 50 mM NaCl, 5 mM EDTA, 20 mM NaButyrate if for H3K27Acetyl-ChIP), elution buffer (IP) (20 mM Tris-Cl pH 7.5, 5 mM EDTA, 50 mM NaCl), HeLa lysis buffer (1% SDS, 50 mM Tris-Cl pH 8.0, 10 mM EDTA, freshly added PIC 1x and PMSF 1mM), HEPES buffer (IP) (50 mM HEPES pH 7.5, 500 mM NaCl, 1% TritonX, freshly supplemented with 1x PIC and 1mM PMSF), high salt wash buffer (FLAG-ChIP) (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM EDTA, 1% TritonX, 0.1% SDS), Laemmli Buffer 1x (192 mM glycine, 25 mM Tris, 0.1% SDS), LB-Agar (LB-Luria + 1.5% Agar-Agar provided by CF), LB-Lennox (0.5 % NaCl, 1% Bacto-Tryptone, 0.5% Bacto-Yeast extract, pH 7,0 adjusted with NaOH provided by CF), LB-Luria (1 % NaCl, 1% Bacto-Tryptone, 0.5% Bacto-Yeast extract, pH 7,0 adjusted with NaOH provided by CF), LiCl wash buffer (FLAG-ChIP) (250 mM LiCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA, 0.5% IGEPAL-CA630, 0.5% Na-deoxycholate), low salt wash buffer (FLAG-ChIP) (150 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM EDTA, 1% TritonX, 0.1% SDS), lower SDS-gel buffer 4x (1.5M Tris-Cl pH 8.8, 0.4 % SDS), lysis buffer (histone-ChIP) (1% SDS, 50 mM Tris-Cl pH 8.0, 10 mM EDTA, freshly supplemented with 1x PIC and 1mM PMSF, 20 mM NaButyrate if for H3K27Acetyl-ChIP), Miniprep P1 (50 mM Tris-Cl pH 8.0, 10 mM EDTA), Miniprep P2 (200 mM NaOH, 1% SDS), Miniprep P3 (3M KOAc, pH 5.5 adjusted by AcOH), NP buffer (10 mM

TRIS-base pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % NP40, 0.1% Na-deoxycholate, 0.1 % SDS, freshly supplemented with 1x PIC and 1mM PMSF), NTMT (50 mM MgCl2, 100 mM NaCl, 100 mM Tris-Cl pH 9.0-9.5, 0.1 % Tween20), nuclear lysis buffer (FLAG-ChIP) (0.5 M NaCl, 10 mM Tris-Cl pH 8.1, 1 mM EDTA, 1% TritonX 0.5% Na-deoxycholate, freshly supplemented with 1x PIC and 1mM PMSF), PBS 10x (80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4 x $2H_2O$, 2.4 g KH_2PO_4 in 1 L H_2O. pH adjusted to 7.4 by HCl), PK buffer (PLA) (0.05 M Tris-Cl, 2 mM $CaCl_2$, pH 7.8, 0.01 M EDTA, 0.01 M NaCl, add 0.02 µg/µl PK freshly), PLA hybridization buffer (2x SSC, 50 % formamide, 10 % dextran sulfate, store at -20 °C), PLA WashB-A (0.15 M NaCl, 10 mM TRIS-base, 0.05 % Tween20, HCl to pH 7.4, 4°C long term storage, use at RT), PLA WashB-B (0.1 M NaCl, 250 mM TRIS-base, HCl to pH 7.5, 4°C long term storage, use at RT), PLA WashB-L1 (2x SSC, 50 % formamide), PLA WashB-L2 (4x SSC, 0.1 % Tween20), Ponceau red staining solution (0.5% Ponceau red, 1 % Acetic acid), RIPA buffer (10 mM TRIS-base pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% TritonX, 0.1% Na-deoxycholate, 0.1 % SDS, freshly supplemented with 1x PIC and 1mM PMSF, 20 mM NaButyrate if for H3K27Acetyl-ChIP), RLT-buffer (from the Qiagen RNeasy kit), RNA-ISH blocking solution (1x Boehringer blocking solution, 10 % goat serum, TBSX), RNA-ISH hybridization buffer (5 x SSC, 50 % Formamide, 10 % Dextran sulfate, 1 x Denhardt, 100 µg/ml Heparin, 100 µg/ml tRNA boiled at 95 °C for 5', 5 mM EDTA, adjusted with citric acid to pH 4.5-5.0), SDS-lysis buffer (WB, IP) (50 mM Tris-Cl pH 8.0, 10 mM EDTA, 1% SDS, freshly added PIC 1x, 1mM PMSF), SDS-PAGE loading buffer 4x (4 % SDS, 240 mM pH 6,8 Tris, 40 % glycerol, 20 % ß-mercaptoethanol, 0,04 % bromphenol blue), sonication buffer (FLAG-ChIP) (0.1% SDS, 10 mM Tris-Cl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate), SSC 20x pH 4.5 (3 M NaCl, 0.3 M Na_3Citrate, pH adjusted to 4.5 by citric acid), SSC 20x pH 7 (3 M NaCl, 0.3 M Na_3Citrate, pH adjusted to 7.0 by HCl), TAE buffer 50x (50 mM EDTA, 1 M acetic acid, 2 M TRIS-base, pH 8.3), tail buffer (20mM Tris-HCl pH8.8, 10mM $(NH_4)_2SO_4$, 10mM KCl, 10mM EDTA, 0.1% Triton X-100), TBS (2 mM KCl, 150 mM NaCl, 100 mM Tris-Cl pH 7.5), TBSX (TBS, 0.1 % TritonX), TE-buffer (FLAG-ChIP) (10 mM Tris-Cl pH 8.0, 1 mM EDTA), TE-buffer (H3K27me3 final protocol) (20 mM Tris-Cl pH 8.0, 1 mM EDTA), TE-buffer (IP, histone-ChIP) (10 mM Tris-Cl pH 8.0, 10 mM EDTA), upper SDS-gel buffer 4x (0.5 M Tris-Cl pH 6.8, 0.4 % SDS)

8.2.4 Kits

Alexa Fluor® 594 ULS™ labeling reagent (TFS, #U21654), Duolink In Situ Detection Reagents FAR RED (SA, #DUO92013), Duolink In Situ Detection Reagents RED (SA, #DUO92008), Golden Gate TALEN and TAL Effector Kit 2.0 (Addgene, #1000000024), High sensitivity DNA chip and reagents (Agilent, #5067-4626), Illustra MicroSpin G-50 Columns (GE healthcare life sciences, #27-5330-01), Longarm Photoprobe Biotin reagent (Vector laboratories, #SP-1020), peqGOLD gel extraction kit (Peqlab, #12-2501-02), peqGOLD Xchange Plasmid Mid Kit (Peqlab, #12-7401-02), PLA probe 5x antiGoat PLUS (SA, #DUO92003), PLA probe 5x antiRabbit MINUS (SA, #DUO92005), PLA probe 5x antiRabbit PLUS (SA, #DUO92002), Qubit dsDNA HS Assay-Kit (TFS, #Q32851), Qubit RNA HS Assay-Kit (TFS, #Q32852), Superscript II reverse transcriptase kit (TFS, #18064014), SuperSignal West Pico Chemiluminescent Substrate (TFS, #34087), Biozol Vectastain ABC kit (Biozol, #PK6101).

8.2.5 Antibodies

antibody	host	conjugate	company	cat. no.	dilution
biotin	goat		Biozol	SP3000	1:100 for IF by PLA-protocol
CD24	rat	PE-Cy7	Sigma-Aldrich	560536	1:200 for FACS and IF
CD31	rat	PE	BD Pharmingen	561073	1:400 for FACS
CD326 (Epcam)	rat	PerCP eFluor710	eBioscience	46-579182	1:200 - 1:100 for FACS, 1:1000 IF
CD326 (Epcam)	rat	APC	eBioscience	17-579183	1:200 - 1:100 for FACS, 1:1000 IF
CD45	rat	PE	BD Pharmingen	553081	3.5:1000 for FACS
Cre	rabbit		Novagen	69050-3	1:1500 for PLA
DIG	goat		Biozol	MB7000	1:1000 for PLA
DIG AP		AP	Roche	11093274910	1:1000 for RNA-ISH
Dynabeads G		Protein G	LifeTechnologies	10003	
EzView RED A affinity gel		Protein A	Sigma-Aldrich	P6486	
EzView RED G affinity gel		Protein G	Sigma-Aldrich	E3403	
FLAG-Ab	mouse		Sigma-Aldrich	F3165	1:2000 for WB, 1:500 for IF / confocal
GFP	rabbit		LifeTechnologies	A11122	1:1000-1:2000 for IF
goat AF647	donkey	AF647	LifeTechnologies	A21447	1:1000 for IF by PLA-protocol
H2A.Z	rabbit		Active Motif	39113	1:300-1:1000 for IF, 0.25 µl / 20 µl beads for ChIP
H3K27ac	rabbit		Abcam	4729	1:1000 for PLA and IF by PLA-protocol, 0.5 µl / 20 µl beads for ChIP
H3K27me3	rabbit		Millipore	17-622	2 µl / 20 µl beads for NP-buffer ChIP
H3K4me2	mouse		Millipore	05-1338	1:1000 for IF by PLA-protocol
H3K4me3	rabbit		Millipore	CS200580	1:500 for IF by PLA-protocol
H3K4me3	rabbit		Abcam	8580	1:50-1:250 for IF by PLA-protocol
mouse AF488		AF488	Invitrogen	A10684	1:1000 for IF by PLA-protocol
mouse AF594		AF594	LifeTechnologies		1:1000 for IF / confocal
mouse POD	goat	POD	Sigma-Aldrich	A9917	1:3000 for WB
pan-histone	rabbit		Abcam	1791	1:500 for PLA
rabbit AF488	goat	AF488	LifeTechnologies	A11070	1:300 for IF
rabbit AF594		AF594	Invitrogen	A11072	1:1000 for IF by PLA-protocol
rabbit biotinylated	goat	biotin	Biozol Vectastain ABC kit	PK6101	1:200 for IF by PLA-protocol
rat AF568	goat	AF568	Invitrogen	A11077	1:300 for IF
Streptavidin-AF488		AF488	Invitrogen	S32354	1:1000 for IF
Streptavidin-HRP		HRP	Perkin Elmer	FP1047	1:500 for Dot Blot
UEA1		FITC	Sigma-Aldrich	L9006	1:200 for FACS, 1:200-1:1000 for IF
UEA1		TRITC	Sigma-Aldrich	L4889	1:200 for FACS, 1:1000 for IF

8.3 Oligonucleotides

ChIP-qPCR Primer	**sequence**	**target**
For-Actg1	ctgagtgagtgagtagcaggggc	*Actg1*
Rev-Actg1	tgcagagtgtgggtgtggga	*Actg1*
414 Arid5b full qPCR FOR	ctccccctcagcaccaaaat	*Arid5b* full
415 Arid5b full qPCR Rev	ccatctttggcagcaaaccc	*Arid5b* full
416 Arid5b short qPCR FOR	tccactccagaaaacctgcc	*Arid5b* short
417 Arid5b short qPCR Rev	tggcgactggaatgggtaac	*Arid5b* short
axin25mc_fwr	tggtttgtgcctggttgagt	*Axin25mc*
axin25mc_rev	accagcaggtcacttatgcc	*Axin25mc*
foxa15mc_fwr	cagtccagtccacgctaagg	*Foxa15mc*
foxa15mc_rev	gcaaccctggaagcaaaagg	*Foxa15mc*
Foxg1 311F	acctgctttatttcgggactgtt	*Foxg1*
Foxg1 361R	agagctgccccgggc	*Foxg1*
626 Galr2 qPCR Fwr	gtcatcgtggcggtactctt	*Galr2*
627 Galr2 qPCR rev	ggaaagcgaccaaaccacac	*Galr2*
For-GAPDH	agtcctagctcaagggcgca	*GAPDH*
Rev-GAPDH	gctactcgcggctttacggg	*GAPDH*
410 Hnf4a full qPCR FOR	cagtggcgagtccttatggt	*Hnf4a* full
411 Hnf4a full qPCR Rev	acctgaccgtggtcctacc	*Hnf4a* full
412 Hnf4a short qPCR FOR	ggggagaatgcgactctctaa	*Hnf4a* short
413 Hnf4a short qPCR Rev	ttcaaactccagggtggtgtag	*Hnf4a* short
622 Hoxa3 qPCR Fwr	tgaaggctaagtgggggtact	*Hoxa3*
623 Hoxa3 qPCR rev	cctaccctccaattgatcccg	*Hoxa3*
624 Hoxb7 qPCR Fwr	gctcgaaccgagttccttcaa	*Hoxb7*
625 Hoxb7 qPCR rev	caggggtagatccggaagtt	*Hoxb7*
Hoxd13 217F	cccctcccaacctgaacttc	*Hoxd13*
Hoxd13 267R	ggttagctcatcgcgggac	*Hoxd13*
h17_38468200_fwr	agggccttcccctagggctg	intergenic 11/12
h17_38468200_rev	aggctactgtgcccccaccc	intergenic 11/12
h2_234119650_fwr	gcagttttgtgccccacggc	intergenic 7/8
h2_234119650_rev	gccgaatctcgcagggacagg	intergenic 7/8
h3_195348800_fwr	cgctggccatgatccagtccaaa	intergenic 9/10
h3_195348800_rev	gcagcaggaagccctgggtc	intergenic 9/10
kcnq15mc_fwr	ggaggtggttgattcccctg	*kcnq15mc*
kcnq15mc_rev	aggcccatcaccacattctg	*kcnq15mc*
577 Lbh qPCR Fwr	ggtgagtatcctgcgccttt	*Lbh*
578 Lbh qPCR rev	ttaacccgcactgggtgtc	*Lbh*
Lhx5 549F	gaacgtactagaccgcgcgt	*Lhx5*
Lhx5 599R	cactcgcagcattgaacaca	*Lhx5*
mdk5mc_fwr	tcctgagcccatcctatccc	*Mdk5mc*
mdk5mc_rev	gcttcttccttctcgccctt	*Mdk5mc*
meis15mc_fwr	aggctctccgctcctatctc	*Meis15mc*

meis15mc_rev	cttcccgggcatatctggtc	*Meis15mc*
Nbs1 118R	cgggacgtgcgcgtt	*Nbs1*
Nbs1 68F	ttcccgcctttccgaag	*Nbs1*
olfm45mc_fwr	aagccagaggaatggctgtc	*Olfm45mc*
olfm45mc_rev	agccactagccagcaacttt	*Olfm45mc*
prelp5mc_fwr	cggaagtcctctttcccacc	*Prelp5mc*
prelp5mc_rev	cgctggaggagtccaaacat	*Prelp5mc*
rnf435mc_fwr	gctctgtggttgggaactga	*Rnf435mc*
rnf435mc_rev	gcatttttgggtgctcctgg	*Rnf435mc*
628 Sdk2 qPCR Fwr	taaagtgccccctctttgcg	*Sdk2*
629 Sdk2 qPCR rev	cgctttacggaatagccagc	*Sdk2*
sema75mc_fwr	agagaaagacagtggcaggc	*Sema75mc*
sema75mc_rev	tgatggggaggtcctagtgg	*Sema75mc*
serpinh5mc_fwr	ggccgttagaccatacaggg	*Serpinh5mc*
serpinh5mc_rev	cattcctgagcctccagacg	*Serpinh5mc*
Shh enh 232F	ttcacccagcagaaccagact	*Shh* enhancer
Shh enh 663R	aattagttgcactgaccaggtgg	*Shh* enhancer
shh5mc_fwr	gggaaaagcttgcgatgtgg	*Shh5mc*
shh5mc_rev	tccatcctaccctcgctctt	*Shh5mc*
Six6 116F	tccagctgcccattttgaat	*Six6*
Six6 166R	cacataccccggctacttgct	*Six6*
418 Smarca2 full qPCR FOR	gccccggattatgcatcctt	*Smarca2* full
419 Smarca2 full qPCR Rev	cccgcctgtttgttttggtc	*Smarca2* full
420 Smarca2 short qPCR FOR	ggacccgaaagagtgcagtg	*Smarca2* short
421 Smarca2 short qPCR Rev	ggtgaggcacaactacaggg	*Smarca2* short
Stag1 378F	attggcgtgtggaaaatgc	*Stag1*
Stag1 429R	aacggagctgcaatccta	*Stag1*
365 test TALE 13 Fwr	gtcctgagccgaaacctacc	TALE #13
366 test TALE 13 Rev	ttgtcttctccgcacgactg	TALE #13
375 test TALE 18 Fwr	tcttgggatggggagtctgt	TALE #18
376 test TALE 18 Rev	tcgtgtgacaaggccatgag	TALE #18
579 Wnt3 qPCR Fwr	ttccaactattgggggcgtc	*Wnt3*
580 Wnt3 qPCR rev	gtggggctccattagaagca	*Wnt3*
575 Xbp qPCR Fwr	gcgtagacgtttcctggcta	*Xbp*
576 Xbp qPCR rev	taagagtagcactttgggggc	*Xbp*

DNA ISH-probe primer for PLA	sequence	target + probe size
332 Hnf4a full FOR	caatccccagccacaggttc	218 bp *Hnf4a* full
333 Hnf4a full Rev	catgaccaagggtacccagc	218 bp *Hnf4a* full
334 Hnf4a short FOR	tagtgagcttcctccgggtg	219 bp *Hnf4a* short
335 Hnf4a short Rev	cacaccgctatgttggcctt	219 bp *Hnf4a* short
457 Epcam Pr8 for	ctcagagaggtgaggcgtgg	222 bp *Epcam*
458 Epcam Pr8 rev	caagttggagtcgccagacc	222 bp *Epcam*
459 Shh Pr9 for	tcttcccttcatatctgccgc	229 bp *Shh*
460 Shh Pr9 for	gctcacaagtcctcaggttcc	229 bp *Shh*
338 Arid5b short FOR	acgcagctaaggcaacactt	241 bp *Arid5b* short
339 Arid5b short Rev	tcaggcagatgtttggcgac	241 bp *Arid5b* short
336 Arid5b full FOR	ggattgcatcaggtcctggc	245 bp *Arid5b* full
337 Arid5b full Rev	atccattctcttcccgccga	245 bp *Arid5b* full
340 Smarca2 full FOR	tctcttctctccacggctcc	267 bp *Smarca2* full
341 Smarca2 full Rev	tagaacggcgaggttcggtt	267 bp *Smarca2* full
342 Smarca2 short FOR	gctgtgatgggggatgtcac	269 bp *Smarca2* short
343 Smarca2 short Rev	gtagcggtgctgacaaggac	269 bp *Smarca2* short

Genotyping Primers	sequence	application	target
GFP1_for	atcatggccgacaagcagaa	gDNA-qPCR	*GFP*
GFP1_rev	ctcgatgttgtggcggatct	gDNA-qPCR	*GFP*
TL_fwr	gcctcccggatccgatgt	gDNA-qPCR	*lsd1-TALE*
TL_rev	tgcttcctgagaggtcattcg	gDNA-qPCR	*lsd1-TALE*
ndufaf3 259F	cctccagcgacagtccctaa	gDNA-qPCR	*ndufaf3*
ndufaf3309R	ccgtactagcaggtgtttttccc	gDNA-qPCR	*ndufaf3*
CMV enh_prom FOR	ccaagtacgccccctattga	PCR	CMV enh_prom 257
CMV enh_prom rev	agtcaaaccgctatccacgc	PCR	CMV enh_prom 258
GFP_Fwr	aggacgacggcaactacaag	PCR	*GFP* embryo
GFP_Rev	gcccgggattctcttctacg	PCR	*GFP* embryo
Hprt-cre mut fwr	gcggtctggcagtaaaaactatc	PCR	*Hprt-Cre*
Hprt-cre mut rev	gtgaaacagcattgctgtcactt	PCR	*Hprt-Cre*
Hprt-cre wt fwr	cacagtagctcttcagtctgataaaa	PCR	*Hprt-Cre*
Hprt-cre wt rev	tttctataggactgaaagacttgctc	PCR	*Hprt-Cre*
JK_neo_stop FOR	tgggcacaacagacaatcgg	PCR	JK_neo_stop 254
JK_neo_stop rev	gatgcgatgtttcgcttggt	PCR	JK_neo_stop 255
For-Lgr5-egfp-WT-com	ctgctctctgctcccagtct	PCR	*Lgr5-GFP*
Rev-Lgr5-eGFP	gaacttcagggtcagcttgc	PCR	*Lgr5-eGFP*
SP6-primer	catttaggtgacactatag	Sanger-sequencing	pGEM-T insert
T7-primer	taatacgactcactataggg	Sanger-sequencing	pGEM-T insert
pCR8_F1	ttgatgcctggcagttccct	Sanger-sequencing	1st TALE-assembly
pCR8_R1	cgaaccgaacaggcttatgt	Sanger-sequencing	1st TALE-assembly
SeqTALEN 5-1	catcgcgcaatgcactgac	Sanger-sequencing	2nd TALE-assembly
TAL_R2	ggcgacgaggtggtcgttgg	Sanger-sequencing	2nd TALE-assembly

8.4 Plasmids

plasmid	type	details
RP23-118H24	BAC	BACPAC DH10B *E.coli* clone
pTAL5 (and TALE assembly vectors)		from Golden Gate TALEN and TAL Effector Kit 2.0
pGEM-T	cloning vector	by Promega
pcDNA 3.1	eukaryotic	by Thermo Fisher Scientific CMV enhancer, NeomycinR, LacZ
pBlue Axin2	RNA ISH-probe	
pcDNA 3.1 hygro	eukaryotic	TALE-expression vector, HygroR
pcDNA-egfp-P2A-Flag3x	eukaryotic	as pcDNA3.1 and eGFP, P2A, FLAG3x
pGEM-T Lgr5	RNA ISH-probe	
pGEM-T Picalm	RNA ISH-probe	
pGEM-T Sfrp5	RNA ISH-probe	
pGEM-T Shh	RNA ISH-probe	
pGEM-T-easy-Lsd1		source for *Lsd1* sequence
pJKPr1	DNA-probe for PLA	pGEM-T, contains *Hnf4a* full probe
pJKPr2	DNA-probe for PLA	pGEM-T, contains *Hnf4a* short probe
pJKPr3	DNA-probe for PLA	pGEM-T, contains *Arid5b* full probe
pJKPr4	DNA-probe for PLA	pGEM-T, contains *Arid5b* short probe
pJKPr5	DNA-probe for PLA	pGEM-T, contains *Smarca2* full probe
pJKPr6	DNA-probe for PLA	pGEM-T, contains *Smarca2* short probe
pJKTALE#10	eukaryotic	TALE targets *MYH9*
pJKTALE#11	eukaryotic	TALE targets *MYH9*
pJKTALE#12	eukaryotic	TALE targets *MYH9*
pJKTALE#13	eukaryotic	TALE targets *LDHB*
pJKTALE#14	eukaryotic	TALE targets *LDHB*
pJKTALE#15	eukaryotic	TALE targets *LDHB*
pJKTALE#16	eukaryotic	TALE targets *ACTB*
pJKTALE#17	eukaryotic	TALE targets *ACTB*
pJKTALE#18	eukaryotic	TALE targets *ACTB*
pSlax eGFP-P2A		source for *eGFP-P2A* sequence

8.5 Software

software	application
2100 Expert	Agilent Bioanalyzer readout
CLC Sequence Viewer 6.7.1	in silico plasmid assembly, in silico restriction analysis, evaluation of Sanger sequencing data
Endnote X5	bibliography
FijI/ImageJ	image processing, qualitative and quantitative analysis
FinchTV 1.4.0	Sanger sequencing quality assessment
Galaxy (usegalaxy.org)	processing of annotated ChIP-seq and RNA-seq data, peak calling
Gene Cluster 3.0	k-means clustering of ChIP-seq, MBD-seq and RNA-seq data, processing of datasets for visualization with Treeview
ImageLab 3.0	gel electrophoresis imaging and analysis, WB/Dot blot imaging and analysis
Java Treeview	visualization of gene clusters
Leica imaging software	AF7000 operation, initial image processing and assessment
MS Excel 2010	ChIP/RNA/MBD-seq QC, numerical troubleshooting and analysis, qPCR analysis, graphical processing, text processing, mathematical data analysis, ChIP-seq clustering
MS PowerPoint 2010	graphical processing
MS Word 2010	text processing, sequence analysis and in silico cloning
NCBI Primer-BLAST	PCR and qPCR primer design
Pyrat	monitoring and coordination of animal stock and breeding progress
UCSC browser	visual track analysis and graphical processing

9. Abbreviations

'	minute	**kbp**	1,000 base pairs
''	second	**kDa**	1,000 Dalton
3D	3-dimensional	**Lsd1**	lysine specific demethylase 1
AE	adult enterocyte	**MBD**	methyl binding domain
AF	AlexaFluor fluorescent Ab conjugate	**Mbp**	1,000,000 base pairs
AP	alkaline phosphatase	**mRNA**	messenger RNA
BAC	bacterial artificial chromosome	**n.s.**	not significant
BIO	biotin	**neg. ctrl.**	negative control
BMP	bone morphogenetic protein	**NGS**	next generation sequencing
bp	base pairs	**NLS**	nuclear localization sequence
cDNA	reversely transcribed RNA	**ON**	overnight
ChIP	chromatin immunoprecipitation	**PAGE**	polyacrylamide gel electrophoresis
DIG	digoxigenin	**PCR**	polymerase chain reaction
E12.5/14.5	day 12.5/14.5 after gestation	**PLA**	proximity ligation assay
F0	founder generation	**pos. ctrl.**	positive control
F0f3-6	female founder no. 3-6	**qPCR**	quantitative PCR
F0m1/2	male founder no. 1 and 2	**rpkm**	reads per kilobase per million mapped reads
F1/2	first/second offspring generation	**rpm**	rotations per minute
FISH	fluorescent ISH	**RT**	room temperature (~22-23°C)
FLAG	DYKDDDDK-peptide	**RT-qPCR**	reverse transcriptase qPCR
FSC-A	forward scatter area	**SC**	stem cell
gDNA	genomic DNA	**SILAC**	stable isotope labeling by amino acids in cell culture
HH	hedgehog	**SN**	supernatant
HRP	horse radish peroxidase	**SSC-A/W**	side scatter area/width
IP	immunoprecipitation	**TA cell**	transit-amplifying cell
ISC	adult intestinal stem cell	**TALE**	transcription activator-like effector
ISH	*in situ* hybridization	**WB**	Western blot

10. Literature

1 Trapnell, C. Defining cell types and states with single-cell genomics. *Genome research* **25**, 1491-1498, doi:10.1101/gr.190595.115 (2015).

2 Wolpert, L. *Principles of development*. 3rd edn, (Oxford University Press, 2007).

3 Alberts, B., Wilson, J. H. & Hunt, T. *Molecular biology of the cell*. 5th edn, (Garland Science, 2008).

4 Grewal, S. I. & Jia, S. Heterochromatin revisited. *Nature reviews. Genetics* **8**, 35-46, doi:10.1038/nrg2008 (2007).

5 Green, M. R. Eukaryotic transcription activation: right on target. *Molecular cell* **18**, 399-402, doi:10.1016/j.molcel.2005.04.017 (2005).

6 Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23**, 4051-4060, doi:10.1038/sj.emboj.7600385 (2004).

7 Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7**, 29-59, doi:10.1146/annurev.genom.7.080505.115623 (2006).

8 Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**, 1413-1415, doi:10.1038/ng.259 (2008).

9 Ben-Dov, C., Hartmann, B., Lundgren, J. & Valcarcel, J. Genome-wide analysis of alternative pre-mRNA splicing. *The Journal of biological chemistry* **283**, 1229-1233, doi:10.1074/jbc.R700033200 (2008).

10 Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47, doi:10.1016/j.cell.2006.06.023 (2006).

11 Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes & development* **23**, 781-783, doi:10.1101/gad.1787609 (2009).

12 Deans, C. & Maggert, K. A. What do you mean, "epigenetic"? *Genetics* **199**, 887-896, doi:10.1534/genetics.114.173492 (2015).

13 Waddington, C. H. The epigenotype. 1942. *International journal of epidemiology* **41**, 10-13, doi:10.1093/ije/dyr184 (2012).

14 Waddington, C. H. *The strategy of the genes; a discussion of some aspects of theoretical biology*. (Allen & Unwin, 1957).

15 Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nature reviews. Molecular cell biology* **15**, 703-708, doi:10.1038/nrm3890 (2014).

16 Elliott, E. N. & Kaestner, K. H. Epigenetic regulation of the intestinal epithelium. *Cellular and molecular life sciences : CMLS* **72**, 4139-4156, doi:10.1007/s00018-015-1997-9 (2015).

17 Maze, I., Noh, K. M., Soshnev, A. A. & Allis, C. D. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature reviews. Genetics* **15**, 259-271, doi:10.1038/nrg3673 (2014).

18 Rossetto, D., Avvakumov, N. & Cote, J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098-1108, doi:10.4161/epi.21975 (2012).

19 Du, H. N. Transcription, DNA damage and beyond: the roles of histone ubiquitination and deubiquitination. *Current protein & peptide science* **13**, 447-466 (2012).

20 Tetteh, P. W., Farin, H. F. & Clevers, H. Plasticity within stem cell hierarchies in mammalian epithelia. *Trends in cell biology* **25**, 100-108, doi:10.1016/j.tcb.2014.09.003 (2015).

21 Yamada, Y., Haga, H. & Yamada, Y. Concise review: dedifferentiation meets cancer development: proof of concept for epigenetic cancer. *Stem cells translational medicine* **3**, 1182-1187, doi:10.5966/sctm.2014-0090 (2014).

22 Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly review of biology* **84**, 131-176 (2009).

23 van der Heijden, G. W. *et al.* Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Developmental biology* **298**, 458-469, doi:10.1016/j.ydbio.2006.06.051 (2006).

24 Gonzalez-Recio, O., Toro, M. A. & Bach, A. Past, present, and future of epigenetics applied to livestock breeding. *Frontiers in genetics* **6**, 305, doi:10.3389/fgene.2015.00305 (2015).

25 Desai, M., Jellyman, J. K. & Ross, M. G. Epigenomics, gestational programming and risk of metabolic syndrome. *International journal of obesity* **39**, 633-641, doi:10.1038/ijo.2015.13 (2015).
26 Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nature reviews. Molecular cell biology* **16**, 593-610, doi:10.1038/nrm4048 (2015).
27 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080, doi:10.1126/science.1063127 (2001).
28 Brown, C. J. & Robinson, W. P. XIST expression and X-chromosome inactivation in human preimplantation embryos. *American journal of human genetics* **61**, 5-8, doi:10.1086/513914 (1997).
29 Harmston, N. & Lenhard, B. Chromatin and epigenetic features of long-range gene regulation. *Nucleic acids research* **41**, 7185-7199, doi:10.1093/nar/gkt499 (2013).
30 Capuano, F., Mulleder, M., Kok, R., Blom, H. J. & Ralser, M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Analytical chemistry* **86**, 3697-3702, doi:10.1021/ac500447w (2014).
31 Tang, Y., Gao, X. D., Wang, Y., Yuan, B. F. & Feng, Y. Q. Widespread existence of cytosine methylation in yeast DNA measured by gas chromatography/mass spectrometry. *Analytical chemistry* **84**, 7249-7255, doi:10.1021/ac301727c (2012).
32 Kaaij, L. T. *et al.* DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome biology* **14**, R50, doi:10.1186/gb-2013-14-5-r50 (2013).
33 Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333**, 143-149, doi:10.1016/j.gene.2004.02.043 (2004).
34 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).
35 Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome research* **20**, 320-331, doi:10.1101/gr.101907.109 (2010).
36 Barres, R. *et al.* Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. *Cell metabolism* **10**, 189-198, doi:10.1016/j.cmet.2009.07.011 (2009).
37 Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905, doi:10.1126/science.1237905 (2013).
38 Patil, V., Ward, R. L. & Hesson, L. B. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* **9**, 823-828, doi:10.4161/epi.28741 (2014).
39 Ichiyanagi, T., Ichiyanagi, K., Miyake, M. & Sasaki, H. Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic acids research* **41**, 738-745, doi:10.1093/nar/gks1117 (2013).
40 Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome research* **23**, 555-567, doi:10.1101/gr.147942.112 (2013).
41 Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417, doi:10.1073/pnas.0510310103 (2006).
42 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics* **13**, 484-492, doi:10.1038/nrg3230 (2012).
43 Weisenberger, D. J. *et al.* Analysis of repetitive element DNA methylation by MethyLight. *Nucleic acids research* **33**, 6823-6836, doi:10.1093/nar/gki987 (2005).
44 Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**, 245-254, doi:10.1038/ng1089 (2003).
45 Marx, A., Kahan, T. & Simon, I. Integrative analysis of methylome and transcriptome reveals the importance of unmethylated CpGs in non-CpG island gene activation. *BioMed research international* **2013**, 785731, doi:10.1155/2013/785731 (2013).
46 Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086, doi:10.1038/nature08924 (2010).
47 Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *The EMBO journal* **30**, 3028-3039, doi:10.1038/emboj.2011.210 (2011).
48 Coleman-Derr, D. & Zilberman, D. Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS genetics* **8**, e1002988, doi:10.1371/journal.pgen.1002988 (2012).

49 Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics* **9**, 465-476, doi:10.1038/nrg2341 (2008).

50 Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916-919, doi:10.1126/science.1186366 (2010).

51 Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature genetics* **39**, 61-69, doi:10.1038/ng1929 (2007).

52 Coleman-Derr, D. & Zilberman, D. DNA methylation, H2A.Z, and the regulation of constitutive expression. *Cold Spring Harbor symposia on quantitative biology* **77**, 147-154, doi:10.1101/sqb.2012.77.014944 (2012).

53 Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-79, doi:10.1038/nature10442 (2011).

54 Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388-392, doi:10.1038/nature09147 (2010).

55 Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Molecular cell* **44**, 17-28, doi:10.1016/j.molcel.2011.08.026 (2011).

56 Bock, C. *et al.* DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Molecular cell* **47**, 633-647, doi:10.1016/j.molcel.2012.06.019 (2012).

57 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

58 Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: how is it established, and why does it matter? *Developmental biology* **339**, 258-266, doi:10.1016/j.ydbio.2009.06.012 (2010).

59 Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Current opinion in genetics & development* **12**, 162-169 (2002).

60 Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nature reviews. Molecular cell biology* **11**, 264-275, doi:10.1038/nrm2861 (2010).

61 Turinetto, V. & Giachino, C. Histone variants as emerging regulators of embryonic stem cell identity. *Epigenetics* **10**, 563-573, doi:10.1080/15592294.2015.1053682 (2015).

62 Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of biological chemistry* **273**, 5858-5868 (1998).

63 Fernandez-Capetillo, O., Lee, A., Nussenzweig, M. & Nussenzweig, A. H2AX: the histone guardian of the genome. *DNA repair* **3**, 959-967, doi:10.1016/j.dnarep.2004.03.024 (2004).

64 Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature genetics* **41**, 941-945, doi:10.1038/ng.409 (2009).

65 Ku, M. *et al.* H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome biology* **13**, R85, doi:10.1186/gb-2012-13-10-r85 (2012).

66 Hu, G. *et al.* H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell stem cell* **12**, 180-192, doi:10.1016/j.stem.2012.11.003 (2013).

67 Yang, X. *et al.* Gene reactivation by 5-aza-2'-deoxycytidine-induced demethylation requires SRCAP-mediated H2A.Z insertion to establish nucleosome depleted regions. *PLoS genetics* **8**, e1002604, doi:10.1371/journal.pgen.1002604 (2012).

68 Li, Z. *et al.* Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**, 1608-1616, doi:10.1016/j.cell.2012.11.018 (2012).

69 Guillemette, B. *et al.* Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS biology* **3**, e384, doi:10.1371/journal.pbio.0030384 (2005).

70 Li, B. *et al.* Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18385-18390, doi:10.1073/pnas.0507975102 (2005).

71 Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**, 233-248, doi:10.1016/j.cell.2005.10.002 (2005).

72 Creyghton, M. P. *et al.* H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell* **135**, 649-661, doi:10.1016/j.cell.2008.09.056 (2008).

73 Conerly, M. L. *et al.* Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome research* **20**, 1383-1390, doi:10.1101/gr.106542.110 (2010).

74 Fujimoto, S. *et al.* Proteome analysis of protein partners to nucleosomes containing canonical H2A or the variant histones H2A.Z or H2A.X. *Biological chemistry* **393**, 47-61, doi:10.1515/BC-2011-216 (2012).
75 Tolstorukov, M. Y., Kharchenko, P. V., Goldman, J. A., Kingston, R. E. & Park, P. J. Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes. *Genome research* **19**, 967-977, doi:10.1101/gr.084830.108 (2009).
76 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annual review of genetics* **43**, 559-599, doi:10.1146/annurev.genet.032608.103928 (2009).
77 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).
78 Rodriguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nature medicine* **17**, 330-339, doi:10.1038/nm.2305 (2011).
79 Sevilla, A. & Binda, O. Post-translational modifications of the histone variant H2AZ. *Stem cell research* **12**, 289-295, doi:10.1016/j.scr.2013.11.004 (2014).
80 Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72**, 73-84 (1993).
81 Gorisch, S. M., Wachsmuth, M., Toth, K. F., Lichter, P. & Rippe, K. Histone acetylation increases chromatin accessibility. *Journal of cell science* **118**, 5825-5834, doi:10.1242/jcs.02689 (2005).
82 Chen, J., Ghazawi, F. M. & Li, Q. Interplay of bromodomain and histone acetylation in the regulation of p300-dependent genes. *Epigenetics* **5**, 509-515 (2010).
83 Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8695-8700, doi:10.1073/pnas.082249499 (2002).
84 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).
85 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
86 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).
87 Kim, T. H. *et al.* Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. *Nature* **506**, 511-515, doi:10.1038/nature12903 (2014).
88 Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature reviews. Cancer* **7**, 823-833, doi:10.1038/nrc2253 (2007).
89 Roguev, A. *et al.* The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* **20**, 7137-7148, doi:10.1093/emboj/20.24.7137 (2001).
90 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).
91 Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953, doi:10.1016/j.cell.2004.12.012 (2004).
92 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).
93 Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880, doi:10.1038/nature03877 (2005).
94 Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735-745, doi:10.1016/j.cell.2007.02.009 (2007).
95 Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353, doi:10.1038/nature04733 (2006).
96 Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes & development* **27**, 1318-1338, doi:10.1101/gad.219626.113 (2013).
97 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).
98 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).

99 Sheaffer, K. L. & Kaestner, K. H. Transcriptional networks in liver and intestinal development. *Cold Spring Harbor perspectives in biology* **4**, a008284, doi:10.1101/cshperspect.a008284 (2012).
100 de Santa Barbara, P., van den Brink, G. R. & Roberts, D. J. Development and differentiation of the intestinal epithelium. *Cellular and molecular life sciences : CMLS* **60**, 1322-1332, doi:10.1007/s00018-003-2289-3 (2003).
101 Crosnier, C., Stamataki, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature reviews. Genetics* **7**, 349-359, doi:10.1038/nrg1840 (2006).
102 Madison, B. B. *et al.* Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* **132**, 279-289, doi:10.1242/dev.01576 (2005).
103 Apelqvist, A., Ahlgren, U. & Edlund, H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Current biology : CB* **7**, 801-804 (1997).
104 Sukegawa, A. *et al.* The concentric structure of the developing gut is regulated by Sonic hedgehog derived from endodermal epithelium. *Development* **127**, 1971-1980 (2000).
105 Ogaki, S., Shiraki, N., Kume, K. & Kume, S. Wnt and Notch signals guide embryonic stem cell differentiation into the intestinal lineages. *Stem cells* **31**, 1086-1096, doi:10.1002/stem.1344 (2013).
106 Kim, B. M., Mao, J., Taketo, M. M. & Shivdasani, R. A. Phases of canonical Wnt signaling during the development of mouse intestinal epithelium. *Gastroenterology* **133**, 529-538, doi:10.1053/j.gastro.2007.04.072 (2007).
107 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).
108 Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265, doi:10.1038/nature07935 (2009).
109 Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134-144, doi:10.1016/j.cell.2010.09.016 (2010).
110 Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822-825, doi:10.1126/science.1196236 (2010).
111 Moossavi, S. Heterogeneity of the level of activity of lgr5+ intestinal stem cells. *International journal of molecular and cellular medicine* **3**, 216-224 (2014).
112 He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nature genetics* **36**, 1117-1121, doi:10.1038/ng1430 (2004).
113 Noah, T. K. & Shroyer, N. F. Notch in the intestine: regulation of homeostasis and pathogenesis. *Annual review of physiology* **75**, 263-288, doi:10.1146/annurev-physiol-030212-183741 (2013).
114 Tian, H. *et al.* Opposing activities of Notch and Wnt signaling regulate intestinal stem cells and gut homeostasis. *Cell reports* **11**, 33-42, doi:10.1016/j.celrep.2015.03.007 (2015).
115 Haramis, A. P. *et al.* De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* **303**, 1684-1686, doi:10.1126/science.1093587 (2004).
116 Yin, X. *et al.* Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. *Nature methods* **11**, 106-112, doi:10.1038/nmeth.2737 (2014).
117 Marshman, E., Booth, C. & Potten, C. S. The intestinal epithelial stem cell. *BioEssays : news and reviews in molecular, cellular and developmental biology* **24**, 91-98, doi:10.1002/bies.10028 (2002).
118 Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415-418, doi:10.1038/nature09637 (2011).
119 van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annual review of physiology* **71**, 241-260, doi:10.1146/annurev.physiol.010908.163145 (2009).
120 van Es, J. H. *et al.* Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nature cell biology* **7**, 381-386, doi:10.1038/ncb1240 (2005).
121 Brennan, P. C., McCullough, J. S. & Carr, K. E. Variations in cell and structure populations along the length of murine small intestine. *Cells, tissues, organs* **164**, 221-226, doi:16662 (1999).
122 Al Alam, D. *et al.* Fibroblast growth factor 10 alters the balance between goblet and Paneth cells in the adult mouse small intestine. *American journal of physiology. Gastrointestinal and liver physiology* **308**, G678-690, doi:10.1152/ajpgi.00158.2014 (2015).
123 Anderle, P. *et al.* Changes in the transcriptional profile of transporters in the intestine along the anterior-posterior and crypt-villus axes. *BMC genomics* **6**, 69, doi:10.1186/1471-2164-6-69 (2005).
124 Boffa, L. C., Vidali, G., Mann, R. S. & Allfrey, V. G. Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. *The Journal of biological chemistry* **253**, 3364-3366 (1978).

125 Shin, J. H., Li, R. W., Gao, Y., Baldwin, R. t. & Li, C. J. Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. *Functional & integrative genomics* **12**, 119-130, doi:10.1007/s10142-012-0263-6 (2012).
126 Goodell, M. A., Nguyen, H. & Shroyer, N. Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. *Nature reviews. Molecular cell biology* **16**, 299-309, doi:10.1038/nrm3980 (2015).
127 Theodosiou, N. A. & Tabin, C. J. Wnt signaling during development of the gastrointestinal tract. *Developmental biology* **259**, 258-271 (2003).
128 Sherwood, R. I., Maehr, R., Mazzoni, E. O. & Melton, D. A. Wnt signaling specifies and patterns intestinal endoderm. *Mechanisms of development* **128**, 387-400, doi:10.1016/j.mod.2011.07.005 (2011).
129 Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327-339, doi:10.1016/j.cell.2011.01.024 (2011).
130 Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499-502, doi:10.1038/nature05295 (2006).
131 Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nature biotechnology* **31**, 1133-1136, doi:10.1038/nbt.2701 (2013).
132 Dillon, R., Gadgil, C. & Othmer, H. G. Short- and long-range effects of Sonic hedgehog in limb development. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10152-10157, doi:10.1073/pnas.1830500100 (2003).
133 Machold, R. *et al.* Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches. *Neuron* **39**, 937-950 (2003).
134 Ihrie, R. A. *et al.* Persistent sonic hedgehog signaling in adult brain determines neural stem cell positional identity. *Neuron* **71**, 250-262, doi:10.1016/j.neuron.2011.05.018 (2011).
135 Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797-803, doi:10.1242/dev.01613 (2005).
136 Zhao, J. *et al.* HnRNP U mediates the long-range regulation of Shh expression during limb development. *Human molecular genetics* **18**, 3090-3097, doi:10.1093/hmg/ddp250 (2009).
137 Tiecke, E., Turner, R., Sanz-Ezquerro, J. J., Warner, A. & Tickle, C. Manipulations of PKA in chick limb development reveal roles in digit patterning including a positive role in Sonic Hedgehog signaling. *Developmental biology* **305**, 312-324, doi:10.1016/j.ydbio.2007.02.017 (2007).
138 Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics* **12**, 1725-1735 (2003).
139 Amano, T. *et al.* Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell* **16**, 47-57, doi:10.1016/j.devcel.2008.11.011 (2009).
140 Lettice, L. A. *et al.* Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Developmental cell* **22**, 459-467, doi:10.1016/j.devcel.2011.12.010 (2012).
141 Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F. & Chiang, C. Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* **418**, 979-983, doi:10.1038/nature01033 (2002).
142 Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nature reviews. Genetics* **15**, 321-334, doi:10.1038/nrg3686 (2014).
143 Bogdanove, A. J. & Voytas, D. F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843-1846, doi:10.1126/science.1204094 (2011).
144 Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nature protocols* **7**, 171-192, doi:10.1038/nprot.2011.431 (2012).
145 Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Molecular cell* **36**, 178-191, doi:10.1016/j.molcel.2009.09.018 (2009).
146 Myer, V. E. & Young, R. A. RNA polymerase II holoenzymes and subcomplexes. *The Journal of biological chemistry* **273**, 27757-27760 (1998).
147 Bataille, A. R. *et al.* A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Molecular cell* **45**, 158-170, doi:10.1016/j.molcel.2011.11.024 (2012).
148 Xue, X. & Lehming, N. Nhp6p and Med3p regulate gene expression by controlling the local subunit composition of RNA polymerase II. *Journal of molecular biology* **379**, 212-230, doi:10.1016/j.jmb.2008.03.069 (2008).

149 Landry, J. J. *et al.* The genomic and transcriptomic landscape of a HeLa cell line. *G3* **3**, 1213-1224, doi:10.1534/g3.113.005777 (2013).
150 Ayoubi, T. A. & Van De Ven, W. J. Regulation of gene expression by alternative promoters. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **10**, 453-460 (1996).
151 Gomez, D., Shankman, L. S., Nguyen, A. T. & Owens, G. K. Detection of histone modifications at specific gene loci in single cells in histological sections. *Nature methods* **10**, 171-177, doi:10.1038/nmeth.2332 (2013).
152 Roussis, I. M., Guille, M., Myers, F. A. & Scarlett, G. P. RNA Whole-Mount In situ Hybridisation Proximity Ligation Assay (rISH-PLA), an Assay for Detecting RNA-Protein Complexes in Intact Cells. *PloS one* **11**, e0147967, doi:10.1371/journal.pone.0147967 (2016).
153 Vikas B. Palhan, C. K., and Gregory D. Davis. *Visualizing DNA Methylation (5mC) on a Specific Genomic Locus (SEPTIN 9) in Individual Human Cancer Cells using In-situ Hybridization and Proximity Ligation Assays*, <http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma-Aldrich/Application_Notes/1/duolink-epi-apps.pdf> (2010).
154 Tsukiji, N., Amano, T. & Shiroishi, T. A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mechanisms of development* **131**, 127-136, doi:10.1016/j.mod.2013.09.003 (2014).
155 Sheaffer, K. L. *et al.* DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes & development* **28**, 652-664, doi:10.1101/gad.230318.113 (2014).
156 Jadhav, U. *et al.* Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* **165**, 1389-1400, doi:10.1016/j.cell.2016.04.031 (2016).
157 Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273-287, doi:10.1016/j.virol.2010.12.007 (2011).
158 Friedli, M. & Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annual review of cell and developmental biology* **31**, 429-451, doi:10.1146/annurev-cellbio-100814-125514 (2015).
159 Guo, H. *et al.* Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nature protocols* **10**, 645-659, doi:10.1038/nprot.2015.039 (2015).
160 Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Molecular cell* **58**, 610-620, doi:10.1016/j.molcel.2015.04.005 (2015).
161 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature biotechnology* **33**, 1165-1172, doi:10.1038/nbt.3383 (2015).
162 Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 179-184, doi:10.1073/pnas.1013004108 (2011).
163 Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nature genetics* **40**, 915-920, doi:10.1038/ng.165 (2008).
164 Vincent, A. *et al.* Cryosectioning the intestinal crypt-villus axis: an ex vivo method to study the dynamics of epigenetic modifications from stem cells to differentiated cells. *Stem cell research* **14**, 105-113, doi:10.1016/j.scr.2014.12.002 (2015).
165 Mariadason, J. M. *et al.* Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. *Gastroenterology* **128**, 1081-1088 (2005).
166 Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics* **14**, 390-403, doi:10.1038/nrg3454 (2013).
167 Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* **10**, 757, doi:10.15252/msb.20145625 (2014).
168 Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39**, e82, doi:10.1093/nar/gkr218 (2011).
169 Fujita, T. *et al.* Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Scientific reports* **3**, 3171, doi:10.1038/srep03171 (2013).
170 Fujita, T. & Fujii, H. Direct identification of insulator components by insertional chromatin immunoprecipitation. *PloS one* **6**, e26109, doi:10.1371/journal.pone.0026109 (2011).
171 Fujita, T., Ryser, S., Tortola, S., Piuz, I. & Schlegel, W. Gene-specific recruitment of positive and negative elongation factors during stimulated transcription of the MKP-1 gene in neuroendocrine cells. *Nucleic acids research* **35**, 1007-1017, doi:10.1093/nar/gkl1138 (2007).
172 Behringer, R. *Manipulating the mouse embryo : a laboratory manual*. Fourth edition. edn, (CSH Press, 2014).

173 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
174 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).

11. Acknowledgements

The described projects were performed at the Institute of Molecular Biology (IMB), Mainz. For this opportunity, I thank my supervisor, who introduced me to the most challenging and interesting fields of developmental biology and epigenomics, providing outstanding guidance, advice and support throughout the past 4 years.

I am very thankful to the members of my thesis advisory committee, who took regularly the time to assess my progress and to provide me with good advice and plentiful new ideas.

This work would not be possible without the active support and hands-on teaching by the IMB core facilities and the TARC Mainz. Especially I would like to thank the animal caretakers, the CF Cytometry, Bioinformatics, Genomics and Microscopy, with each of whom I was working on a daily basis.

The past 4 years would not be the same without my dear colleagues, with whom I shared the joys of research, all positive and negative results as well as quite an amount of coffee, tea and sweets. As much as your scientific input was helpful, I enjoyed the time with you and I hope that we will stay in touch in the years to come.

Last but not least, I would like to thank all my friends and my family, who were so supportive and patient with me. A very special thanks goes to my little niece and nephew, who gave me a smile any time I saw, heard or thought about them.