Mechanisms of Regulatory Adaptation in the Evolving Genome

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by
Alessa R. Ringel
Berlin 2023

The present work was conducted from 04.12.2017 to 04.01.2023 at the Max Planck Institute for Molecular Genetics in the research group of Prof. Dr. Mundlos.

1st reviewer: Prof. Dr. Stefan Mundlos
2nd reviewer: Prof. Dr. Sigmar Stricker

Date of defense: 26.05.2023

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. Intellectual property of other authors has been marked accordingly. I also declare that I have not applied for an examination procedure at any other institution and that I have not submitted the dissertation in this or any other form to any other faculty as a dissertation.

TABLE OF CONTENT

SUMMARY

The development from a single cell into a complex organism requires the precise control of gene expression in space and time. To achieve this, the activity of genes is governed by large regulatory chromatin landscapes that when disrupted can cause gene mis-regulation and disease. However, at the same time, the successful modification of these landscapes is thought to be a major driver of phenotypic innovation during evolution. Given the vulnerability of these landscapes in disease settings, it remains largely unknown how their integrity is maintained when novel genes are “safely” incorporated during evolution, which is addressed in this work.

Specifically, here, multiple mechanisms are dissected that adapted the *Fat1* regulatory landscape to maintain its integrity while simultaneously incorporating a novel gene, *Zfp42*, during evolution. First, comparative evolutionary genomics was used to reconstruct the history of the locus (section 1). Second, the three-dimensional chromatin configuration of the locus was examined in relationship to the gene activities using genomics-technologies (HiC, DamID) combined with super resolution microscopy and *in silico* modeling (section 2). Finally, the mechanisms that adapted the landscape in ESCs (section 3) and embryonic limbs (section 4) for the emergence of *Zfp42* were investigated using genome engineering and genomics.

Two tissue-specific mechanisms were identified that enabled the independent activities of *Zfp42* and *Fat1* despite sharing the same regulatory chromatin landscape: In ESCs, the landscape physically restructures and isolates the genes together with their regulatory information, from one another, thereby allowing their independent regulation. Surprisingly, this restructuring is not driven by the most recognized chromatin structuring force, loop extrusion, but rather by the underlying epigenetic state of chromatin. A different mechanism operates in embryonic mouse limbs where both genes are exposed to the same regulatory information driving *Fat1* activation, but surprisingly not *Zfp42*. The inactivity of *Zfp42* cannot be explained by nuclear envelopment attachment nor by enhancer-promoter specificity. Instead, *Zfp42* is kept inactive by a highly context-dependent silencing mechanism driven by DNA methylation. As such, *Zfp42* is ectopically active and responsive to the surrounding regulatory information when DNA methylation is removed or when the gene is slightly repositioned within its domain.

Combined, we find that *3D-restructuring* and *context-dependent silencing* adapted the *Fat1* landscape to integrate *Zfp42*. More generally, this demonstrates that even single regulatory landscapes harbor an enormous regulatory complexity and, thus can accommodate multiple independently regulated genes. We believe that this has significant consequences for human genetics where similar genomic alterations do not drive disease in patients. This is possible, because additional, yet still unknown, mechanisms control how regulatory information is used in the genome.

ZUSAMMENFASSUNG

Die Entwicklung von einer einzelnen Zelle zu einem komplexen Organismus erfordert die präzise räumliche und zeitliche Aktivierung von Genen. Um dies zu erreichen, wird die Expression von Genen durch große regulatorische Chromatinlandschaften kontrolliert. Die Störung dieser genetischen Organisationseinheiten kann zu einer Fehlregulation und Krankheiten führen. Allerdings wird gleichzeitig angenommen, dass die Modifikation dieser Landschaften eine wichtige Triebkraft für die Ausprägung neuer phänotypischer Merkmale während der Evolution darstellt. Angesichts der Bedeutung dieser Landschaften bei der Entstehung von Krankheiten ist weitgehend unklar, wie sie im Laufe der Evolution „sicher" verändert werden können, um neue Gene zu integrieren, was in dieser Arbeit untersucht wurde.

Konkret wurden hier mehrere molekulare Mechanismen untersucht, die die regulatorische Landschaft des Gens *Fat1* so angepasst haben, dass sie ihre Integrität aufrechterhalten und gleichzeitig ein neues Gen, *Zfp42*, während der Evolution inkorporieren konnte. Zunächst wurde durch Genomik-Vergleiche mit anderen Spezies die Evolutionsgeschichte der Chromatinlandschaft rekonstruiert (Kapitel 1). Anschließend wurden die Aktivität der Gene und die dreidimensionale Chromatinfaltung der Landschaft mittels Genomik-Technologien (HiC, DamID), hochauflösender Mikroskopie und Computersimulationen charakterisiert (Kapitel 2). Schließlich wurden die Mechanismen, die die Landschaft in Stammzellen (Kapitel 3) und Gliedmaßenknospen der Maus angepasst haben (Kapitel 4), um die Integration von *Zfp42* zu ermöglichen, mittels Genom-Engineering- und Genomik-Technologien untersucht.

Dies führte zur Identifizierung von zwei gewebespezifischen Mechanismen, die die unabhängige Regulation von *Zfp42* und *Fat1* ermöglichen, obwohl sie dieselbe regulatorische Chromatinlandschaft teilen: In Stammzellen wird die Landschaft physikalisch umstrukturiert, sodass die Gene zusammen mit ihren regulatorischen Informationen voneinander getrennt werden, was unabhängige Aktivitäten ermöglicht. Interessanterweise wird diese alternative Chromatinfaltung nicht durch den bekannten Prozess der Loop-Extrusion angetrieben, sondern durch den epigenetischen Zustand des Chromatins. In den Gliedmaßenknospen hingegen wirkt ein anderer Mechanismus. Hier sind beide Gene den gleichen regulatorischen Informationen ausgesetzt, die zwar zur Aktivierung von *Fat1*, überraschenderweise aber nicht von *Zfp42* führen. Die Inaktivität von *Zfp42* lässt sich weder durch seine Interaktion mit der Kernhülle noch durch Enhancer-Promoter-Spezifität erklären. Stattdessen wird *Zfp42* durch DNA-Methylierung inaktiv gehalten, welche stark von der Sequenzzusammensetzung in der direkten Umgebung des Gens abhängt. So ist *Zfp42* ektopisch aktiv und reagiert auf die umgebenden regulatorischen Informationen, wenn die DNA-Methylierung entfernt oder das Gen innerhalb seiner Domäne leicht verschoben wird.

Insgesamt haben wir zwei molekulare Mechanismen, *3D-Umstrukturierung* und *kontextabhängiges Silencing*, gefunden, die die Landschaft von *Fat* für eine erfolgreiche Integration von *Zfp42* angepasst haben. Allgemein zeigt dies, dass sogar einzelne Chromatinlandschaften eine enorme regulatorische Komplexität besitzen, die es ihnen erlaubt, mehrere voneinander unabhängig regulierte Gene zu beherbergen. Wir glauben, dass dies bedeutende Konsequenzen für die Humangenetik hat, wo viele ähnlich veränderte Landschaften bei Patienten und Patientinnen nicht immer zu Krankheiten führen. Dies ist möglich, weil mehrere, teilweise noch unbekannte Mechanismen kontrollieren, auf welche Weise Gene ihre regulatorischen Informationen im Genom nutzen.

INTRODUCTION

The tightly coordinated spatiotemporal activities of genes enable a single cell to develop into a highly complex organism and, later, to maintain tissue homeostasis. As such, disrupting gene activities drives various pathologies from congenital malformation and cancer to congenital diseases (Chakraborty and Ay 2019; Anania and Lupiáñez 2020). However, at the same time, the successful modification of gene activities is a major driver of the evolutionary innovation and phenotypic diversity seen across the animal kingdom (Long et al. 2016b). Combined, this raises the puzzling question of how gene activities so sensitive to disruption can be "safely" modified during evolution. Moreover, what mechanisms are used to sustain existing gene activities and simultaneous integrate novel ones? This thesis addresses these questions and reveals how the genome´s regulatory complexity evolves over time.

In metazoans, precise gene regulation is encoded within cis-regulatory elements (CREs) that include gene promoters and the distal enhancer elements that control them from surrounding chromatin (Robson et al. 2019). The communication between promoters and enhancers is, on one hand, controlled by the binding of epigenetic modifiers and their instructive modifications. On the other hand, the precise structural folding of chromatin in 3D space plays a critical role (Fig. 1). As such, chromatin loops create enhancer-promoter contacts that modulate cell-type-specific gene expression (Bolt et al. 2022). These loops occur within topologically-associated domains (TADs) that partition the genome into discrete functional blocks (Dixon et al. 2012; Nora et al. 2012). Finally, interactions between TADs with similar epigenetic states further spatially separates chromatin by activity into active (A) and inactive (B) compartments (Lieberman-Aiden et al. 2009). Moreover, some chromatin dynamically attaches to the nuclear envelope (NE) in Lamina associated domains (LADs) which is intrinsically linked to a gene´s activity (van Steensel and Belmont 2017). Thus, collectively CREs, epigenetic regulators and chromatin structure collectively define a gene's regulatory landscape and so can all serve as substrates for evolutionary innovation.

This introduction will give an overview about the current state of knowledge of how regulatory landscapes are built, how they are disrupted in disease and modified in evolution. The first part focuses on the functional components and the second part on the structure of chromatin regulatory landscapes.

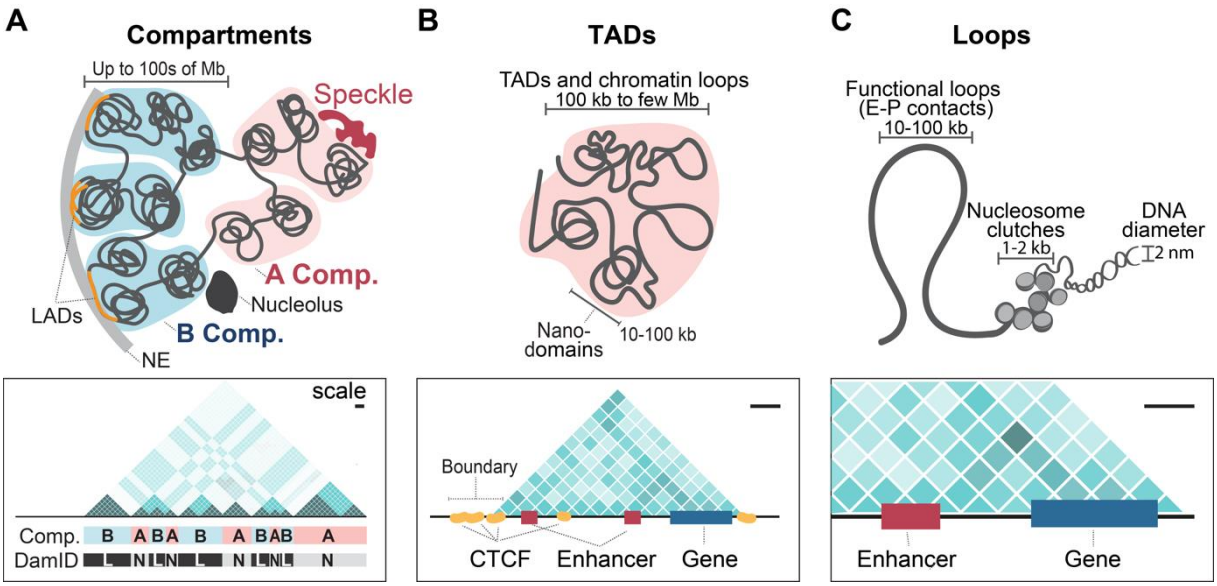


Figure 1 Overview of 3D genome organization
Schematic of structural feature (upper panel) with corresponding HiC views (lower panel). **A**. Below chromosomal territories, chromatin spatially separates into active A- (red) and inactive B- (green) compartments that are megabases in size (Lieberman-Aiden et al. 2009). A subset of B compartments is anchored at the NE in LADs (orange). In HiC maps higher order interaction of compartments (A-A or B-B) are visible as plaid pattern. The NE association of chromatin can be detected by DamID-seq (L: LAD, N: nonLAD). **B**. TADs are presence at the submega-base to megabase scale and are delimited by boundaries frequently consisting of multiple CTCF sites (yellow) (Nora et al. 2012; Dixon et al. 2012). Sub-structures such as nanodomains are found in a range of 10-100kb and represent globular structures of active and inactive chromatin (Szabo et al. 2020). **C**. In some cases, chromatin loops can be detected as distinct dots in HiC representing an increased contact frequency of the loop anchors frequently occupied by genes and enhancer.

1. The building blocks of regulatory landscapes

1.1 Types of CREs

CREs are non-coding DNA sequences that control gene transcription. There are multiple types of CREs namely, promoters, enhancers and silencers (Wittkopp and Kalay 2011). While promoters drive transcription of a gene itself, enhancers and silencers modulate where, when, and how strongly this transcription occurs. In this way, CREs fine-tune and coordinate the activity of genes in specific cell-types and at distinct developmental time points. Hence, they also represent ideal templates for modifying gene expression in evolution and disease.

Promoters

Promoters are regions upstream of genes where transcription starts. Promoter sequences consist of a proximal and core promoter element. The core promoter is an around 100 bp region around the transcriptional start site (TSS) that is sufficient for transcription initiation (Haberle and Stark 2018). In metazoans, promoters recruit and assist the accurate positioning of the pre-initiation complex (PIC) which includes general transcription factors

(TF) and RNA polymerase II (Pol II) (Haberle and Stark 2018). Once the transcription apparatus is assembled, RNA Pol II can initiate transcription and subsequently transition to an elongating form that transcribes the gene body until reaching a terminator sequence (Shandilya and Roberts 2012). Generally, 2/3 of all mammalian promoters contain CpG islands, dense clusters of cytosines followed guanines (CpG), or alteratively display a TATA-box binding motive (Deaton and Bird 2011; Mishal and Luna-Arias 2022). Moreover, one can distinguish between ubiquitously active housekeeping promoters and tissue-specific developmental promoters displaying stronger sequence conservation (Farré et al. 2007; Haberle et al. 2019). Thus, the core promoter itself contains a considerable amount of information for gene regulation.

Enhancers

In bacteria, promoters contain all information that is needed to recruit the transcription machinery for gene expression (Chen et al. 2021). Yet, this is not the case in metazoans. Here, when, where and to what level promoters transcribe is governed by activating CREs, enhancers. First identified in the simian virus, enhancers are defined as elements that *enhance* gene transcription in an orientation- and distance-independent manner (Moreau et al. 1981; Banerji et al. 1981). Thus, while promoters are found at a gene´s TSS, the roughly 0,2-1 kbp large enhancer elements can be widely positioned in the surrounding landscape (Kvon 2015). Enhancers contain dense clusters of sequence motifs that act as binding platforms for a combination of TFs which collectively determine the enhancer´s tissue-specific activity (Shlyueva et al. 2014; Panne 2008). Through not yet fully understood, it is thought that the distinct enhancer activities are encoded in common rules like the number, spacing, orientation, affinity and/or order of TF binding sites (“enhancer grammar”) (Jindal and Farley 2021). Following TF binding, transcriptional co-activators including p300 and Mediator complex are recruited to stimulate PIC assembly at the enhancer site (Shlyueva et al. 2014; Bulger and Groudine 2011). Once an enhancer is active, bidirectional transcription is often, but not always, initiated to create short, non-polyadenylated enhancer RNAs (De Santa et al. 2010; Kim et al. 2010). Much is left to learn of how enhancers are defined and function. However, clearly enhancers are crucial regulators of spatiotemporal gene expression and provide suitable substrate for evolution to modify gene expression.

Silencers

The negative modulation of gene activity occurs through repressive CREs, called silencers, which are far less understood than activating CREs. Similar to enhancers, silencers contain multiple TF binding sites and can be positioned anywhere within a

regulatory landscape (Zhang et al. 2022). Generally, they can be separated into *silencing elements* and *negative regulatory elements*. (Ogbourne and Antalis 1998). In this view, *silencer elements* drive direct repression by directly blocking binding sites for activating TFs. (Ogbourne and Antalis 1998; Zhang et al. 2022). In contrast, *negative regulatory elements* drive passive repression by indirectly preventing the binding of activating TFs through the establishment of a repressive chromatin state which often involves histone deacetylation and methylation (Zhang et al. 2022; Ogbourne and Antalis 1998). As a result, both interfere directly or indirectly with the assembly of the transcriptional machinery and its function (Ogbourne and Antalis 1998; Zhang et al. 2022). Of note, it is increasingly noticed that CREs display bimodular functions as an activator or repressor depending on the cellular contexts suggesting a broad range of action for CREs (Reynolds et al. 2013; Ogbourne and Antalis 1998; Zhang et al. 2022). Collectively, silencers represent another major CRE type that inhibits gene expression.

Combined, regulatory landscapes contain an assembly of gene promoters that are controlled by enhancers and silencers. However, identifying their sequences and functions within the genome and, moreover, how they collectively operate remains challenging.

1.2 Examining CREs

In the following sections, I will describe how CREs can be identified. As the knowledge about silencers is limited I will mainly focus on promoters and enhancers CREs. Generally, CREs can be distinguished and characterized by a number of features including their epigenetic signature, conservation, functional assays and, recently, machine learning approaches trained on a subset of the aforementioned features.

Epigenome

Putative CREs and their functional states can be identified by their unique epigenetic signatures through multiple genomics approaches. Specifically, CREs display a unique combination of chromatin accessibility, protein binding, histone modifications and DNA methylation. Chromatin accessibility is detected by sequencing-based assay including ATAC-seq, DNase-seq and MNase-seq (Crawford et al. 2006; Schones et al. 2008; Buenrostro et al. 2013). Histone modifications and DNA binding proteins can be assayed by CUT&TAG or chromatin immunoprecipitation followed by sequencing (ChIP-seq) while DNA methylation can be analyzed with whole genome bisulfide sequencing (WGBS) (Solomon et al. 1988; Kaya-Okur et al. 2019; Olova et al. 2018). Collectively, these tools allow for systematic mapping of putative CREs at a genome-wide scale.

Large scale analyses have revealed CREs can be distinguished by distinct epigenetic states (Heintzman et al. 2009). Active promoters and enhancers both overlap with chromatin accessible regions and can be distinguished though the binding of distinct transcription factors (Palstra and Grosveld 2012). Further, promoters possess H3K27ac marks, high levels of H3K4me3 and an enrichment of stalled or paused RNA-Pol II while the transcribed gene itself is marked by H3K36me3 (reviewed in (Bannister and Kouzarides 2011; Calo and Wysocka 2013). In contrast, enhancers are marked by H3K27ac, H3K4me3 and high levels of H3K4me1 (reviewed in (Bannister and Kouzarides 2011; Calo and Wysocka 2013). Additionally, most promoters display a high CpG density and enhancers a low CpG density (Gardiner-Garden and Frommer 1987; Carter and Zhao 2021). The majority of promoter CpG islands are unmethylated independent of a gene's activity, but if methylated, it strongly correlates with gene repression (Luo et al. 2018; Deaton and Bird 2011). In contrast, enhancers frequently display tissue specific dynamic methylation with hypomethylation linked to activity, thereby enabling the identification of key lineage-specific regulators (Ziller et al. 2013) (Andersson et al. 2014; Schultz et al. 2015; Hon et al. 2013; Luo et al. 2018). Lastly, silencers are seemingly marked by accessible chromatin and the repressive histone modification H3K27me3 (Huang et al. 2019; Zhang et al. 2022).

Collectively, these features can be used for the systematic identification of putative CREs by diverse computational approaches including the CRUP algorithm (Condition-specific Regulatory Units Prediction) (Ramisch et al. 2019).

Functional assays

Though genomics methods can identify putative CREs, functional assays are critical for testing their actual activities. While promoters are easily identified by their proximity to a gene´s TSS, putative enhancer elements need further validations. Typically, enhancers are tested for two properties; namely (i) their ability to drive reporter gene transcription and (ii) their requirement for endogenous gene expression. For the former reporter assay, a putative enhancer is cloned next to a reporter gene whose transcription is driven by a minimal promoter (Kvon 2015). By lacking autonomous activity, transcription at this minimal promoter thus relies on the presence of an enhancer. Subsequently, these enhancer-reporter constructs can be tested either episomally or by their integration in a random or site-specific manner into the genome (Kvon 2015). The spatiotemporal enhancer activity can then be visualized by the reporter´s transcription, either *in vitro* or *in vivo* (Fig. 2A) (Visel et al. 2007; Kvon 2015). In this way, hundreds of individual putative elements have been tested as enhancers.

A disadvantage of these assays is their limitation to systematically test enhancers as they are generally low throughput. In contrast, sequencing based approaches allow to asses enhancer activities of thousands of elements in parallel in MPRAs (Massively Parallel Reporter Assays) performed *in vivo* or *in vitro* (Santiago-Algarra et al. 2017). MPRAs often use plasmid-based enhancer-promoter constructs with unique barcodes allowing the later assignment of transcriptional outputs to the corresponding tested element. In this way, CREs could be compared at large scale revealing that enhancers display a broad range of activities and can be classified by common features (Zabidi et al. 2015; Trauernicht et al. 2020; Vanhille et al. 2015). Moreover, using an MPRA recently allowed the evaluation of the evolutionary sequence architecture of 30,000 human enhancers, their functional activity and correlation with human trait variations (Fong and Capra 2021).

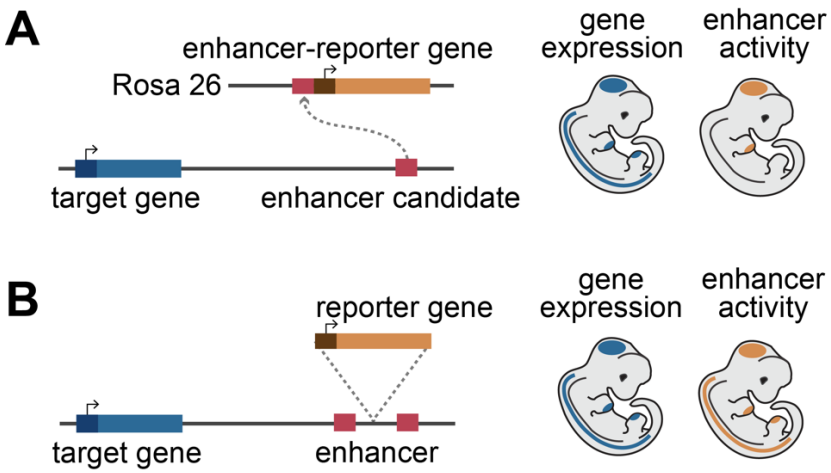


Figure 2 Mapping enhancer activities
A. Enhancer-reporter assay: A putative enhancer is cloned together with a minimal promoter and reporter gene into a safe harbor locus (e.g. *Rosa 26*). Generating mouse embryos carrying such constructs enables to test putative enhancer elements for their ability to drive a subset of a gene´s overall expression pattern. **B.** Sensor integrations: Sensors consist of a reporter gene under the control of a minimal promote. When knocked into a genomic landscape, they report the availability of enhancer activities at their integration site (Robson et al. 2019; Symmons et al. 2014).

Though powerful tools, reporter assays remove CREs from their endogenous chromatin context and this can potentially influence their reported activities (Bolt et al. 2022). Moreover, regulatory landscapes are frequently composed of many CREs (Osterwalder et al. 2018; Will et al. 2017). As such, it is not possible from reporter assays alone to identify which enhancers regulate which gene(s). Instead, elements can be test by eliminating or epigenetically-modulating them at their endogenous location by targeted genome editing technologies (e.g. CRISPR, CRISPRa and CRISPRi) (Liu et al. 2022; Kraft et al. 2015). Gene expression and phenotypic consequences can then be mapped by multiple approaches, including RNA-seq or creation of mutant embryos (Fig. 3 and 26). However, as regulatory landscapes frequently harbor redundant enhancers, observing no

gene expression effect upon element deletion, activation or repression does not exclude its regulatory function.

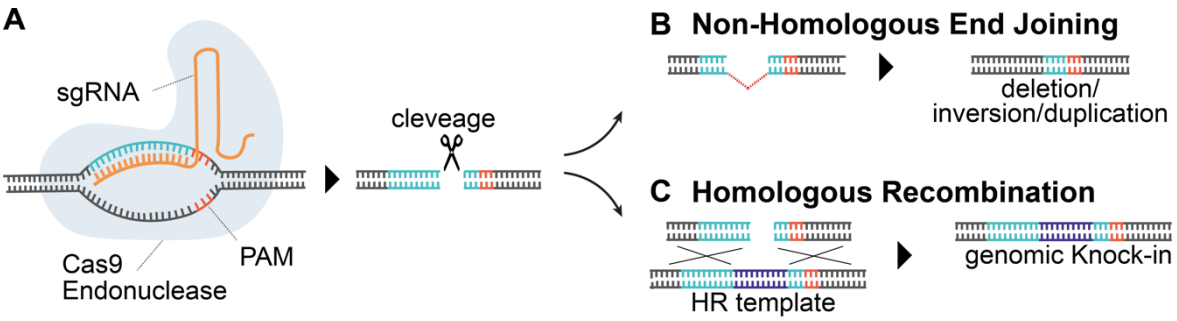


Figure 3 CRISPR/Cas9 genome engineering
Strategies for targeted genome engineering using CRISPR/Cas9 **A.** CRISPR/Cas9 creates targeted double strand breaks using a Cas9 nuclease complex combined with a sgRNA. The sgRNA is around 20 nucleotides long and contains a protospacer-adjacent motif (PAM) sequence that defines the cutting site of Cas9. The introduced double stand break can be repaired through either Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR) (Scully et al. 2019; Kass and Jasin 2010). **B.** In the more error-prone NHEJ DNA ends at the cleavage site are re-ligated whereby structural variations including deletions, insertions and duplications can occur (Kass and Jasin 2010). Through the use of two sgRNAs one can aid the generation of desired structural variants at regulatory landscapes (CRISVar) (Kraft et al. 2015). **C.** In HDR the second allele is used as template to repair the double stand break. As such this mechanism can be exploited to generate knock-ins by providing templates containing the transgene flanked by homology regions of the insertion site (Scully et al. 2019; Kass and Jasin 2010).

In summary, there are numerous tools to map and test the functional components of regulatory landscapes, CREs. This led, based on numerous reports systematically profiling CREs, to the estimation that the human genome overall contains hundreds of thousands of enhancers which by far outnumbers the ~ 20.000 protein-coding genes (Pennacchio et al. 2013). Of these, a subset is active in a given tissue. For example, examining CREs in livers across 20 mammalian species revealed that enhancers quickly change between species and that on average 12.500 promoters and 22.500 enhancers are active (Villar et al. 2015). Combined, this reveals that a single developmental gene is controlled by multiple enhancers enabling an enormous combinatorial complexity. Further the fast-evolutionary turnover of enhancers suggests they can be fertile targets to modulate gene activities (Long et al. 2016b). However, this leads to the question of how newly evolved and pre-existing enhancers act together to define a gene´s activity? To better understand this question, next I introduce the modular behavior of enhancers and how they are altered in disease and evolution.

1.3 Enhancers act modular to control gene expression

Many genes are regulated by multiple enhancers with distinct activities that operate together additively, redundantly or co-operatively and/or inter-dependently (Fig. 4)

(Thomas et al. 2021; Osterwalder et al. 2018). For example, *Shh* expression in multiple tissues is driven by the composite of at least 11 enhancers (Anderson et al. 2014). Loss of the ZRS enhancer eliminates specifically *Shh´s* limb expression demonstrating a single enhancer drives a distinct subset of the overall expression pattern (Anderson et al. 2014). However, most situations are far more complicated. As such, 9 enhancers driving *Ihh* expression in the limbs, growth plates and skull sutures display an additive and redundant behavior (Will et al. 2017). Nevertheless, these enhancers are not entirely interchangeable demonstrating complex interdependencies between them (Will et al. 2017). Similarly, redundant behaviors and hierarchical functional importance of enhancers has been observed at other loci (Shin et al. 2016b; Montavon et al. 2011; Osterwalder et al. 2018). Finally, using synthetic regulatory genomics to design deletions, inversions and translocations of enhancers at the LCR driving *Sox2* expression revealed a complex interplay of 27 DHS elements including inter-dependent behaviors (Brosh et al. 2022). As such, the enhancers DHS19+20 and DHS25+26 themselves drive only little *Sox2* expression but when linked together overall *Sox2* expression is increased more than their additive activities (Brosh et al. 2022). Thus, collectively, multiple enhancers modularly coordinate the accurate regulation of genes which when altered can drive disease.

Depending on the mode of action, disrupting enhancers can have different effects. Where redundant, so called “shadow” enhancers, are present, elimination of enhancers has no affect. For example, nine individual deletions of conserved enhancers, proximal to genes associated with congenital limb malformations, caused no obvious phenotype indicating they are dispensable for limb morphogenesis (Osterwalder et al. 2018). However, when acting cooperatively, lost or gain of elements can drive diverse pathologies and in fact, most disease-causing mutations are found in enhancer sequences (Visel et al. 2009; Tak and Farnham 2015; Maurano et al. 2012). For example, microdeletions in an enhancer driving *Sox9* expression is associated with the Pierre Robin syndrome while duplicating another *Sox9*-enhancer causes human sex reversal (Benko et al. 2009; Croft et al. 2018). Additionally, sometimes already epigenetic modifications alone or alteration of a critical part of enhancer can be sufficient to cause strong effects. As such, aberrant enhancers methylation in neurons negatively influences the pathological onset of Alzheimer and even single nucleotide changes in the ZRS enhancer drives a broad range of limb malformations (Furniss et al. 2008; Lettice et al. 2003b) (Li et al. 2019).

In addition to their importance for proper development, enhancers also represent an ideal playground for evolutionary modifications of gene activities. Indeed, modulating enhancers seems to be effective and simultaneously “safe”, as the tissue-specificity and pervasive redundancy of enhancers likely limits the risk for lethal genetic alterations (Long

et al. 2016b; Levine 2010; Wittkopp and Kalay 2011). Accordingly, enhancers display a high evolutionary turnover where they are frequently lost, gained or modified driving phenotypical traits (Villar et al. 2015; Schmidt et al. 2010). As such, enhancer losses contributed to new body plans including the pelvic elimination in different stickleback fish populations and limb-less snakes (Chan et al. 2010; Kvon et al. 2016). Likewise, an enhancer duplication and fusion leading to increased androgen production contributed to the evolutionary beneficial masculinizing in female moles (Real et al. 2020). Finally, endogenous retroviruses drove the lineage specific emergence of enhancers carrying SRF motifs that allows them to respond to female hormones and thus likely contributed to mammalian placenta evolution (Sun et al. 2021).

Thus, collectively, the accurate regulation gene activities is driven by an assembly of multiple enhancers which therefore provide a rich source for evolutionary innovations and diseases in humans.

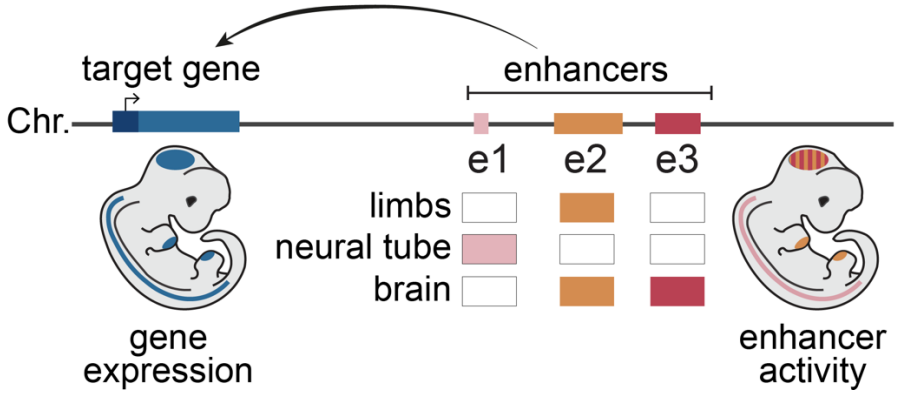


Figure 4 Gene activities are controlled by multiple enhancers
Complex developmental gene expression is controlled by a composite of multiple enhancers (e1, e2, e3) that act over very long distances (Lettice et al. 2003a; Spitz et al. 2003). These enhancers can act additively, redundantly and/or co-operatively to drive a subset of the overall gene expression pattern.

1.4 Enhancer-promoter specificity

Large regulatory landscapes, few genes and many enhancers: How does the genome control which enhancers interact with which gene? Indeed, over 60% of developmental enhancers do not regulate their nearest genes, but instead a more distal gene suggesting that a mechanism exists dictating the desired communications (Chen et al. 2022). One such mechanism that could determine a correct pairing is the widely and controversially discussed specificity mechanism that acts as a lock-and-key principle for enhancers and promoters. Indeed, in drosophila an intrinsic selectivity for some enhancers and promoters has been documented (Li and Noll 1994; Merli et al. 1996). As such, *dpp* but not the more proximal *oaf* gene responds to *dpp*-enhancers. However, replacing *oaf* ´s promoter with

that of *dpp´*s results in *oaf* expression clearly demonstrating a strict compatibility (Merli et al. 1996). Similarly, in mammals a selective compatibility for some distinct enhancer-promoter pairs might exist (Jing et al. 2008; Bertolino and Singh 2002; Vakoc et al. 2005). Moreover, MPRAs in drosophila revealed that promoters of developmental and housekeeping genes show more than tenfold preferences towards different classes of enhancers (Arnold et al. 2017); (Zabidi et al. 2015). Further, these two promoter classes recruit different transcription cofactors and differ in their mechanisms for transcription initiation (Haberle et al. 2019); (Serebreni et al. 2022). Thus, in Drosophila, different biochemical features of distinct enhancers and promoters inherently generates a degree of specificity.

Unlike drosophila, the extent to which enhancer-promoter compatibility operates in mammals is still hotly debated. Recently, two studies used MPRAs to systematically address this in mammals and came to contradicting conclusions (Martinez-Ara et al. 2022); (Bergman et al. 2022). While Bergman et al. report that most enhancers activate all promoters to similar degrees, Martinez-Ara et al. found that thousands of enhancers and promoters display intrinsic compatibilities (Martinez-Ara et al. 2022; Bergman et al. 2022). Both studies noted a separation of promoters into a housekeeping and developmental classes (Martinez-Ara et al. 2022; Bergman et al. 2022). Martinez-Ara et al. found that such a classification cannot explain their observed specificity (Martinez-Ara et al. 2022). In contrast, Bergman et al. reports generally no compatibilities, but noted subtle differences between enhancers and promoters of these two classes which are qualitatively not comparable to differences observed in Drosophila (Bergman et al. 2022). Thus, enhancer-promoter specificity is an appealing idea that certainly would allow evolution to readily incorporate new information without perturbing pre-existing relationships, but currently it is unknown to what extent this operates in mammals.

2. The 3D genome is structured at multiple scales

So far, this chapter has introduced the building blocks of regulatory landscapes, their properties and evolutionary modifiability. Yet, understanding how chromatin landscapes control gene activities and their changes in evolution, requires consideration of their 3D structure. The following section introduces the methods used to study chromatin structure, followed by how enhancer-promoters communicate through loop formation in 3D space. Afterwards, TADs, LADs and compartments and how these different structural features influence gene regulation will be described.

2.1 Methods to study chromatin structure

Methods to analyze the 3D genome includes sequencing-based genomic technologies, microscopy and computational approaches. Genomics approaches have been particularly significant, allowing mapping of chromatin's spatial positioning within the nucleus through technologies such as TSA-seq, pADamID and DamID-seq. (Chen et al. 2018b; Vogel et al. 2007; van Schaik et al. 2020; van Steensel and Henikoff 2000). Further chromatin-chromatin interactions can be studied through a variety of approaches, of which chromosome conformation capture techniques (C-technologies) including GAM, SPRITE and HiC which have become most prevalent (Dekker et al. 2002; Dekker et al. 2013; Kempfer and Pombo 2020). Specifically, HiC generates genome-wide contact matrixes which reflect the interaction frequency of chromatin regions revealing many structural features of the genome (Fig. 5) (Jerkovic and Cavalli 2021).

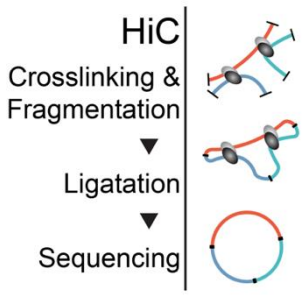


Figure 5 Principle of HiC
HiC is a sequencing-based method that assays chromatin interactions. DNA fragments in close proximity in the 3D nuclear space are cross-linked and fragmented by digesting enzymes. Afterwards free chromatin ends that are in close physical proximity are ligated to generate a chimeric sequence. The abundance of chimeric DNA sequences coincides with the relative contact frequency of these sequences. Following sequencing a genome-wide contact matrix can be generated displaying chromatin interactions.
There are multiple variations of C-technologies. Here used ones are: genome-wide HiC and capture-HiC where a contact matrix is generated for only a distinct genomic region. For review of C-technologies (Jerkovic and Cavalli 2021; Dekker et al. 2013); technical details see methods 15.

Though capable of single cell analyses, the low resolution of these approaches limits C-technologies to average readouts from millions of cells (Nagano et al. 2015; Nagano et al. 2013). Instead, imaging approaches are increasingly deployed to map single allele structures with ever-improving labeling methods and resolution (Maslova and Krasikova 2021). Unlike Hi-C, imaging using fluorescence in situ hybridization (FISH) of DNA with fluorescently labelled probes that hybridize to genomic sequences can be applied directly on cells to visualize structures (Kempfer and Pombo 2020). As such, the use of oligopaint probes and sequential labelling chromatin allows to trace entire chromosome structures (DNA-MERFISH) (Nir et al. 2018; Beliveau et al. 2012; Su et al. 2020; Nguyen et al. 2020; Sawh and Mango 2020). Moreover, recent coupling of such chromatin tracing with RNA-MERFISH and immunofluorescence staining allowed the simultaneous visualization of chromatin, RNA and the nuclear lamina in mammalian cells (Liu et al. 2020; Liu et al. 2021a). Finally, computational approaches that model the chromatin fiber are frequently used to predict

and test the mechanisms shaping chromatin structure (Esposito et al. 2021). Combined with sequencing approaches, these methods have revealed a multifaceted structural organization of genomes.

2.2 Enhancer-promoter communication requires proximity

Just how enhancers select their target genes is an open question, so too is it unclear how enhancers can communicate with promoters that are positioned megabases away (Lettice et al. 2003a). However, now, accumulating evidence suggests that enhancers activate genes by moving into their close physical proximity though chromatin looping. For example, the distal enhancer locus control region (LCR) loops and sequentially contacts different globin genes to progressively activate them in development (Deng et al. 2014; Palstra et al. 2003). Accordingly, artificially forcing the LCR to contact the *ß-globin* gene at the incorrect time leads to its inappropriate activation, thereby suggesting contact is crucial for transcription (Deng et al. 2012). Similarly, live cell imaging in Drosophila demonstrated that transcription occurs when proximity between labelled enhancers and promoters is achieved (Chen et al. 2018a). Moreover, imaging of the ZRS enhancer and its target gene *Shh* revealed their constant physical proximity within a range of ~ 400 nm independent of their activity state (Williamson et al. 2016). However, when active, *Shh* and ZRS move even closer to within ~ 200 nm (Williamson et al. 2016). Here, physical proximity of ZRS and *Shh* might be so important for transcriptional activation that they move into close vicinity even prior to activation in preparation for (or following) gene activation. Combined, these studies demonstrate that physical proximity is important for enhancer-promoter communication.

Despite this, the relationship between proximity and transcription may be non-linear or even indirect. At the *Sox2* locus, transcriptional activity did not correlate with immediate proximity suggesting that direct physical contact per se is not required for transcription (Alexander et al. 2019). Likewise, increased average separation has been observed between *Shh* and its ZRS enhancer upon activation (Benabdallah et al. 2019). In both cases, this may indicate that enhancer-promoter pairs need only to be in a permissive physical range that is sufficient for transcriptional activation.

Although widely accepted that proximity is important for enhancer function, the exact mechanism through which their proximity induces gene transcription is unknown (Beagrie and Pombo 2016). The classical view so far is that functional enhancer-promoter contacts involve Mediator; a conserved multi-protein complex that rapidly evolved in eukaryotes to likely increase regulatory capacity (Soutourina 2018). Upon its recruitment to promoter and enhancer sites, Mediator acts as a physical bridge between TFs bound at

the enhancer and the PIC at the promoter (Soutourina 2018) (Fig. 6A). As such, the Mediator is involved in the formation of chromatin loops and, once present, recruits RNA Pol II and supports transcription initiation and elongation through as yet unclear mechanisms (reviewed in (Soutourina 2018; Richter et al. 2022). Combined, enhancer-promoter proximity at many loop anchors involves bridging by the Mediator. However, recently an alternative model has been proposed for enhancer function; liquid-like condensates.

Liquid-like condensates has been suggested to couple the formation of classical chromatin loops with the generation of a microenvironment that facilitates enhancer function (Hnisz et al. 2017) (Fig. 6B). These dynamic and reversible condensates are formed by the de-mixing specific biomolecules into drop-like membrane-less compartments (Rippe 2022). The formation of such condensates has been demonstrated for many transcription-associated proteins carrying intrinsically disordered regions including TFs, RNA-Pol II, chromatin remodelers and Mediator (Boehning et al. 2018); (Cho et al. 2018); (Chong et al. 2018); (Sabari et al. 2018); (Lu et al. 2018); (Fasciani et al. 2020); (Shi et al. 2021); (Daneshvar et al. 2020). Interestingly, it is proposed that the Mediator might have been evolutionarily optimized for its property to form such condensates (Richter et al. 2022); (Nagulapalli et al. 2016). As such, condensates may allow for less strict physical proximity and the simultaneous communication of multiple enhancers with several genes (Kyrchanova and Georgiev 2021; Hsieh et al. 2020). Regardless of their mechanism of formation or action, enhancer-promoter contacts reflect an important feature to implement regulatory evolution.

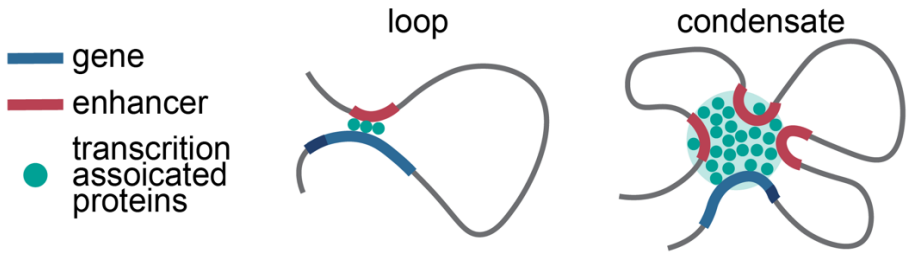


Figure 6 Enhancer promoter communication
Left: Classical view where the enhancer is brought into physical proximity to its target gene by the formation of a chromatin loop. The loop anchor is decorated with proteins required for transcription which includes TFs, co-activators, RNA Pol II and the Mediator serving as indirect physical bridge between enhancer and promoter (Soutourina 2018). **Right**: Emerging view where chromatin loops are combined with the formation of condensates. The condensates contain transcription-associated proteins and create at the multi-loop anchor a microenvironment which likely enhances the transcriptional process (Hnisz et al. 2017).

2.3 TADs constrain enhancer-promoter communication

As enhancers can regulate genes found far along the chromatin fiber, additional features are needed to restrict their activities to only the correct target gene(s). TADs represent one such feature that achieves this. TADs are chromatin regions of preferred self-interaction ranging from 100kb to several megabases in size (Dixon et al. 2012; Nora et al. 2012). TADs are separated from one another by boundary regions, which are typically enriched by convergently orientated (facing ><) CTCF binding sites (Merkenschlager and Nora 2016). Alternatively, boundaries can be formed by highly transcribing genes or retrotransposon elements (Zhang et al. 2019; Zhang et al. 2020a; Bonev et al. 2017). Accordingly, imaging approaches have shown TADs as discrete physical units (Bintu et al. 2018; Szabo et al. 2018).

Functionally, TADs appear to facilitate transmission of enhancer activities while simultaneously restricting them to their correct target gene(s). Supporting this, enhancer mobilization experiments have demonstrated that the 2-3 fold higher contact frequency found within a TAD permits enhancer-promoter communication (Zuin et al. 2021). However, the reduced contact frequency found beyond a TAD's boundaries blocks it. Consistently, sensors (minimal promoters + reporter gene) which were integrated throughout the mouse genome produce similar expression patterns when found in the same TAD but not in neighboring domains (Fig. 2B) (Despang et al. 2019; Symmons et al. 2014). In other words, 3D TAD structures transmit and compartmentalize regulatory information, resulting in two distinct effects. Namely, genes within the same TAD tend to be co-regulated and the regulatory effects of sequence variants are delimited to genes within TAD boundaries (Le Dily et al. 2014; Shen et al. 2012; Delaneau et al. 2019). Importantly, these properties grant TADs a crucial regulatory role in evolution and disease.

TADs can change in disease and evolution

Disrupting TADs can have profound pathogenic consequences. Specifically, genomic rearrangements affecting TADs can create inappropriate enhancer-promoter communications, leading to gene mis-expression and diseases from developmental disorders to cancers (Anania and Lupiáñez 2020; Spielmann et al. 2018). For instance, inversions containing *Epha4*-enhancers and a TAD boundary results in new assignment of these enhancers to genes. Specifically, while *Epha4* loses its enhancers and their resulting limb expression pattern, *Wnt6* in the neighboring TAD gains them to drive abnormal limb development (Lupianez et al. 2015) (Fig. 7). Such enhancer "adoption" or "hijacking" events are also observed in cancer and other pathologies and might explain ~ 12% of copy-number variation phenotypes caused by deletions affecting TAD

boundaries (Ibn-Salem et al. 2014; Northcott et al. 2014; Xu et al. 2022). Another example is a duplication at the *Sox9* locus which creates a neo-TAD resulting in the inappropriate communication of the *Kcnj2* gene with *Sox9*-enhancers driving a limb malformation (Franke et al. 2016). Collectively, these studies demonstrate the regulatory and clinical importance of TADs.

Nevertheless, not all genomic rearrangements cause gene mis-expression and disease despite creating novel enhancer-promoter pairings. Multiple loci with TAD rearrangements did not display pathogenic gene mis-expression including translocations that repositioned 34 genes near enhancers (Despang et al. 2019; Laugsch et al. 2019; Ghavi-Helm et al. 2019; Dong et al. 2018). Similarly, a recent large-scale analysis of somatic structural variations (SVs) in cancer revealed that only 14% of boundary deletions changes expression of nearby genes by more than two-fold (Xu et al. 2022). Thus, it is clear that additional mechanisms must be in place to direct enhancer activities to their targets to sustain gene expression in rearranged TADs.

Given their importance in proper development of an organism, TADs are generally believed to be stable in evolution. Indeed, many specific TADs have been reported to be conserved across species (Dixon et al. 2012; Harmston et al. 2017; Gómez-Marín et al. 2015; Renschler et al. 2019). Likewise, analyzing rearranged breakpoints across species showed that TADs are more likely to be reshuffled as a whole regulatory unit rather than being broken during vertebrate evolution (Krefting et al. 2018; Vietri Rudan et al. 2015). Thus, several TADs, particularly those containing developmental genes, are conserved and this is likely a product of purifying selection. Nevertheless, changes to TADs are also an ideal source for new traits in evolution (Acemel et al. 2016; Acemel et al. 2017). For instance, an inversion at the *Fgf9* locus drives higher *Fgf9* expression and is associated with gonadal masculinization in female moles (Fig 7) (Real et al. 2020). Likewise, TAD-rearrangements in humans relative to macaque monkeys are associated with altered gene expressions, which is potentially related to our greater neurological complexity (Luo et al. 2021). Thus, collectively the effects of rearranged TADs for gene activities cannot yet be reliably predicted as we lack a full picture of all mechanisms governing gene expression.

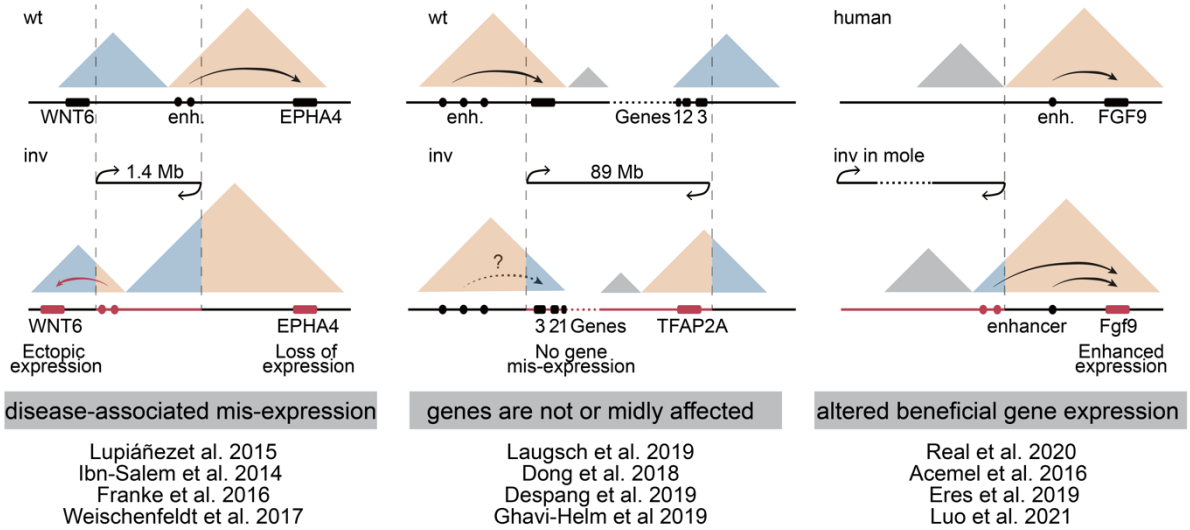


Figure 7 Predicting effects of genomic rearrangements is difficult
To predict the effects of genomic rearrangements including deletions, inversion, duplication that affect TADs is still challenging. Such rearrangements can be: disease-causing (A), unnoticed/ cause only minor expression effects (B) or are beneficial for novel traits in evolution (C). **A.** An inversion at the *Epha4* locus relocates its enhancers into the *Wnt6* TAD. The lack of enhancers in the *Epha4* TAD causes *Epha4*-loss-of-function while their gain in the *Wnt6* TAD causes a *Wnt6*-gain-of-function resulting in abnormal limb development (Lupianez et al. 2015). **B.** An inversion in patients with Branchio-Ocular-Facial Syndrome (BOFS) creates new enhancer-promoter parings in two TADs. The *Tfap2a* gene loses its enhancers which leads to its downregulation causing BOFS. The three genes 1,2,3 (MMS22L, NDUFAF4, GPR63) are placed into the other shuffled TAD containing active *Tfap2a*-enhancer, which has unexpectedly no effect on their expression (Laugsch et al. 2019). **C.** An inversion in moles compared to humans brings additional gonadal enhancers into the Fgf9 TAD increasing *Fgf9* expression which likely contributed to the beneficial masculinization of female mole gonads (Real et al. 2020).

Loop extrusion drives TAD formation

In the current prominent model, TADs are established by loop extrusion, a chromatin structuring force created by the interplay of the insulator protein CTCF and the SMC protein cohesin. Here, cohesin progressively extrudes chromatin to create a growing loop until it is blocked by CTCF or unloaded by WAPL (Fig 8) (Fudenberg et al. 2016; Sanborn et al. 2015). The dynamics of loop extrusion has been directly visualized for human cohesin *in vitro* at an average rate of 0.5 kbp/sec and up to 2.1 kbp/sec (Davidson et al. 2019; Kim et al. 2019). As a result, the TADs as observed by bulk Hi-C actually reflect the population average of ensemble structures adopted by loci (de Wit 2020). In other words, HiC TADs are the product of nested loop extrusion between boundaries and as such represent a preferential behavior of chromatin interactions occurring at a locus over time, rather than physical structures per se.

Multiple lines of evidence support this dynamic behavior. First, the formation of chromatin loops and TADs is dynamic with cohesin residing on the chromatin fiber for ~ 20 min due to frequent loading and unloading (Hansen et al. 2017). Further live-cell imaging of chromatin dynamics demonstrated that TAD loops are transient. On average

they exist for ~5 min and occur every ~10 min during an ~15 hours cell cycle in ESCs (Mach et al. 2022). However, chromatin loops stabilized with convergent CTCFs, occur more frequent, are less variable, and persist for ~16 min (Mach et al. 2022). This matches the median live time of 10-30 min observed for the CTCF-stabilized *Fbn2* loop in another study (Gabriele et al. 2022). Likewise, polymer simulations estimate cohesin-mediated loops stabilized by CTCF to last for 5–15 min existing ~20–31% of time while non-CTCF stabilized interactions last less than a minute and occur only 1-3% of the time (Mach et al. 2022). In summary, loop extrusion is a dynamic and variable process, where constraints it through barrier-forming proteins allow for the enrichment of desired enhancer-promoter interactions (Mirny et al. 2019).

The loop extrusion hypothesis allowed to experimentally test a number of predictions by eliminating components of the process. Specifically, elimination of the blocking factor CTCF leads as predicted to unconstrained loop extrusion consequently TADs are lost genome-wide (Nora et al. 2017; Wutz et al. 2017; Szabo et al. 2020). When the unloading factor Wapl is missing cohesin´s residence time on chromatin increases which leads to the formation of larger loops that accumulate and form elongated structures (“Vermicelli”) reminiscent of mitotic chromosomes. (Wutz et al. 2017; Haarhuis et al. 2017; Tedeschi et al. 2013; Kueng et al. 2006). Finally, when eliminating loop extrusion itself, by depleting cohesin or its loading factor Nipbl, all loops and TADs disappear in HiC and imaging approaches (Rao et al. 2017; Szabo et al. 2020; Schwarzer et al. 2017). However, interestingly when TADs are lost, HiC interactions above TADs persist an become even more prominent revealing structural features, namely compartments, build by a different force (Goel et al. 2022; Hsieh et al. 2021; Rao et al. 2017; Schwarzer et al. 2017). Thus, collectively all experimental evidence supports the loop extrusion model which in interplay with other forces shapes the chromatin topology to structurally guide enhancer-promoter interactions.

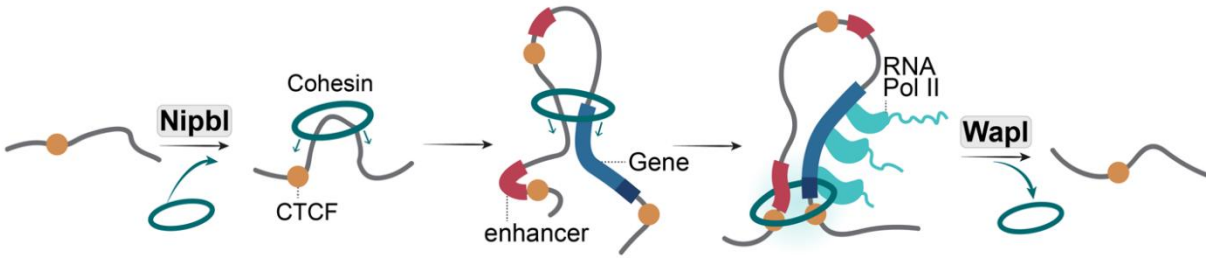


Figure 8 The loop extrusion model
Upon loading by NIPBL, cohesin moves along the fiber and extrudes a chromatin loop which grows in size it reaches a CTCF where it is stalled or unloaded by WAPL. Cohesin blocking occurs often at TAD boundaries or at the loop anchor of enhancer-promoter contacts (Fudenberg et al. 2016; Sanborn et al. 2015).

2.4 Compartments separate active and inactive chromatin at large scale

Chromosomes are divided into higher-order structural features, so-called A- and B-compartments, that provide profoundly different chromatin environments for gene. Compartments are detectable microscopically and manifest above TADs in Hi-C maps as an alternating checkerboard or plaid pattern in which regions of the same type (A-A, B-B) self-interact, as revealed by PCA analysis (Lieberman-Aiden et al. 2009; Rao et al. 2014; Nir et al. 2018; Wang et al. 2016). Such interactions partition chromatin by activity into A (active) and B (inactive), which can be further sub-divided into A1, A2, B1, B2, B3, B4 sub-compartments in higher-resolution Hi-C (Rao et al. 2014; Lieberman-Aiden et al. 2009). Matching this spatial separation, such A and B compartments also differ in gene density, replication timing, epigenetic signature, and differing strengths of chromatin-chromatin interactions (Belmont 2022); (Rao et al. 2014); (Belaghzal et al. 2021); (Chen et al. 2018b). As such, compartments reflect chromatin in fundamentally distinct biochemical environments for gene regulation that extensively reconfigure in development. Indeed, 1/3 of the genome switches compartments from A to B or vice versa concomitantly with changes in gene transcription during development (Bonev et al. 2017; Sima et al. 2019; Fraser et al. 2015; Dixon et al. 2015). In short, compartments reflect a higher order organization between which regulatory landscapes must transition.

While TADs are formed by loop extrusion, the mechanisms establishing compartments are less clear. In recent years, biophysical concepts of self-organizing phase separation were proposed to drive the segregation of active and inactive chromatin (Hildebrand and Dekker 2020; Erdel and Rippe 2018). One model suggests that phase separation might occur through bridging factors that connect chromatin regions of the same type (Hildebrand and Dekker 2020). Such a bridging function has been proposed for HP1 alpha which connects heterochromatic regions by binding two H3K9me3 nucleosomes, thereby likely promoting the formation of B compartments (Machida et al. 2018; Singh and Newman 2020; Kilic et al. 2018; Hiragami-Hamada et al. 2016). Similarly, analogous bridging functions are suggested for the architectural proteins YY1 and PRC1 (Weintraub et al. 2017; Isono et al. 2013). Despite this, an alternative model suggests that compartments are driven by liquid–liquid phase separation generated through weak multivalent interactions between proteins with intrinsically disordered regions (Romero et al. 2001; Vucetic et al. 2003). Interestingly, strong evidence suggests a number of nuclear bodies including the nucleolus, are established through this mechanism (Lafontaine et al. 2021). However, unfortunately, testing these mechanisms as establishing compartments is extremely challenging due to the intrinsic relationship between chromatin's activity and their structuring in compartments (Hildebrand and Dekker 2020). As the mechanism of

formation remains unclear, likewise it is not yet understood how their profoundly different chromatin states may influence gene regulation.

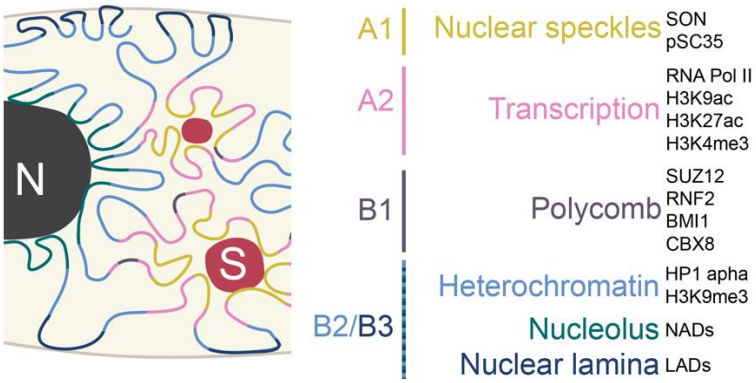


Figure 9 Chromatin sub-compartments
Left. Schematic representation of the spatial position of sub-compartments relative to nuclear landmarks: lamina of the NE (L), transcription associated speckles (S) and the nucleolus (N). Sub-compartments differ in their epigenetic features, DNA replication timing and different strengths of intrinsic chromatin-chromatin interactions (Chen et al. 2018b; Belaghzal et al. 2021; Rao et al. 2014). **Right.** List of selected known features that correspond to distinct sub-compartment: A1: Loci near the speckle-associated proteins pSC35 and SON; A2: histone marks of transcriptionally active loci and RNA Polymerase II; B1: polycomb group proteins CBX8, BMI; B2/B3: heterochromatin associated histone mark H3K9me3 and HP1apha protein, LADs and nucleolus associated domains (NADs) (Belaghzal et al. 2021).

2.5 Chromatin interactions with the NE create a repressive environment

A large fraction of B compartments attaches to the NE in lamina-associated domains (LADs) (van Steensel and Belmont 2017; Guelen et al. 2008; Reddy et al. 2008; Finlan et al. 2008). LADs can be mapped by DamID-seq and are generally classified according to their NE-association across different tissues (Fig. 10) (Guelen et al. 2008; Peric-Hupkes et al. 2010; Kind et al. 2015; Vogel et al. 2007). Constitutive LADs are enriched for H3K9me2/3 marks and are permanently NE-attached in cell types investigated thus far (Pickersgill et al. 2006; Peric-Hupkes et al. 2010; Guelen et al. 2008). In contrast, facultative LADs are enriched with H3K27me3, and often contain developmental genes which dynamically transition from and to the NE during development, concomitantly with their altered activities (Peric-Hupkes and van Steensel 2010; Leemans et al. 2019; Guelen et al. 2008).

Though is unclear how regions are driven to form LADs, current evidence suggests interactions between heterochromatin and NE proteins play a critical role (Belaghzal et al. 2021; Robson et al. 2016; van Steensel and Belmont 2017; Belmont 2022). For mammalian cells, LAD-anchoring was found to be dependent on at least the partially redundant NE-proteins LBR which binds heterochromatin through HP1 and laminA/C (Holmer and Worman 2001; Solovei et al. 2013). As such, loss of LBR in mice results in

an inverted chromatin architecture with euchromatin A compartments replacing heterochromatic B compartments at the periphery (Solovei et al. 2013). Excitingly, this inverted genome organization occurs uniquely in the rod receptors of nocturnal mammals where it acts as an evolutionary adaptation to enhance retinal light transmission (Solovei et al. 2009; Subramanian et al. 2019). Thus, this modification highlights the alteration of chromatin structure as a phenotypic innovation during evolution. However, more generally, the fact that B compartments persist in inverted nuclei highlights that NE-attachment and compartmentalization are related but distinct structural features of chromatin.

As NE-association strongly correlates with gene activities, many studies have addressed their repressive potential (Lund et al. 2013); (Robson et al. 2016); (Robson et al. 2017). On one hand, forcing transcriptional activation of promoters in LADs drives NE-release (Brueckner et al. 2020); (Therizols et al. 2014). However, on the other hand, artificial tethering to the NE drives transcriptional repression of some promoters, but not others (Finlan et al. 2008); (Kumaran and Spector 2008); (Reddy et al. 2008). Furthermore, a recent high-throughput screen tested hundreds of promoters for their transcriptional capacity when positioned in LAD or non-LAD chromatin environments (Leemans et al. 2019). Interestingly, hundreds of *LAD-sensitive* promoters were more active when removed from an LAD environment, thereby indicating that heterochromatin can suppress transcription (Leemans et al. 2019). Indeed, insertion of specific genes in LADs promoted a 40-130 fold reduction in expression and failure to achieve this during development causes impaired repression and differentiation (Leemans et al. 2019; Robson et al. 2016; Poleshko et al. 2017). Nevertheless, other promoters - referred to as *escapers*- were significantly less sensitive to repression in LADs indicating they are not universally repressive (Leemans et al. 2019). In these cases, *escaper* promoters underwent local NE-release indicating locally-restricted chromatin remodeling might be sufficient for their activities (Leemans et al. 2019).

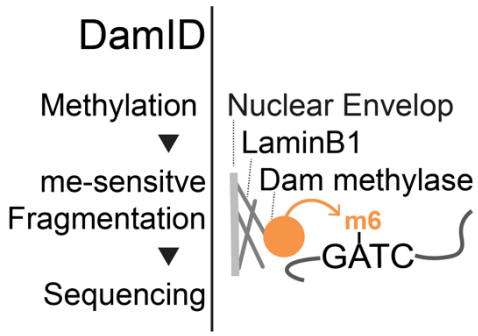


Figure 10 Principle of DamID
DamID is a sequencing-based method that assays chromatin interactions with the NE. DNA adenine methyltransferase (Dam) is fused to nuclear lamin B1 and methylates adenin of NE-proximal GATC sequences. Metylation-sensitive *DpnI* enzymes cuts between 5´TC and 3´GA^{me}. Unmethylated fragments are digested with *DpnII* and thus excluded from following PCR based amplification and preparation for subsequent sequencing.
For details see methods part 14 and (Vogel et al. 2007)

Combined, compartments and LADs are critical components of chromatin structure that are intrinsically related to gene activity. However, whether a gene's activity is a cause

or consequence of its incorporation into a LAD or compartment is unknown. In short, it is an open question what role these structures play in gene regulation or how this can be modified in evolution.

Collectively, complex gene regulation is controlled through multiple interconnected epigenetic and structural mechanisms that all could be modified to adapt a landscape for the emergences of new genes during evolution.

AIM

Modifying or incorporating new traits is essential to evolution but how they are successfully implemented at a genetic level remains poorly understood. Complex gene expression is driven by enhancers and promoters operating within structurally constrained TAD regulatory landscapes (Robson et al. 2019). Accordingly, such regulatory landscapes are often conserved across evolution and disruptions to them can drive diseases (Spielmann et al. 2018; Dixon et al. 2012; Fraser et al. 2015; Harmston et al. 2017). Considering this, it is unclear how new genes and functions can be safely incorporated into pre-existing landscapes during evolution. Although an astounding number of potential epigenetic and structural mechanisms exists which could influence genome evolution many unknowns remain. Do regulatory conflicts arise if a new gene emerges in a TAD but requires completely different regulatory needs as the already existing genes? If so, which of the numerous regulatory mechanisms are used to then resolve such conflicts? How are existing chromatin landscapes modified to deal with the evolutionary novelties? In this thesis, I address these questions by dissecting the mechanisms allowing a new gene, *Zfp42*, to emerge within an ancient locus without adopting or disrupting the functions of its pre-existing gene, *Fat1*.

MATERIALS & METHODS

Indicated and indented parts are taken from Ringel et al. 2022

1. Materials

Chemicals

If not stated differently, standard chemicals were obtained in analytical grade quality from Roth, Merck or Sigma-Aldrich.

Bacterial strains

General cloning was performed using chemically competent E. coli Top10 cells (internal production by Asita C. Stiege, MPIMG or Thermo Fisher, #C404006)

Software

Software programs and algorithms with corresponding source are listed in table 2.

2. Standard laboratory procedures

2.1 Molecular biological methods

All molecular standard procedures including cloning, bacterial transformations, gel electrophoresis, PCRs were performed according to standard protocols from Green and Sambrook (Green and Sambrook 2012).

2.2 Isolation of DNA and RNA

Isolation of genomic DNA

Cells and ear biopsies from adult mice were lysed in 50 µl lysis buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS freshly supplemented with 0.08 µg/ µl proteinase K (Roche Diagnostics, #1000144)) and incubated for 4-16 h at 56°C. Afterwards 25 µl 5M NaCl were added and mixed with two volumes of 100 % ethanol. After incubation for 30 min at -20°C., DNA was pelleted by centrifugation (max speed / 30 min / 4 °C). The pellet was washed twice with 70 % ethanol, air-dried and afterwards dissolved in 10 mM Tris/HCl.

Isolation of plasmid DNA

Plasmid DNA was isolated using QIAprep Miniprep Kit (Qiagen, #27104) according to the manufacturer's guidelines.

Isolation of BAC DNA

Isolation of BAC DNA was performed using the NucleoBond BAC 100 Kit (Machery-Nagel, #740579) according to the manufacturer's instruction.

Simultaneous isolation of RNA and genomic DNA

RNA and genomic DNA from E11.5 forelimb samples were simultaneously obtained by using Quick-DNA/RNA Microprep Plus Kit (Zymo, #D7005) according to the manufacturer's instructions.

Isolation of RNA

Snap-frozen mouse limbs or feeder-free ESCs were processed with the RNeasy Mini Kit (Qiagen, #74104) according to the manufacturer's instructions.

3. Animal models

3.1 Mice

Mutant embryos came either from crossing of adult transgenic mice or from morula aggregations, respectively (see 5). Line establishment and maintenance was achieved by crossing with wildtype C57Bl.6/J mice. All mice were housed at the animal care facility of the MPIMG in a centrally controlled environment at 20–22.2 °C, 30–50% humidity combined with a 12 h dark/light cycles. Animal procedures were conducted under the license numbers G0176/19, G0247/13 and G0243/18 issued by LAGeSo Berlin.

3.2 Chicken

Fertilized chicken eggs were received from Valo Biomedia. The fertilized eggs have been incubated at 37.8 °C, 45% humidity and embryos dissected at the developmental stages HH22 and HH24.

3.3 Opossum

Adult opossums (*Monodelphis domestica)* were maintained under the ZH104 permission issued by LAGeSO at the animal care facility of the Museum für Naturkunde, Berlin. Opossums were housed in a centrally controlled environment at 24–26 °C, and 60-65% humidity combined with a 12 h dark/light cycles. Female opossums were euthanized with an Isoflurane overdose under the LAGeSo-issued T0198/13 permission according to national and international standards.

4. 3. Cell culture

Culturing of cell was performed according to standard procedures described in (Behringer et al. 1994) and (Kraft et al. 2015).

4.1 Feeder cells

Tissue culture dishes were gelatinized by adding 0.1 % gelatin/H_20 (Sigma, #G-1393) until the grow area was covered followed by incubation at 37°C for 15 min up to overnight. Afterwards excess liquid was aspirated and Mitomycin-inactivated CD1 mouse embryonic fibroblast were seeded at a density of $4x10^4/cm^2$ in regular Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ ml glucose, without sodium pyruvate (Lonza, #BE12-733F), containing 10 % regular FCS (Biochrom, #S0615), 1x glutamine (Lonza, #BE17-605E) and 1x penicillin/streptomycin (Lonza, #DE17-603).

Feeders were derived from E13.5-E14.5 CD1 and DR4 embryos, cultured until passage 5 and tested for Mycoplasma contamination with the Mycoalert assay control set (Lonza, #LT07-518) and Mycoalert detection kit (Lonza, #LT07-118) (Tucker et al. 1997; Rice and O'Brien 1980). Afterwards feeders were mitotically inactivated with mitomycin C and then frozen at a density of $2.5x10^6$ cells/ vial in medium containing 10 % FCS and 10 % DMSO (Sigma, #D-2650).

4.2 Wildtype and CRISPR modified ESCs

Wildtype and CRISPR modified mouse G4 ESCs (XY, 129S6/SvEvTac x C57BL/6Ncr F1 hybrid) were grown as colonies in a density of $2x10^4$ cells/cm^2 on top of a feeder monolayer in ESC-culture medium at 37 ^oC, 7.5% CO_2 (ESC-culture medium: DMEM (Gibco, #10829-018) supplemented with 15 % FCS (PANSera ES, #P30-2600, Lot 130407ES), 1x glutamine (Lonza, #BE17-605E), 1x penicillin/streptomycin (Lonza, #DE17-603), 1x non-essential amino acids (Gibco, #11140-35), 1x nucleosides (Chemicon, #ES-008D), 0.1 mM β-mercaptoethanol (Gibco, #3150-010) and 1000 U/ ml LIF (Murine Leukemia Inhibitory Factor ESGRO™ (10^7 U/ml) (Chemicon, #ESG1107) (George et al. 2007).The culture medium was changed every 24 h and cells regularly transferred onto new culture dishes using Trypsin. Cells were frozen at densities of 0,5-3 $x10^6$ cells/cryovial in ESC medium containing 20% FCS and 10% DMSO.

4.3 293FT

293FT cells (Thermofisher, #R70007,) were cultured in DMEM (Gibco, #41966-029) supplemented with 20% FCS, 4 mM L-Glutamine, 1x penicillin/streptomycin and used for lentiviral production.

4.4 AID -tagged ESCs

CTCF-AID-GFP and Rad21-AID-GFP E14 ESCs were kindly provided by Elphège Nora (UCFS, San Francisco) These cells were cultured feeder-free on gelatinized dishes in

ESC-culture medium (see 4.2). Degradation of AID-tagged proteins was induced with 500 µM auxin (Abcam, #ab14642) for 48 h and between 1-6 h, respectively, as previously described in Nora et al. and Liu et al. (Liu et al. 2021b; Nora et al. 2017). All cells were frozen at densities of 1.5 x10^6 cells/cryovial in ESC medium containing 20% FCS and 10% DMSO.

4.5 Mouse limb cells culture

Limb cells were derived from wildtype or CRISPR-modified E11.5 embryos. Limbs were dissected in ice cold PBS and dissociated with trypsin at 37°C to receive a single cell suspension. Cells were filtered (40 µm), centrifuged (300 g / 3 min / 4°C) and resuspended in DMEM/F12 supplemented with 10% FCS, 4 mM L-Glutamine, 1x penicillin/streptomycin, 250 ng/ml Recombinant Mouse Wnt-3a protein (R&D Systems, #1324-WN-010/CF) and 150 ng/ml Recombinant Human/Mouse FGF-8b Isoform (R&D Systems, #423-F8-025/CF). Following cells were plated onto gelatinized dishes and incubated at 37^{o}C in 5.0% CO2.

4.6 Opossum fibroblasts

Access to adult Opossums was provided by Peter Giere (Museum für Naturkunde, Berlin, Germany). Opossum embryonic fibroblast cells were derived from stage 30 embryos. Embryos were dissociated, their heads removed and the remaining body dissociated with trypsin. Cells were filtered (100 µm), centrifuged (300 g / 3 min / 4°C) and cultured in DMEM (Lonza, #BE12-733F), containing 10 % FCS (Superior, Biochrom, #S0615), 1x glutamine (Lonza, #BE17-605E) and 1x penicillin/streptomycin (Lonza, #DE17-603) at 32^{o}C in 7.0% CO2. Medium was changed every 2 – 3 days. When passaging these lines, confluency was never below 70%. All cells were frozen at densities of 1.5 x10^6 cells/cryovial in culture medium supplemented with 20% FCS and 10% DMSO.

5. Morula Aggregation to generate mutant embryos and mice

Mutant embryos and adult mice were generated in diploid and tetraploid morula aggregations. For aggregation mutant ESCs were cultured on a CD1 feeder layer until medium sized colonies (15-30 cells) were grown. These colonies were then used for morula complementation as described by Eakin et al. and Artus et al. (Artus and Hadjantonakis 2011; Eakin and Hadjantonakis 2006). The principle of diploid and tetraploid aggregation is briefly descripted in figure 11.

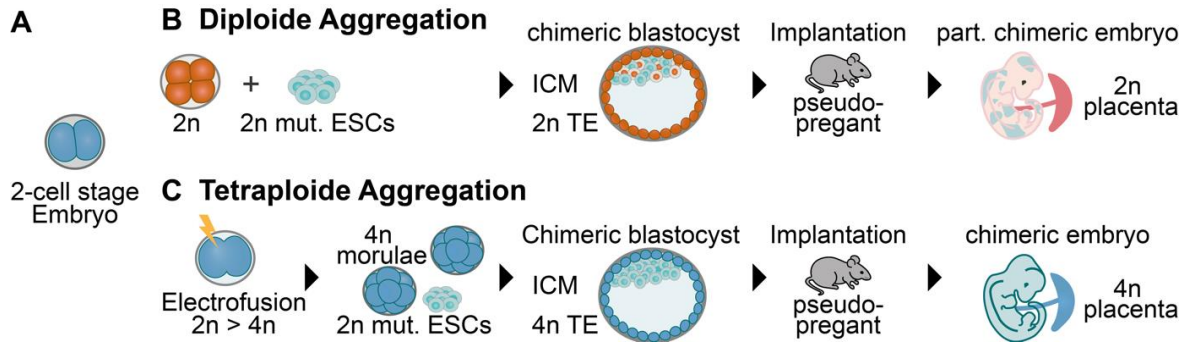


Figure 11 Principle of morula aggregation
A. 2-cell stage embryo after first cell division. **B**. Diploid aggregation: Mutant ESCs are mixed with a wildtype morula in a ratio 1:1. This mixture develops into a chimeric blastocyst where the mutant and wildtype ESCs contribute to the ICM. The blastocyst is retransferred into pseudo-pregnant CD1 female mice where it implants and further develops *in vivo.* Born embryos are partially mutant and wildtype. **C.** Tetraploid aggregation: Electrofusion at the 2-cell-stage generates a tetraploid cell that further develops into a tetraploid morula. In the following step 4n morulae are mixed with diploid mutant ESCs in a ratio of 2:1. This mixture develops into a chimeric blastocyst where the mutant ESCs populate the ICM. Subsequently, the blastocyst is retransferred into pseudo-pregnant CD1 female mice where it implants and further develops *in vivo.* The diploid mutant ICM gives rise to a full chimeric embryo while the tetraploid cells give rise to extraembryonic tissues such as the placenta (Artus and Hadjantonakis 2011; Eakin and Hadjantonakis 2006).

6. Samples preparations

6.1 Preparation of samples for ATAC-seq, ChIP-seq, DamID-seq, cHi-C/HiC

ESCs were trypsinized and heavily feeder depleted. Specifically, feeders and ESCs were separated by resuspending cells in a small volume and by letting the cells settle three times for 40 min each. As feeders settle and attach at the culturing dish quicker than the ESCs the majority of feeders will bet lost when transferring the cell suspension after each 40 min to a new dish. Afterwards a tiny fraction of feeder-free ESCs was plated onto a gelatinized dish to access on the next day based on the cell morphology if all feeders were successfully removed. The majority of cells was pelleted by centrifugation (300 g / 5 min / 4°C) and further processed dependent on the downstream application.

Limb buds of chicken, opossum and mouse were dissected from the embryos in cold PBS. Isolated limbs were then for trypsinized 5 min at 37°C with regular pipetting the solution until no clumps were visible. Afterwards the limb cell suspensions were passed through a 40 µm filter, centrifuged (300 g / 5 min / 4°C) further processed dependent on the downstream application. Also see 9, 10, 14 and 15.

6.2 Preparation of samples for 3D-SIM microscopy

Cover slides (Zeiss 18mm x 18mm, 0.170 ± 0.005 mm) for FISH have been autoclaved and then treated with 0.01% poly-lysine/H_2O for 2-4 hours. After several washes with H_20 and they have been air-tried and were ready to use. Single cell suspensions of E11.5 limbs

or feeder-free ESCs were applied as 80-100 µl drop (1,5-2 x10^6 cell/ ml) onto the coverslip and incubated for 2 h at 37°C, 7.5% CO2. After 2 hcells were fixed with 4% PFA/PBS for 10 min at room temperature followed by several washes with PBS. Coverslips were placed in a cover slip carrier (in-house manufacture by Karsten Beyer, MPIMG) and send to Montpellier where they have been processed within 3 weeks after fixation.

6.3 Preparation of AID-tagged cell lines for c-HiC and FISH

The 48h-induced CTCF-AID-GFP were GFP-negative sorted using FACs and fixed for c-HiC. RAD21-AID-GFP induced for c-HiC fixation were sorted according to GFP signal loss and additional DAPI staining to collect cells outside the G2/M cell cycle phase. Cells induced for microscopy were directly plated onto poly-lysin coated coverslips. Also see 4.4, 13 and 15.

6.4 Preparation of embryos for WISH

Entire embryos were independent of their developmental stage fixed for overnight up to 16 hin 4% PFA/PBS. On the next day, embryos were washed at least three times with PBST (0.1% Tween20/PBS) and step-wise dehydrated. Dehydration was performed at 4°C by incubating the embryos in 25%, 50%, and 75% methanol/PBST for 30 min up to 2 hours. Finally, they were placed in 100% methanol and stored at −20°C until further processing. See 7.5.

6.5 Preparation of embryos for LacZ staining

Transgenic embryos were dissected in cold PBS and dependent on the developmental stage fixed for 20 - 30 min in 4% PFA/PBS rotating at 4°C. After several washes with PBS, LacZ staining was started. See 7.4.

7. Analysis of gene activities

7.1 c-DNA synthesis and qRT-PCR

E12.5 hindlimbs from somite staged embryos were dissected in cold PBS. Samples were snap-frozen and stored at -80°C. For further processing RNA was isolated with the Rneasy Mini Kit (Qiagen, #74104) according to the manufacturer's instructions. Optional DNA digesting with Dnase (Roche Diagnostics, #4716728001) was performed. cDNA was generated using the Superscipt III (Thermo Fischer, #18080093). LacZ mRNA levels were quantified by qPCR using Biozym Blue Sybr Green (Biozym, #331455S) for at least 3 biological replicates per condition each as technical triplicates. 2-ΔΔCt method has been

used for analysis of relative lacZ expression levels. qPCR primers used: qPCR_LacZ_F, 5′-TTCAACATCAGCCGCTACAG-3′; qPCR_LacZ_R, 5′-CGTCGATATTCAGCCATGTG -3′; qPCR_mGAPDH_F, 5′-TCAAGAAfinaGGTGGTGAAGCAG-3′ and qPCR_mGAPDH_R 5′-ACCACCCTGTTGCTGTAGCC-3′. These primers were designed using the online tool Primer3Plus and the efficiency of each primer pair was calculated by preforming a standard dilution series.

Standard cDNA synthesis mix: 1 µl oligodT (50 µM), 1 µl dNTP mix (10 mM), 1 µg RNA, H_20 to a final volumne of 13 µl. This mixture was incubated for 5 min at 65 °C and then immeadetly placed on ice for 2 min. Next, remaining reaction mixture was added: 4 µl 5X First-strand buffer, 1 µl DTT (0.1 M), 1 µl RNaseOUT (40 U/µl), 1 µl SuperScriptIII RT (200 U/µl) and subsequent cDNA was synthesized. Thermocycler: 5 min at 25 °C; 60 min at 50 °C; 15 min at 72 °C. Standard qPCR mix: 6 µl 2X SYBR Green, 2.5 µl forward and reverse primer mix (2.5 µM each), 5.5 µl cDNA (1.5-2 ng/µl). Efficiency of each primer pair was calculated by preforming a standard dilution series.

7.2 RNA-seq

Mouse limbs or feeder-free ESCs were collected and snap-frozen until further processing. Rneasy Mini Kit (Qiagen, #74104) was used to extract total RNA according to the manufacturer's instructions. Optional DNA digesting with DNase (Roche Diagnostics, #4716728001) was performed. Libraries were prepared for sequencing using the KAPA HyperPrep kit for NGS DNA Library Prep (Roche, #7962363001) with poly-A enrichment. Subsequent sequencing on a Novaseq2 was performed as 75 or 100 bp paired-end reads aiming for at least 50 million reads per sample. STAR mapper was used to mapped sequenced reads to the reference genome (mm10) (splice junctions based on RefSeq; options: --alignIntronMin20 --alignIntronMax500000 --outFilterMismatchNmax 10). Differential gene expression analysis was performed using the DESeq2 package from Love et al. (Love et al. 2014). Altered gene expression was considered as being significant with an adjusted P value cut-off of at 0.05.

7.3 Single cell RNA-seq

Three published single cell RNA-seq (scRNA-seq) datasets were used to examine *Triml2*, *Zfp42*, and *Fat1* genes expression. Covering early mammal development data sets from whole embryo gastrulation (Pijuan-Sala et al. 2019), whole embryo organogenesis (Cao et al. 2019) and whole placenta (Marsh and Blelloch 2020) were re-mapped. These data were visualized by Uniform Manifold Approximation and Projection (UMAP) embeddings for the whole placenta and the gastrulation datasets, or t-Distributed Stochastic Neighbor

Embedding (tSNE) for the organogenesis dataset. For clustering of cell types either reported cell type definitions, or ‘‘integrated_snn_res.0.6’’ was used for the whole placenta data. For *Triml2*, *Zfp42*, and *Fat1* UMI counts were plotted in the range 0 to >2. Re-plotting of scRNA-seq data sets was done by Cesar A. Prada-Medina.

7.4 CAGE analysis

CAGE analysis was done by **Dermot Harnett**. The following method part is taken from (Ringel et al. 2022):

> *To calculate gene-expression correlations, we downloaded FANTOM stage 5’ CAGE TPM data (https://fantom.gsc.riken.jp/5/data/). We discarded samples annotated as belonging to ‘reference’ ‘whole body’ or similar samples, and also excluded testis and related tissues from the analysis. We also removed all libraries with fewer than 1 million reads, and all peaks with less than 32 reads across all samples. Overlapping each peak with the Gencode M23 annotation, we assigned peaks to genes if they overlapped a Gencode exon for that gene, or were less than 200bp upstream of a TSS. Peaks not overlapping a gene were discarded, and the counts for all of a gene’s peaks were summed.*
>
> *Since the FANTOM data contained the resulting gene x sample count matrix was then normalized as per (Alam et al., 2020) – normalized counts-per-million for each sample.*

7.5 Spatial mRNA mapping by whole amount in-situ hybridization (WISH)

mRNA expression in embryos was visualized by WISH using digoxigenin (DIG) -labelled antisense RNA probes.

Generation of DIG-labelled *in-situ* probes

Fat1 probes were generated using PCR amplification with cDNA from mouse, chicken or gDNA from opossum embryos as template. *Zfp42* and *Triml* probes were PCR amplified from ESC cDNA. Primers for amplification can be found in table 9. PCR reaction was performed with Tag polymerase to generate A-overhangs that allowed sub-cloning into pTA vector (provided by Asita C. Stiege, MPIMG). After sequencing confirmation, the plasmids DNA served as template to PCR amplification of the probe sequence with SP6 and T7 primers for subsequent *in vitro* transcription. For *in vitro* transcription 200 ng of purified PCR products were used as template combined with DIG RNA labelling mix (Roche, #11277073910), transcription buffer, RNA polymerases SP6 or T7 (Roche, #10999644001) to generate an anti-sense and a sense probe as control, respectively. This mixture was than incubated for 2 h at 37°C. Next, DNA was digested by DNAse I treatment (Roche, #04716728001) for 15 min at 37°C. The reaction was stopped by adding 2 mM

EDTA/ H_2O -DEPC (pH 8.0). Finally, DIG-labeled RNA probes were precipitated with 1/10 volume 0.4 M LiCl and 3 volumes 100% ethanol and subsequent incubation over night at -80 °C. Probes were pelleted by centrifugation (max speed / 20 min / 4°C), washed with 70% ethanol, air tried and resuspended in 10mM Tris.

WISH procedure in embryos

Before starting the WISH procedure embryos were rehydrated by step-wise incubation in 75%, 50%, 25% methanol/PBST (PBST: PBS with 0.1% Tween20) and PBST for each 20 min on ice. In a next step, embryonic pigmentation was eliminated by bleaching the embryos for 1h on ice with 6% H_2O_2/PBST. After washing the embryos in PBST, embryonic tissue was made accessible for DIG-labeled RNA probes with proteinase K (10 µg/ml) treatment for 3 min. Proteinase K treatment was quenched with glycine (2 mg/ml). Following three PBST washes the embryos were re-fixed in 4% PFA for exactly 20 min. Afterwards embryos were incubated for 15 min in L1 buffer (50% formamide, 5XSSC, 1% SDS, 0.1% Tween20 in DEPC-H_2O, pH 4.5) at 68°C. After adding 0.1% transfer RNA (RNA type III from baker's yeast, Sigma, #R6750) and 0.05% heparin (Sigma, #H3149) to the L1 buffer to block unspecific binding sites the embryos were incubated for additional 2 h at 68 °C. Subsequently the DIG- labelled RNA probes were added in a ratio of 1:500 followed by overnight incubation. On the next day, unbound probes were removed with several washing steps: 3 x 30 min in L1 at 68°C, 3 x 30 min in L2 (50% formamide, 2X SSC pH 4.5, 0.1% Tween20) at 68°C, 1 x 20 min in L3 (2X SCC pH 4.5, 0.1% Tween20) at 68°C and 1 x 20 min at RT. Subsequently, RNA is being digested by treating embryos for 1 h with RNase with (0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween20, 100 µg/ml RNase A in H_2O), followed by several PBST washes. Next, embryos were blocked for 2 h at room temperature in blocking solution (1X TBS, 1 % Tween20, 2% serum (Lamb Serum, Gibco, # 16070), 0.2% BSA (Sigmar, #A2153)) followed by incubation at 4 °C overnight in blocking solution containing anti-digoxigenin-alkaline phosphatase antibody (1:5000) (Roche, #11093274910). At the next day, unbound antibodies were removed by excessive washing, 2x 1 h and 2x 30 min in TBST2 buffer (1X TBS, 0.1% Tween20, 0.05% levamisole/tetramisole (Sigma, #L9756)). After this, the embryos are prepared for subsequent staining by incubation for 15 min in alkaline phosphatase buffer (20 m M NaCl, 50 m M MgCl2, 0.1 % Tween20, 0.1 M Tris-HCl, 0.05 % levamisole/tetramisole). Finally, embryos are stained in BM Purple AP Substrate (Roche, #1442074) until the desired coloring is achieved. When staining was completed, embryos were washed with PBST, then fixed in 4 % PFA/PBS supplemented with 0.2 % glutaraldehyde and 5 mM EDTA for long term storage at 4°C. Imaging of embryos was performed in the following 4 weeks

using the ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera.

7.6 Embryonic LacZ staining

The LacZ transcription of reporter constructs were visualized as previously described by Lobe et al. with minor modifications (Lobe et al. 1999). For staining the embryos were kept in the dark at 37°C in lacZ staining buffer supplemented with 0.5 mg/ml X-gal, 5 mM potassium ferrocyanide and 5 mM potassium ferricyanide. When the desired staining was reached the embryo as washed few times in PBS and subsequently fixed with 4% PFA/PBS supplemented with 0.2% glutaraldehyde and 5mM EDTA for long term storage at 4°C. Embryos were imaged within the next 3 weeks using a S SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera.

8. Genome engineering

8.1 Single guide RNA design and cloning

Single guide RNAs (sgRNAs) were designed using the Benchling sgRNA design tool (https://www.benchling.com/). This tool implemented the scoring system from the Feng Zhang lab to predict sgRNAs with a high on target specificity and low off-targeting. Both score values should be as high as possible and in this study used guides had a score of at least 50 or higher (https://zlab.bio/guide-design-resources#score) (Hsu et al. 2013). Two oligonucleotides consisting of a 20-nt specific guide sequence with vector-specific restriction recognition sites as overhangs were annealed, phosphorylated with a T4 Polynucleotide Kinase (Thermo Fisher Scientific, #EK0032). Subsequently the nucleotides were cloned into the pX459 vector (pSpCas9(BB)-2A-Puro) or pX489 (pSpCas9(BB)-2A-GFP) directly upstream of a protospacer adjacent motif (PAM)(Gratz et al. 2013). Each vector contains the CRISPR-associated nuclease Cas9 and trans-activating crRNA (tracrRNA). Prior to using the sgRNAs their sequence was confirmed by sanger sequencing using a vector specific primer (ColR: CACGCGCTAAAAACGGACTA).

8.2 CRISPR/Cas9-based genome editing in ESCs

Desired genome modifications in ESCs were acquired by using the CRISPR/Cas9 method and following the steps displayed in figure 12.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
PREPARATION	prepare CD1 monolayers for G4 ESCs	seed 3.5-4 x10^5 G4 cell per 6cm dish	remove 2 h before transfection penicillin/streptomycin	prepare DR4 feeders: 2x 6cm per CRISPR	start puromycin selection	puromycin selection	stop puromycin selection after 48h; start recovery	24h recovery	48h recovery				freezing medium for 2x 96-well plates	ESC medium without LIF				
ACTION			***CRISPR/Cas9 Transfection***		***Splitting*** each CRISPR onto 2x 6 cm DR4 dishes, ***start selection***	***puromycin selection and recovery***				***Picking*** depending on colony size			***Split&Freeze*** depending on the colony size			***gDNA extraction***	***Genotyping***	***Expanding*** ESCs colonies with desired modification

Figure 12 Timeline for CRISPR/Cas9 experiments
Exemplary timeline for a CRISPR experiment using the px459 vector (Addgene). Cell are transfected and subsequently split for selection with 2 µg/ml puromycin (Sigma-Aldrich, #P8833). After recovering from the selection, single colonies are picked and grown in 96 well formats. 2/3 of the grown colonies will be frozen for later expansion and 1/3 is kept in culture for gDNA extraction and genotyping. When using px458 vector selection on day 5 will occur according to GFP signal by FACs rather than puromycin selection.

Transfection

2 hbefore transfection penicillin/streptomycin was removed from the medium. In one tube, sgRNA containing vector(s) were either mixed solely or combined with a KI-construct containing vector in a total volume of 125 µl OptiMEM (Gibco, #51985-026). In another tube 100 µl OptiMEM were mixed with 25 µl of FuGENE HD transfection agent (Promega, #E2311). Both tubes were combined, vortexed vigorously and incubated for 15 min at room temperature. Afterwards the transfection solution was dropwise added onto the ESCs. To generate genomic SVs (duplications, inversions, deletions) two sgRNAs were used with a total plasmid amount of 4 µg each. To integrate a construct at a distinct genomic site 8 µg of an sgRNA containing vector and 4 µg of the KI-construct containing vector was used. In this way generated cell lines with coordinates of sgRNA cutting site and corresponding KI-constructs are listed in table 4.

Picking

With the help of binocular microscopes single ESC-colonies (15-30 cells) were picked in PBS. Using a pipette tip, single colonies were transferred into U-bottom 96-well plates containing 0.2% Trypsin-EDTA (Gibco, #25300-054) and incubated for 10 min at 37°C to dissociate the colony into single cells. Trypsin reaction was stopped with ESC medium and single cell solutions were seeded onto CD1 monolayers in flat bottom 96well-plates and cultured for 3 days.

Split and freeze

96-well-plates were split into triplicates. Cells of each well were washed with PBS, trypsinized and resuspended in 150 µl of which 50 µl each were transferred into three 96-well plates. Two 96-well plates were frozen with bicarbonate-free DMEM (Gibco, #52100) containing 10% DMSO in U-bottom 96-well plates. One plate was left in culture and after ESCs robustly grow genomic DNA for genotyping was harvested (see 2.2).

Genotyping by PCR and qRT-PCR

Standard PCR procedures with Taq-polymerase (internal production by Asita C. Stiege) was used to screen picked mESC clones regarding potential SVs and KIs. PCR products were analyzed on 1-2 % TAE-agarose gels.

Standard 25 µl PCR reaction: 19.2 µl H_2O, 2.5 µl 10x Taq buffer (750 mM TRIS/HCl pH 8.0; 200 mM (NH4)2SO4; 0.1% Tween20; 15 mM MgCl2), 0.1 µl dNTPs (12.5 mM), 0.1 µl forward primer (100 µM), 0.1 µl reverse primer (100 µM), 2.5 µl template (around 50 ng genomic DNA), 0.5 µl Taq enzyme. Thermocycler: 5 min at 95°C, 25-30 cycles (30 sec at 95°C, 30 sec at 60°C, 45 sec at 72°C); 5 min at 72°C. PCR-positive candidate ESC clones were analyzed regarding their copy number for the desired CRISPR/Cas9 modification by qPCR.

Standard 12 µl qPCR reaction: 6 µl SYBR Green Master Mix (Applied Biosystems), 2 µl PrimerMix (1.5 µM each) and 4 µl gDNA (50 ng total). qPCRs were set up in 384well-plate always in triplicates. The efficiency of primers was tested by standard dilution series of control genomic DNA. qPCRs were run on an ABIPrism Quant7 thermocycler. Relative presence of a genomic region has been determined with the 2-ΔΔCt method. The ratios between tested region and a control region were compared between wildtype and mutants.

Genotyping by western blot

Western Blot was used to genotype *Dnmt3b*$^{-/-}$ mutants and confirm the loss of proteins. Around 2 million feeder-depleted ESCs were used for Western Blot analysis. After centrifugation (300 g / 5 min / 4°C) cells were lysed using lysis buffer (25 mM Hepes pH7.6, 5 mM MgCl2, 1 mM DTT, 25 mM KCl, 0.05 mM EDTA, 10% Glycerol, 0.1% IGEPAL, 1X (Sigma, #4693159001)). Afterwards nuclei were pelleted by centrifugation (3000 g / 5 min / RT), the supernatant was discarded and the pellet resuspended in 150 µl RIPA Buffer. Nuclei in RIPA buffer were vortexed for 20 min at 4°C. Following the mixture was centrifuged 12.000 rpm for 15 min and the supernatant was collected for blotting. Antibodies in the Western blots were used in following concentrations: 1:2000 for rabbit anti-Dnmt3a (Abcam, #ab188470), 1:1000 for rabbit anti-Dnmt3b (Cell Signaling, #cs48488), 1:100 for rabbit antiACTIN (Sigma Aldrich, #A2066) as loading control. Finally, imaging was performed using HRP chemiluminescence.

8.3 Generation of enhancer-reporter lines

Mouse *Fat1-enh*. sequence was PCR-amplified (see tabel 1) and together with a phosphoglycerate kinase promoter integrated at the *Col1a1* locus using the flippase (FLP)-flippase recognition target (FRT) system in C2 ESCs (Beard et al. 2006). Specifically, $1x10^4$ C2 cells/ cm^2 were cultured on a CD1 feeder-monolayer. Cells were transfected with

3 µg FLP-encoding vector and 9 µg of targeting construct. For these vectors were mixed with 1 µl Lipofectamine LTX Plus reagent (Thermo Fisher Scientific, #A12621), 20 µl Lipofectamine LTX in Optimem in a total volume of 250 µl for around $8x10^5$ cells cultured on a 6 well. After one day, C2 cells were spitted and transferred onto hygromycin-resistant DR4 feeder layers. Subsequently hygromycin B (final concentration 150 µg/ml in ESC medium) treatment was started and continued for at least five and up to ten days. Hygromycin B – resistant colonies were picked, cultured, frozen and their DNA genotyped for the desired mutation as descripted for CRISPR/Cas9 editing approach (see 4.2). For coordinates of integration site, primers for enhancer amplification and later genotyping see tables 4 - 6.

8.4 Generation of LacZ-sensor lines

The insertion of LacZ-sensors used asymmetric homology arms matching the flanking sequences of the insertion site. The in combination used sgRNA bisect and so inactivate the sgRNA recognition site of the provided KI construct. Flanking sequences/homology arms were cloned into a vector harboring the inserted lacZ gene under the control of a *β-globin* minimal promoter and a polyA sequence at the gene end for transcription termination. Homology arms were typically synthesized with Genewiz while inserts mostly have been PCR amplified and cloned into the vector by restriction digests. *β-globin* promoter was substituted through restriction cloning with the PCR-amplified or synthesized *Zfp42, Triml1/2,* or *Fat1* promoters (see table 1). All promoters incorporated at least 50 bp downstream and 250 bp upstream of the major TSS as defined by public available FANTOM5 CAGE data (https://fantom.gsc.riken.jp/5/data/). List of used sgRNAs and coordinates of homology arms are provided in table 4.

9. Chromatin modifications

9.1 ChIP-seq

Cells were harvested and fixed with 1% formaldehyde in 10% FCS/PBS for 10 min rotating at RT. Crosslinking was quenched by adding glycine to a final concentration of 125 mM. To isolate nuclei the fixed cells were washed twice with cold PBS and lysed for 10 min on ice using freshly prepared lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA with protease inhibitor (cOmplete, EDTA-free, Sigma, #46931590). Nuclei were centrifuged for (480 g / 5 min / 4°C), washed with PBS and either snap frozen in liquid N_2 or further processed. Chromatin for ChIP-seq was sheared using a Bioruptor until DNA fragment sizes of 200–500 base pairs were reached. Afterwards, samples were processed

with the iDeal ChIP-seq kit for histones/transcription factor (Diagenode, #C01010051, #C01010170,) according to the manufacturer's instructions. For each Histone ChIP 5 µg chromatin was used in combination with each 1 µg of antibodies against H3K4me1 (Diagenode, #C15410037) H3K4me3 (Merck Millipore, #07-473), H3K27ac (Diagenode, #C15410174), H3K27me3 (Merck Millipore, #07-449). For each CTCF ChIP 25 µg chromatin was used in combination with 1.3 µg antibodies against CTCF (Diagenode, #C15410210). Libraries were prepared for sequencing using the KAPA HyperPrep kit (KAPA Biosystems, #07962363001) according the manufactures instructions with following specifications for library construction: ChIP-seq DNA input was set to 10 ng, Adaptor:insert molar ratio was set to 200:1 (1.5 µM adaptors : 5 ng fragmented DNA), Amplification cycles were set to 15. Finally, samples were subjected to a double-sided AMPure XP beads purification with ratios of 0.9 and 0.6. End concentration of the samples were measured using Qubit and prior to sequencing library purity and quality was asset running a High Sensitivity DNA Bioanalyzer ChIP (Agilent, #5067-4627). Libraries were sequenced as 100 bp paired-end reads and raw fastq data processed as following: mapping of reads to the reference genome was performed with STAR (Dobin et al. 2013), filtering, sorting and duplicate removal was performed with SAMtools (Li et al. 2009). Lastly, deepTools was used for the generation of coverage tracks (Ramirez et al. 2016).

9.2 ChIPmentation

Limb cells from embryonic chicken were subjected to ChIPmentation and libraries were prepared as previously described (Schmidl et al. 2015). In PBS dissected embryonic limb cells were dissociated using trypsin. This reaction was stopped with 10% FCS/PBS and the suspension was filtered through a 70 µm strainer (SmartStraine, Miltenyi Biotec, #130098462). Afterwards cells were fixation with 1% MeOH-free formaldehyde/PBS for 10 min at 4 °C followed by quenching the fixation using glycine and pellet collection by centrifugation (300 g / 5 min / 4°C). Subsequently cells were lysed for 10 min at 4°C in lysis buffer (10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA, 0.1% Sodium deoxycholate, 0.5 % N-lauroylsarcosine). Next, chromatin was sheared with a Covaris E220 to reach between 200 and 700 bp fragments. Chromatin was incubated together with the histone antibodies bound for immunoprecipitation with Dynabeads™ Protein G overnight at 4°C. For specific antibodies used see traditional ChIP-seq (see 9.1). At the next day, Tn5-mediated *tagmentation* of pull-downed chromatin was incubated for 5 min at 37°C followed by chromatin de-crosslinking with proteinase K (Roche Diagnostics, #1000144) at 65°C overnight. DNA was then purified using the MinElute Reaction Clean up kit (Qiagen, #28204). In the library amplification Nextera indexing primers (single-

indexed) were added. The number of PCR cycles for each library was estimated using Ct values from qPCR (where number of cycles = rounded up Ct value +1). Fragments purified and selected according to size using AMPure XP beads (left-sided or double-sided selection). Prior to sequencing he concentration of DNA concentration was determined by Qubit HS and fragment sizes checked on a TapeStation D5000HS (Agilent, #G2964AA). All libraries were sequenced as 100 bp paired end reads on a Novaseq2. ChIPmentation was done by Mai H.Q. Phan.

10. ATAC-seq

Performance of ATAC-seq experience was according to the protocols from Buenrostro et al. with minor modifications (Buenrostro et al. 2015; Buenrostro et al. 2013). 1-2x10^5 isolated limb cells were washed in PBS and lysed in freshly prepared lysis buffer that does not contain ionic detergents (10 mM TrisCl pH7.4, 10 mM NaCl, 3 mM MgCl2, 0.1 % (v/v) NP-40). Lysis was incubated for 2 min at 4°C, followed by centrifugation (500 g / 10 min / 4°C) to plellet the nuclei. The supernatant was discarded and the nuclei containing pellets subjected to transposition reaction for 30 min at 37° C and 300 rpm shaking. Transposition miture: 2.5 µl Tn5 Transposase (Illumina, #15027865) in 50 µl 1XTD buffer (Illumina, #15027866). Afterwards DNA was isolated using the MinElute Reaction Clean up kit (Qiagen, #28204) eluting DNA in 11 µl of 10 mM Tris. Sample concentration was determined by Qubit and the sample further prepared for sequencing. Transposed fragments were amplified and marked by barcoded adapters via PCR as following: 10 µl transposed DNA, 2.5 µl of Custom Nextera PCR Primer 1 and 2 (each 25 µM) in total 50 µl 1X NEBNext High-Fidelity PCR Master Mix (NEB, #). Thermocycler: 5min at 72°C, 30 sec at 98°C; 5 cycles (10 sec at 98°C, 30 sec at 63 °C, 1 min at 72°C). To reduce GC and size bias initially PCRs are performed with only 5 cycles. From this, an aliquot of the PCR reaction is subjected to qPCR to identify the number of cycles required to not saturate the PCR. Standard qPCR for ATAC-seq reaction: 5 µl of previously PCR amplified DNA, 4.41 µl Nuclease Free H_2O, 0.25µl of two different Nextera qPCR primers (each 25 µM), 0.09 µl 100x SYBR Green I (Invitrogen, #S-7563), 5 µl NEBNext High-Fidelity 2X Master Mix (New England Labs, #M0541). Thermocycler: 30 sec at 98°C; 20 cycles (10 sec at 98°C, 30 sec at 63 °C, 1 min at 72°C). From this qPCR was determined if additional PCR cycles amplifying the library are needed. Finally, samples were subjected to a double-sided AMPure XP beads purification and eluted in 20 µl of 10 mM Tris. Concentration was measured using Qubit and prior to sequencing library purity and quality was asset running a High Sensitivity DNA Bioanalyzer ChIP (Agilent, #5067-4627). Bioanalyzer results should display periodic peaks according to the DNA length hat wraps around nucleosomes

(e.g. mononucleosomes 150-180 bp, dinucleosomes 300-500 bp). Library were sequenced aiming for at least 50 million reads in oder to assess open versus closed chromatin. Here, 75 bp paired-end sequencing was performed with at least 100 million reads per sample. Analysis of ATAC-seq data was performed as following: Fastq files were subjected to adapter trimming using cutadapt (Martin 2011), mapping was done with Bowtie2 (Langmead and Salzberg 2012), filtering, sorting and duplicate removal was performed with SAMtools (Li et al. 2009). Lastly, deepTools was used for the generation of coverage tracks (Ramirez et al. 2016). For visualization UCSC genome browser was used.

11. DNA methylation

11.1 Bisulfite-cloning sequencing

Bisulfite conversion was performed with the EpiTect Bisulfite Kit (Qiagen, #59104) on 1µg of genomic DNA. Bisulfite-treated DNA was PCR amplified in a nested PCR approach at the *Zfp42* promoter using following primers listed in table 10. Subsequently the amplified DNA was cloned into a pbluescript vector (Stratagene) and sanger sequenced. Per condition 2 replicates with each 10-20 different clones were sequenced and at least 12 CpG were analyzed with BiQ Analyzer software (Bock et al. 2005). Bisulfite-cloning was done by Juliane Glaser.

11.2 WGBS

Genomic DNA from ESCs and E11.5 limb buds was extracted with the PureLink Genomic DNA Mini Kit (Thermo Fisher, #K182002) according to the manufacturer's instructions. Using a Covaris E220 gDNA was sheared in micro TUBE AFA Fiber Pre-Slit Snap-Cap tubes (Covaris, #520077). Afterwards, gDNA was purified with the DNA Clean & Concentrator-5 kit (Zymo, #D4013) following the manufacturer's guidelines. Subsequent bisulfite conversion of the DNA was performed with the EZ DNA Methylation-Gold Kit (Zymo, #D5005) according to the manufacture's protocol. Next, samples were further processed with the Accel-NGS Methyl-seq DNA library kit (Zymo, #DL-ILMMS-12) following the manufacturer's instructions. Finally, WGBS libraries were purified using AMPure XP beads (Agencourt), library quality and absence of adapter primers was checked with the TapeStation D5000HS (Agilent, #G2964AA) prior to loading samples onto a NovaSeq6000 and sequences 150 bp paired-end reads. Fastq raw sequencing reads were subjected to adapter trimming using cutadapt (version 2.4; parameters: --quality-cutoff 20 --overlap 5 --minimum-length 25; Illumina TruSeq adapter clipped from

both reads), followed by trimming of 10 nucleotides at the 5' end of the first read and 15 nucleotides at the 5' end of the second read. This was combined with removing 5 nucleotides at the 3' end of both reads (Kechin et al. 2017). Afterwards were mapped to the mouse mm10 reference genome using BSMAP (version 2.90; parameters: -v 0.1 -s 16 -q 20 -w 100 -S 1 -u -R) (Xi and Li 2009). A command from GATK was used to remove duplicates (version 4.1.4.1; --VALIDATION_STRINGENCY=LENIENT --REMOVE_DUPLICATES=true) (McKenna et al. 2010). The analyses were restricted to only CpGs covered by at least 10 reads and at most 150 reads on autosomes. Rates of methylation were determined the MOABS package and mcall (version 1.3.2; default parameters) (Sun et al. 2014). Differentially methylated region (DMR) were called using metilene (version 0.2-8; parameters: -m 10 -d 0.2 -c 1 -f 1) (Juhling et al. 2016) and filtered by a Q-value < 0.05. Bedtools 'intersect' determined if an DMRs was assigned to promoter (Quinlan and Hall 2010). WGBS was performed by Alexandra Mattei and Sara Hetzel.

12. Simulating chromatin structure with a modified SBS model

mSBS-polymer modelling was performed by **Andrea Chiariello**, **Simona Bianco** and **Andrea Esposito**. The following method part is taken from (Ringel et al. 2022).

> ***SBS-polymer modelling with NE-attachment***
>
> *We simulated the 3D structure of the Fat1/Zfp42 locus in ESC and E11.5 limb buds using a Strings and Binders Switch (SBS) polymer model that incorporates NE-attachment as described below (Barbieri et al. 2012) (Chiariello et al. 2016; Nicodemi and Prisco 2009).*
>
> ***Polymer model***
>
> *Briefly, the SBS polymer model simulates a chromatin filament as a string with N beads, possessing potential binding sites for specific interacting molecules (binders). The binder concentration c and bead-binder interaction energies E_{int} control the system's state through a coil-globule transition occurring when they are above a threshold (Barbieri et al. 2012) (Chiariello et al. 2016). The type and location of binding sites specific for different regions of the Zfp42/Fat1 locus were inferred from ESC or E11.5 limb cHi-C data using PRISMR (mm10 chr8: 40300000 - 46200000; 20 Kb resolution) (Bianco et al. 2018). This machine-learning based algorithm returns the minimal arrangement of binding sites to fit the input. As output, the best polymer modelling the Fat1/Zfp42 locus was generated with 13 distinct types of binding sites in each condition. From these polymers, we obtain a set of 3D structures representing chromatin conformations in ESC and E11.5 limb through standard Molecular Dynamics simulations (see below).*

Details of Molecular Dynamics simulations

In order to build an ensemble of 3D structures representing the Fat1/Zfp42 locus in E11.5 limb and ESC cell lines, we perform extensive Molecular Dynamics (MD) simulations (Chiariello et al. 2016). For simplicity, bead and binders have the same diameter $\sigma = 1$ *and mass* $m = 1$ *in dimensionless units. A standard truncated Lennard-Jones (LJ) potential models the hard-core repulsion between the objects. By contrast, interaction between beads and binders is modelled with an attractive LJ potential with distance cutoff ranging from* $R_{int} = 1.3\sigma$ *to* $R_{int} = 1.5\sigma$ *and an interaction intensity, given by the minimum of the LJ potential, within the range of* $E_{int} = 3.1 - 8.2K_BT$*. An additional non-specific, weaker interaction (in the* $E_{int} = 2 - 3K_BT$ *range) is set among binders and the polymer. Consecutive beads of the polymer are linked by FENE bonds (Kremer and Grest 1990) with standard parameters (length* $R_0 = 1.6\sigma$ *and spring constant* $K_{FENE} = 30K_BT/\sigma^2$*). Beads and binders move through Brownian dynamics according to the standard Langevin equation (Allen and Tildesley 2017) with temperature* $T = 1$*, a friction coefficient* $\zeta = 0.5$ *and an integration time step* $\Delta t = 0.012$ *(dimensionless units). The polymer is first initialized as a Self-Avoiding-Walk and the binders are randomly located in the simulation box, then the system is equilibrated up to approximately* 10^8 *timesteps. From each model, we perform up to* 10^2 *independent simulations in which polymer configurations are sampled every* $5*10^5$ *timestep once equilibrium is reached. Simulations are performed with the LAMMPS package (Plimpton 1995).*

Modelling the nuclear envelope

To model the NE, we introduce a spherical wall of radius R *within the simulation box. Polymer beads can attractively interact with NE though a short range, truncated LJ potential with affinity* E_{NE} *ranging from* $0.0K_BT$ *to* $10K_BT$ *and cutoff distance* $r_{cutoff} = 2.5\sigma$*. Among the NE-bead interaction energies tested, the structures obtained immediately after the NE-polymer adsorption (around* $1.2K_BT$*) generated structural measurements that most closely matched those observed by FISH. Alternatively, beads interact with NE only through a purely repulsive LJ potential. The NE sphere radius is set to* $R = 40\sigma$*. In order to define the interaction state (repulsive or attractive) of each polymer bead with NE, we employ DamID data for each wild or mutant ESC/limb sample. Briefly, we compute the average DamID signal in each 20kb window and evaluate its sign. Polymer beads associated with an average positive DamID signal are classified as attractively interacting with NE. Conversely, beads associated with a negative signal experience only a repulsive interaction. In this way, regions enriched with DamID tend to attach to the NE in the model. In our simulations, the NE is introduced after the SBS (polymer+binders) system is equilibrated, as described in the previous section. Then, in order to ensure the complete interaction of the polymer with the NE, the system is equilibrated up to other* $7*10^7$ *timesteps.*

Quantification of measurements

Pairwise distance distributions are extracted from the population of 3D polymer structures as previously described (Chiariello et al. 2016; Conte et al. 2020). For each pair of objects, we first compute the center of mass of the polymer beads belonging to that object, then we evaluate the distance between the centers of mass. This distance is then averaged over the last 20 frames of each simulation. In order to map dimensionless length scale into physical units we compare pairwise distances measured by FISH. In total, we compare six different probe pairs (D1-D2, Fl1-Fl2, Zfp42R-D1, Zfp42R-D2, Zfp42R-Fl1, Zfp42R-Fl2) both in E11.5 limb and ESCs, for each pair we equalize the model and experimental median and then average over the different probe pairs. The resulting length scale mapping factor is $\sigma = 44nm$*. Distances from NE shown in Figure S4E and F are estimated by computing:* $d_{NE} = R - |\vec{r}_{CM} - \vec{r}_{NE}|$*, where* R *is the model NE radius,* $\vec{r}_{CM}$ *is the position of the center of mass of the object and* $\vec{r}_{NE}$ *is the position of the NE center. Physical distances are then obtained using the mapping factor* σ *previously calculated from the comparison with pairwise FISH distances. Pairwise overlaps between two objects shown in Figure S4I are obtained by using the following expression:* $overlap_{12} = A_{12}/(A_1 + A_2 - A_{12})$*, where* A_1 *and* A_2 *are the surfaces of 2D projections associated to object 1 and object 2 respectively and* A_{12} *is their common area. For simplicity, 2D projections are approximated as circles whose radii* R_1 *and* R_2 *are estimated as gyration radii from the projected coordinates, so* $A_1 = \pi R_1^2$ *and* $A_2 = \pi R_2^2$*. In this way, overlapping areas can be easily estimated using standard geometric relations. Indeed, given the distance* d *between the centers of the projected objects and supposing, without loss generality,* $R_2 > R_1$*, we have a partial overlap if* $R_2 - R_1 < d < R_1 + R_2$*. In this case:* $A_{12} = R_2^2\alpha_1 - d_1\sqrt{(R_2^2 - d_1^2)} + R_1^2\alpha_2 - d_2\sqrt{(R_1^2 - d_2^2)}$*, where* $d_1 = \frac{R_2^2 - R_1^2 + d^2}{2d}$ *and* $\alpha_1 = arccos\frac{d_1}{R_2}$ *(analogous relations hold for* d_2 *and* α_2*). If* $d \geq R_1 + R_2$*, we impose* $A_{12} = 0$*, i.e. objects are well separated in space; finally, if* $d \leq R_2 - R_1$*, we set* $A_{12} = \pi R_1^2$*, i.e. object 1 is completely contained within object 2. Three body overlaps shown in Figures S4E and S4G involving Zfp42R or Fat1 with D1+D2, are defined as:* $overlap_{123} = (A_{12} + A_{13})/(A_1 + A_2 + A_3 - A_{12} - A_{13} - A_{23})$*, where object 1 can be Zfp42R or Fat1. As for 3D distances, overlap values are averaged over the last 20 frames of each simulation. Analogously, a geometric mapping factor of 1.2 is found when comparing with pairwise experimental medians. Sphericity is defined using the standard formula:* $sphericity = \frac{\pi^{1/3}(6V)^{2/3}}{A}$*, where* A *and* V *are area and volume of the object respectively. Area and volume are estimated from the coordinates of the polymer beads belonging to the region under consideration by means of a 3D convex hull approximation, computed with the Python package scipy.spatial.*

Contact maps are computed as previously described (Conte et al. 2020; Chiariello et al. 2016). We first measure the distance r_{ij} between any two beads i and j. If the distance is lower than threshold (7.5σ in Figure S4B, C), the beads are in contact. For each considered condition (without NE and with NE at different interaction energies), aggregated matrices are obtained over the different independent simulations. Visual and quantitative comparisons reveal a general good agreement between model and cHi-C data in both cell lines (Pearson $r = 0.90$ and distance-corrected (Bianco et al. 2018). Pearson $r' = 0.72$ in HL, $r = 0.91$ and $r' = 0.64$ in ESC, genomic distances $> 100kb$). Subtraction matrices D are defined as the simple bin-wise difference $D_{ij} = x_{ij}^{NE} - x_{ij}$, where x_{ij}^{NE} and x_{ij} are the entries of the contact maps with and without NE respectively.

Polymer graphics

Polymer 3D snapshots shown in Figures are representative single molecule structures taken from real MD simulations. Regions corresponding to Fl1, D1, Zfp42R, D2, Fat1, Fl2 are differently colored. A slice of the simulated NE is rendered as a thick spherical wall colored as in FISH imaging. To clarify the relationship between the polymer and NE, each image is presented from the same point-of-view through a geometrically calibrated 3D rotation matrix. For visual purposes, polymers are shown in a coarse-grained version of a smooth third-order polynomial spline passing through bead coordinates.

mSBS-polymer modelling was performed by **Andrea Chiariello**, **Simona Bianco** and **Andrea Esposito**. The following method part is taken from (Ringel et al. 2021).

Models of mutants: *Polymer models of deletions in HL are simulated as described in (Bianco et al. 2018). Basically, we implement in-silico mutations on the polymer model trained on WT data by deleting the portion corresponding to the deleted chromatin regions in experiments. Specifically, polymer model for ΔD1 has $\mathcal{N}$=2130 beads (i.e. without the region corresponding to D1); analogously, polymer model for ΔD2 has $\mathcal{N}$=2190 beads (i.e. without the region corresponding to D2); finally, polymer model for ΔD1+2 has $\mathcal{N}$=1370 beads and it is much shorter as it carries the deletion of both D1 and D2. For each mutation, a population of 3D polymer structures is then obtained through independent MD simulation performed as described above. DamID data specific for each mutation is integrated in the model to simulate NE. Distances and overlap distributions are generated using mapping coefficients estimated from the WT models.*

13. Studying chromatin structure by FISH and 3D-SIM

13.1 Generation of FISH probes

Oligopaint library design: Single-stranded Oligopaints were used to visualize the chromatin region of interest. Oligopaint libraries were designed as described at (https://oligopaints.hms.harvard.edu) and in (Beliveau et al. 2015). The mm10 coordinates, size, number and density of oligonucleotides are listed in table 11. Oligopaint probes consist of 32-41mer genomic sequence with flanking non-genomic regions containing primers for amplification and labelling with fluorophores (Beliveau et al. 2015). Oligonucleotide primer sequences (5′→3′) used are listed in table 12.

Oligopaint probe generation: Oligopaint probes were produced by emulsion PCR using primers listed in table amplification pooled libraries followed by a 'two-step PCR' combined with the lambda exonuclease method described by Beliveau et al. (Beliveau et al. 2015). First PCR mixture was prepared as following: 85 µl H2O, 10 µl Kapa Taq Buffer A, 2 µl KAPA dNTP Mix, 0.5 µl forward primer, 0.5 µl reverse primer (see table 11), 1µl product of ePCR (1ng/µl), 1 µl KAPA Taq DNA Polymerase (5U/µl). The following thermo cycles were used: 1.: 5 min at 95 °C; 2: three cycles: 30 sec at 95 °C, 45 sec at 60 °C, 30 sec at 72 °C; 3.: twenty cycles: 30 sec at 95 °C, 1min at 68 °C, 30 sec at 72 °C; 4.: 5 min at 72 °C. PCR products were purified using the PCR purification kit from Qiagen. Second PCR mixture was prepared as following: 864 µl H2O, 100 µl Kapa Taq Buffer A, 20 µl KAPA dNTP Mix, 2.5 µl forward primer, 2.5 µl reverse primer (coupled with A488 or ATTO647 if Sec1, A565 if Sec6), 1µl product of PCR 1 (1ng/µl), 10 µl KAPA Taq DNA Polymerase (5U/µl). The following thermo cycles were used: 1.: 5 min at 95 °C; 2: twenty-five cycles: 30 sec at 95 °C, 30 sec at 60 °C, 15 sec at 72 °C; 3.: 5 min at 72 °C. Afterwards, samples were purified again with the PCR purification kit from Qiagen and subsequently exonuclease reaction was performed by mixing in 225 µl eluted PCR2 product 2 with 25 µl of exonuclease buffer and 0,5 µl of exonuclease. Reaction was incubated for 30 min at 37 °C and then 10 min at 75 °C. Finally, probes were purified with ethanol precipitation and used for subsequent FISH procedure. The two-step PCR enables signal amplification by adding a secondary oligonucleotide with two additional fluorophores. Thus, each probe contains in total three fluorophores.

BAC probe preparation: The BAC for the *Fat1* region (CHORI/BACPAC, # RP23-451E23) was labeled with AlexaFluor 555 using the FISH Tag DNA Kit according to the manufacturer's instructions (Thermo Fisher Scientific, #MP32947).

13.2 FISH and immunostaining for 3D SIM

FISH procedure: Generally, sample preparation was performed as described in (Szabo et al. 2020). On day one, fixed cells on coverslip (see 5.3) were treated with PBS/0.5% Triton X-100 for 10 min at RT. After, three washes in PBS, samples were incubated for 10 min in 0.1 M HCl/H_2O, washed in PBS. Next, coverslips were twice rinsed in 2X SSCT (2× SSC/0.1% Tween20) and then first incubated in 50% formamide/2× SSCT for 20 min at RT and then, another 20 min at 60 °C. In a next step, hybridization solution containing the Oligopaint probes (primary and secondary probes at 1–3 µM final concentration + 0.8 µl RNase (10 mg/ml) + 12.5 µl FHB (50% formamide, 10% dextran sulfate, 2×SSC and salmon sperm DNA with a final concentration of 0.5 mg/ml) were added onto the side. If needed, 25 ng of *Fat1* BAC probe together with excess of mouse Cot-1 DNA was also added and co-hybridized. Coverslips were sealed with rubber cement, then DNA and probes were co-denatured for 3 min at 80 °C followed by overnight incubation in a humid box at 42 °C. On day 2, the rubber cement was removed and colverslides were washed in 2XSSCT for 15 min at 60 °C, then twice for 10 min at RT. After rinsing the samples several times in PBS, immunostaining was performed. Free binding sides were blocked by incubating the sample for 1 h in PBT/2%BSA. Then lamina was detected with antibodies. Primary antibody, ant-lamin B1 (1:1000), was incubated overnight at 4 °C between coverslips followed by secondary antibody, anti-rabbit-IgG-Atto 647 (1:100) incubated for 1 h at RT. Finally, samples were washed in PBT and stained with DAPI (final concentration 1 µg/ml) for 3 min at RT. After, several washing steps in PBS, coverslides were mounted with Vectashield and then sealed with color-free nail polish.

Image acquisition: The DeltaVision OMX V4 microscope combined with an ×100/1.4 numerical aperture Plan Super Apochromat oil immersion objective (Olympus) and electron-multiplying charge-coupled device (Evolve 512B; Photometrics) camera for a pixel size of 80 nm. Diode lasers at the wavelength of 405, 488, 561 and 647 nm were used with corresponding emission filters. Images were acquired every 125 nm in z (Z-stacks) each from 3 angles and in 5 phases per image plane. SoftWorx v.6.5 from GE Healthcare Systems was used to reconstruct the raw images with the channel-specific optical transfer functions (pixel size of reconstructed images = 40 nm). The SIMcheck plugin of ImageJ v.1.52i tool was used to assess the quality of reconstructed images (Ball et al. 2015). Alignment of different channels were frequently checked and calibrated using TetraSpeck beads (200 nm) (Thermo Fisher Scientific, # T7280).

FISH image analysis was done by **Quentin Szabo**. This method part is taken from (Ringel et al. 2022).

Oligopaint FISH image analyses

Image analysis was performed using Fiji and MATLAB (R2018-2019 and image processing toolbox). For overlap intermingling fraction and combined sphericity measurements, z-stacks of regions of interest (ROIs) of 3×3 µm2 surrounding FISH signals were extracted and smoothed using a 3D Gaussian filter (sigma = 0.5 pixel). FISH channels were then segmented in 3D using automatic Otsu's method. Only ROIs containing 1 FISH segmented object per channel (or at least 1 object for the combined D1+D2 FISH) larger than 0.04 µm3 were kept for further analyses. Object intermingling fraction of Zfp42R or Fat1 with D1+D2 (Figures 3D and S4) was obtained by dividing the overlapping volume between Zfp42R or Fat1 and D1+D2 by the volume of Zfp42R or Fat1. Overlap (Figure 3D and S4) correspond to the Jaccard Index between the two segmented FISH objects. For combined sphericity calculation, FISH segmented objects from the two channels were merged into one, and only ROIs containing 1 merged object were considered for the analysis. Combined sphericity was defined as $= \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$ *where V is the volume of the segmented object and A its surface area. For distance to lamin analysis, z-stacks of ROIs surrounding individual nuclei were extracted and smoothed using a 3D Gaussian filter (sigma = 0.5 pixel). FISH channels were segmented using a threshold value corresponding to 20% of the maximum pixel intensity. For a given FISH channel, only nuclei containing 2 segmented FISH objects larger than 0.04 µm3 were kept for further analysis. For each FISH object, an ROI surrounding its maximum and minimum z-coordinates was extracted and the lamin channel was segmented using Otsu's method. Lamin segmented objects smaller than 0.02 µm3 were discarded and Lamin segmented channel was processed using the MATLAB imfill function. 3D Euclidean distance transform of the segmented Lamin channel was calculated using the MATLAB bwdistsc function and distance to the centroid of the FISH segmented object was extracted.*

14. Studying chromatin-NE associations by DamID-seq

14.1 Production of lentiviruses

Lentivirus production was performed using rapidly growing and easy to transduce 293FT cells. 2.8 µg pMD2.G, 4.6 µg psPAX2, and 7.5 µg of pLgw V5-EcoDam or pLgw-EcoDam-V5-Lamin (sources: Steensel and Bird Lab) were mixed with 36 µl lipofectamine 2000 (Invitrogen, # 11668-027) in a total volume of 3 ml Optimem (Invitrogen, # 31985-047) supplemented with 1:500 protamine sulphate (10 µg/ ml) to enhance the transduction. With this around 6 million 293FT cells cultured in 10 cm tissues were transfected. After 16 h, the medium was replaced. Virus-containing supernatants was collected after 48 h and 72

h. First, cellular debris were removed from viral supernatants by centrifugation (330 g / 10 min / RT). Afterwards suspension was filtered with a 0.45 μm^2 low protein-binding PES syringe filter. Finally, viral supernatant was concentrated using Lenti-X concentrator (Takara, #631232) according to the manufacturer's instructions. After centrifugation (max speed / 1 h / 4 °C) pellet was resuspended in Optimem and either frozen at -80 °C or immediately used.

14.2 DamID-seq

DamID was used to assays chromatin-NE interactions and performed as described in Vogel et al. and Robson et al. (Robson and Schirmer 2016). ESCs and cultured E11.5 limb cells were transduced with either DNA adenine methyltransferase (Dam-only) alone or fused to the nuclear lamin B1 (Dam-Lamin B1). Specifically, 1,5-2 $x10^5$ cells were plated per gelatinized 6 well which were allowed to saddle down for 1 h prior to transduction with lentiviruses. After 24 h, medium was exchanged and cell were split onto feeder-monolayers with a 3 times higher growth area (6cm plates). After 2 days, cells were feeder depleted and subsequently ESCs were pelleted by centrifugation (300 g / 5 min / RT). Feeder-free E11.5 limb bud cells were directly harvested by centrifugation (300 g / 5 min / RT). Cell pellets were snap-frozen and stored at -80°C until further processing.

Processing of DamID samples: gDNA was extracted using DNeasy tissue kit following the protocol for “DNeasy Protocol for Cultured Animal Cells”. First, 2.5 µg gDNA was digested with methylation sensitive enzymes *DpnI* that overnight at 37 °C followed by inactivating the enzymes for 20 min at 65°C. Next, DamID adaptor duplex (dsAdR), generated by hybridization of the oligonucleotides AdRt (5’-CTAATACGACTCACATAGGGCAGCGTGGTCGCGGCCGAGGA-3’) and AdRb (5’-TCCTCGGCCG-3’), were ligated onto *DpnI* (NEB, # R0176L) digested fragments for 2 h at 16°C followed by enzyme heat inactivation for 10 min at 65°C. Final ligation mixture contained 2.5 µg DpnI-digested gDNA, 2 µM Adaptor dsAdR and 5U T4 ligase (Roche, # 799009). Next, fragments were further digested with *DpnII* (NEB, # R0543S) for 1 h at 37°C followed by PCR amplification of only the methylated fragments. Final PCR set up contained 0.5 µg *DpnII* digested DNA, 1.25 µM Adr-PCR primer (5’-GGTCGCGGCCGAGGATC-3’), 0.2 mM dNTPs in 1X PCR advantage enzyme mix (Clonetech, # 639105). Thermocycler: 10 min at 68 °C, 1 min at 94 °C, 5 min at 65 °C, 15 min at 68 °C; 3 cycles: (1 min at 94 °C; 1 min at 65 °C, 10 min at 68 °C); 18 cycles: (1 min at 94 °C; 1 min at 65 °C, 2 min at 68 °C). Generated PCR products were purified using the Qiaquick PCR purification kit followed by libraries preparation using the KAPA HyperPrep kit according to the manufacturer´s protocols. Prior to sequencing quality was checked

running a High Sensitivity DNA Bioanalyzer ChIP (Agilent, #5067-4627). DamID-seq samples were sequenced with 75 or 100 bp paired-end reads on a Nova Seq2 (Illumnia). Each experiment was performed at least in duplicates. Sequencing data were mapped to the mm10 mouse reference genome using the alignment tool BWA-MEM (v.0.7.12) (Li et al. 2009). Read counts overlapping with *DpnII* (GATC) sides were normalized as reads per kilobase, divided by fragment length and per million mapped reads. These counts were used to calculate the log2 fold change of Dam–Lamin B1 and corresponding Dam-only samples. HMMt was used to quantify DamID signals using a modified Baum-Welch algorithm with t emissions which allowed LAD calling for 20 kb bins (https://github.com/gui11aume/HMMt).

15. Studying chromatin interactions by cHiC and HiC

15.1 Capture HiC (c-HiC)

Library design: SureSelect design for cHi-C library was done with the SureDesign tool from Agilent covering chr8: 39022300-48000000 (mm10).

Fixation for HiC/c-HiC: Cell pellets (see 5.1) were resuspended in 10% FCS/PBS. Cells were fixed by adding 37% formaldehyde to a final concentration of 2% followed by 10 min incubation rotating at room temperature. Fixation was quenched by adding glycine to a final concentration of 125 mM and subsequently pelleted (480 g /4°C/ 5min). To isolate nuclei cells were resuspending in freshly prepared lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA with protease inhibitor) and placed on ice for 10 min. Nuclei were centrifuged (480 g / 5 min / 4 °C), washed twice with PBS, snap frozen and stored in liquid N2.

Processing of c-HiC samples: Fixed and lysed were thawed on ice and resuspended in 420 µl 1X *DpnII* restriction buffer (NEB) with 15 µl 10%SDS and incubated for 1 hat 37°C, shaking at 1000 rpm. SDS was quenched by adding 150 µl 10% Triton X-100 (Sigma, #93443) and incubated for an addition hshaking at 900 rpm at 37°C. Next, 40 µl *DpnII* enzyme (10 U/µl) was added first for 6 h and then additional 40 µl of *DpnII* prior to overnight incubation shaking at 1000 rpm at 37°C. On the next day, 20 µl *DpnII* enzyme was added to the samples and incubation was continued for an additional 4 h shaking at 1000 rpm at 37 °C followed by enzyme inactivation by incubation for 25 min at 65°C. Next, digested chromatin was transferred into a 15 ml Falcon tube and diluted with 5.1 ml H_2O, 700 µl 10× ligation buffer, 5 µl T4 DNA ligase (30 U/µl) and incubated rotating for 4 h at 16°C and 30 min at room temperature. Chimeric chromatin products were de-cross-linked overnight with 30 µl and 5 µl proteinase K (10 mg/ml, produced by Asita Stiege) at 65 °C. On the

next day, RNA was digested adding 30 µl of RNase (10 mg/ml) and incubated for 45 min at 37°C. Then, DNA was extracted by adding 7 ml volume phenol-chloroform to the samples, vigorously shaking and centrifugation (4000 rpm / 15 min / RT). Next, the water phase was transferred into a new falcon mixed with 100 µg glycogen and subsequently DNA was precipitated with 1 volume 100% ethanol and 0.1 volume 3M NaAc, pH 5.6. The sample was placed at -80°C for 30 min and afterwards centrifuged (max speed /1 h / 4°C). After washing twice with 70% ethanol, the pellet was air tried for 15 min and resuspension in 50 -100 µl with 10 mM Tris-HCl, pH 7.5 and concentration of library samples was measured using Qubit. Then samples were shared using a Covaris sonicator (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 6 cycles of 60 s each; set mode: frequency sweeping; temperature: 4–7°C) and adaptors were ligated onto sheared DNA and amplified following the manufacturer's instructions for Illumina sequencing (Agilent). Afterwards libraries were enriched and indexed for sequencing using the customer SureSelect beads according to the manufacturer's instructions (Agilent).

15.2 HiC

Processing of HiC samples: Fixed and lysed cells (see above) were thawed on ice and resuspended in 500 µl 1X *DpnII* buffer with 50 µl 0.5% SDS/ H_2O and incubated for 10 min at 62°C. Then, 145 µl H_2O and 25 µl 25% TritonX-100 (Sigma, 93443) was added and incubated for 15 min at 37°C to quench the SDS without rotation. Next, chromatin was digested with 10 µl *DpnII* enzyme (100 U) for 2 hfollowed by heat inactivation for 20 min at 65°C. Afterwards, chromatin was centrifuged (5000 g / 5 min / RT) and the pellet resuspended in 240 µl 1X NEB Buffer 2 to fill in the digested DNA ends with biotin-14-dATP (Life Technologies, #19524-016) using DNA Polymerase I (NEB, #M0210) followed by ligation using T4 DNA ligase (NEB, #M0202). The following mixture for the fill in reaction was incubated for 1 hat 37°C: 37.5 µl of 0.4mM biotin-14-dATP, each 1.5 µl of dCTP/dGTP/dTTP (10 mM), 8 µl (5U/µl) DNA Polymerase I, Large (Klenow), Fragment. The following mixture for the ligation reaction was incubated at 18°C overnight: 663 µl H_2O, 120 µl of 10X NEB T4 DNA ligase buffer, 100 µl 10% Triton X-100, 12 µl 10mg/ml BSA (100X BSA), 5 µl T4 DNA Ligase (400 U/µl). At the next day, samples were centrifuged (5000 g / 5 min / RT) and pellet resuspended in 600 µl 10mM TrisHCL, pH 7.5. Next, crosslinking was reversed by adding 25 µl protease K (10 mg/ml) and 60 µl 10% SDS for 30 min ad 55°C and subsequent incubation in 5 M NaCl for 2 h at 68°C. Afterwards DNA was purified by ethanol precipitation and then sheared into 300–600 bp fragments using S-Series 220 Covaris. In the next step biotin pull-down was performed using Dynabeads MyOne Streptavidin T1 beads (Life technologies, #65602). Specifically, after washing

beads in 1X Tween Washing Buffer (TWB: 9 ml H_2O, 15ml 10mM Tris pH 7.5, 6ml 5M NaCl, 30 µl 0.5M EDTA, 15 µl Tween20) beads were resuspended in 300 µl 2X binding buffer, mix with the 300 µl of sheared chromatin and incubate for 15 min rotating at RT to allow streptavidin beads to bind to the biotinylated DNA. Separate beads using a magnet rack and resuspend in 100 µl 1X NEB T4 DNA Ligase buffer (NEB, #B0202). Then separate beads via magnet rack again and discard buffer.

Next, DNA ends were repaired with T4 DNA polymerase (NEB, #M0203) and the Klenow fragment of DNA polymerase I (NEB, #M0210) and subsequently phosphorylated using T4 Polynucleotide Kinase (NEB, #M0201). The following mixture was prepared and incubated for 30 min at RT: 79.2 µl H_2O, 8.8 µl 10X NEB T4 DNA ligase buffer containing 10mM ATP, 2 µl 25mM dNTP mix, 5 µl T4 PNK (10U/µl), 4 µl T4 DNA polymerase (3U/µl), 1 µl DNA polymerase I (5U/µl). Then, beads were washed twiced as following. Beads were separated using the magnet rack, first resuspended in 1X NEB Buffer 2 and then in 600 µl 1X TWB for 2 min rotating at 500 rpm. Next, following mixture was prepared and incubated for 30 min at 37°C: beads, 90 µl 1X NEB buffer2, 5 µl 10mM ATP, 5 µl NEB Klenow exo minus (5U/µl) (NEB, #M0212). Then beads were washed twice again as described above and further prepared for sequencing by adding adaptors to DNA fragments using the NEBNext Ultra DNA Library Prep Kit from Illumina. Beads were resuspended in 50 µl 1X NEB Quick ligation buffer (NEB, #B6058) with 2 µl DNA Quick ligase (NEB, # M2200) and 3 µl universal adapter from the Library Prep Kit from Illumina and incubate at RT for 15 min. Then, additionally 3 µl user enzyme was added and further incubated for 15 min at 37°C. Final amplification occured in 4-6 PCR cycles using following PCR set up: 15 µl adaptor-ligated DNA fragments on beads, 25 µl NEBNext Ultra II Q5 Master Mix, 5 µl Index Primer/i7 Primer, 5 µl Universal PCR Primer/i5 Primer. Afterwards, purification and size selection were performed using AMPure XP beads. Finally, prior to sequencing libraries were checked on a Bioanalizier ChIP and then loaded onto a NovaSeq2 yielding for 100-200 mio paired-end reads per sample

cHi-C and Hi-C analysis was done by **Robert Schöpflin**. The following method part is taken from (Ringel et al. 2022):

> ***cHi-C analysis***
>
> *Raw fastq files had read lengths of 75 bp and 100 bp, respectively. In a preprocessing step, fastq files with 100 bp read length were trimmed to 75 bp to achieve comparable initial read lengths for all samples. Afterwards, fastq files were processed with the HiCUP pipeline v0.8.1 (no size selection, Nofill: 1, Format: Sanger) for mapping, filtering and deduplication steps (Wingett et al. 2015). The pipeline was set up with Bowtie 2.4.2 for mapping short reads to reference genome mm10 (Langmead and*

Salzberg 2012). If replicates were available, they were merged after the processing with the HiCUP pipeline. Binned and KR normalized cHi-C maps (Knight and Ruiz 2012; Rao et al. 2014) were generated using Juicer tools v1.19.02 (Durand et al. 2016). Only read pairs for region chr8:39,030,001-48,000,000 and with MAPQ≥30 were considered for the generation of cHi-C maps.

In addition to the original cHi-C maps, custom reference genomes were derived from mm10 for the ΔD1+2 deletion line. cHi-C and Hi-C maps were displayed as linear-scaled heatmaps in which very high values were truncated to improve the visualization.

Hi-C analysis: *Fastq files were processed with the Juicer pipeline v1.5.6 (Durand et al. 2016) (CPU version) using bwa v0.7.17 (Li and Durbin 2010) for mapping short reads to the reference genomes mm10 (mouse), hg19 (human), galGal6 (chicken), monDom5 (opossum), susScr11.1 (pig), and AmexG_v6.0-DD (axolotl), respectively. Replicates were merged after the mapping, filtering and deduplication steps of the Juicer pipeline. Juicer tools v1.7.5 (Durand et al. 2016) were used to generate binned and KR normalized Hi-C maps from read pairs with MAPQ≥30.*

For compartment analysis, hic-files were converted at 100kb bin size to the cool format using hic2cool (v0.8.2) (https://github.com/4dn-dcic/hic2cool) and balanced using cooler (v0.8.5) (Abdennur and Mirny 2020). Afterwards, compartment analysis was performed using cooltools (v0.3.0) (https://github.com/open2c/cooltools) and using the GC content as reference track.

TADs were identified by applying TopDom v.0.0.228 on 50-kb binned and KR-normalized maps using a window size of 10 (Shin et al. 2016a). Insulation scores were calculated using Cooltools (https://github.com/open2c/cooltools/tree/v0.4.1)"

16. Enhancer prediction and conservation

Enhancer prediction and analysis of enhancer conservation was done by **Tobias Zehnder.** The following method part is taken from (Ringel et al. 2022):

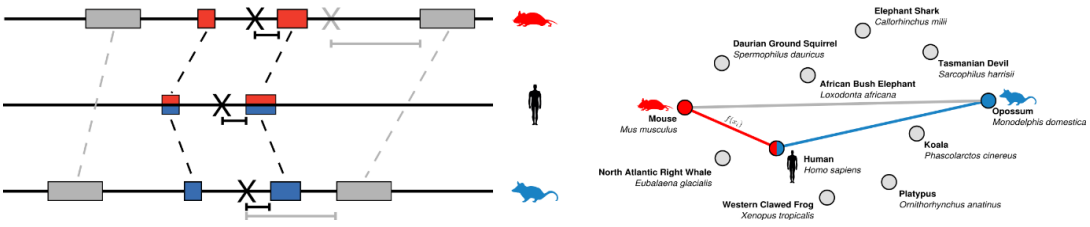
Enhancer prediction

Enhancers were predicted using a series of established tools for ATAC-seq peak prediction and enhancer / promoter prediction. First, Genrich (not published, https://github.com/jsh58/Genrich/) was used to predict ATAC-seq peaks. We filtered for those that overlap a enhancer predicted by CRUP (Ramisch et al. 2019) and do not overlap an annotated TSS (UCSC) or a promoter predicted by eHMM (Zehnder et al. 2019).

Enhancer conservation analysis

ATAC-seq peaks and predicted enhancers were projected between mouse, opossum and chicken using a published stepped pairwise sequence alignment approach across multiple bridging species (Baranasic et al. 2021). For a genomic region with conserved

synteny, any non-alignable coordinate can be approximately projected from one genome to another by interpolating its relative position between two alignable anchor points. The accuracy of such interpolations correlates with the distance to an anchor point. Therefore, projections between species with large evolutionary distances tend to be inaccurate due to a low anchor point density. Including so-called bridging species increases the anchor point density and thus improves projection accuracy. The optimal choice and combination of bridging species may vary from one genomic location to another. This presents a shortest path problem in a graph where every node is a species and the weighted edges between nodes correspond to a scoring function that represents the distances of genomic locations to their anchor points (|x - a|). The scoring function exponentially decreases with increasing distances |x - a|. The shortest path problem is solved using Dijkstra's Shortest Path Algorithm (Dijkstra, 1959). The sets of bridging species used here are described in Table S6. Projected elements from ATAC-seq peaks were then classified into directly (DC), indirectly (IC) or not conserved (NC) according to the following criteria: DC elements overlap a direct sequence alignment between the reference and the target species. IC elements do not overlap a direct alignment, but are projected with a score > 0.99, i.e. either overlapping or in direct vicinity to a multi-species anchor. A score of > 0.99 means that the sum of the distances from the element and its intermediate projections to their respective anchor points is < 150 bp throughout the optimal bridging species path. The remaining peaks are classified as non-conserved (NC).



***Genomic coordinate projection schematic illustration. Left**. An example genomic location X is projected between observed (e.g. mouse) and target species (e.g. opossum) using the direct alignments (grey rectangles) and the alignments via a bridging species (e.g. human, blue and red rectangles). Projections are indicated as a black X in the respective species). Dashed lines connect pairwise sequence alignments. The projected locations of X in observed species are indicated in grey (direct alignments) and black (via bridging species). **Right.** Example graph comprising 13 species (nodes). For any genomic location, the shortest path through*

the species graph yields the combination of species which maximizes projection accuracy.

RESULTS

In the following part, the results of this work are presented in four sections. First, we describe the evolutionary history of the *Fat1* locus and the later incorporation of *Zfp42* together with its unexpected independent expression program (1). In section two, we reveal how this evolutionary change influenced the locus' structure connected to *Fat1´*s and *Zfp42´*s independent activities in ESCs and Limbs (2). Finally, in sections three and four, we reconstruct how these structural alterations cooperate with epigenetic mechanisms to enable independent *Zfp42* and *Fat1* regulation. In particular, section three focuses on the mechanisms in ESCs driving the locus' structural adaption enabling the genes´ simultaneous but independent activity (3). The fourth section focuses on the mechanisms of lamina association, enhancer-promoter specificity and DNA methylation in limbs that do not or do explain *Zfp42´*s inactivity despite *Fat1´*s activity (4).

1. *Zfp42´*s evolutionary emergence in *Fat1´*s ancient regulatory landscape

Considering their importance for delimiting enhancer-promoter interactions, TADs have generally been considered as stable across evolution and during development (Dixon et al. 2012; Harmston et al. 2017; Gómez-Marín et al. 2015; Renschler et al. 2019; Krefting et al. 2018; Vietri Rudan et al. 2015). Moreover, TADs have been thought to be units of gene co-regulation as all genes within a TAD are exposed to the same enhancers (Nora et al. 2012; Shen et al. 2012; Flavahan et al. 2016; Dixon et al. 2016). As such, newly emerged genes would be expected to adapt the regulatory information of the TAD in which they land. In this first section, we investigated the validity of these statements for the *Fat1/Zfp42* TAD. By tracing back when *Zfp42* emerged within *Fat1´s* landscape, we explore how stable the TAD is across evolution and whether *Zfp42* adapted to *Fat1´*s expression. Using HiC/cHiC combined with ATAC-seq, we mapped the locus´ structure and function. With this, we found that throughout evolution *Fat1* is associated with its structurally preserved TAD that drives conserved *Fat1* expression in embryonic limbs across different species. Moreover, this *Fat1* patterning is neither disrupted nor adopted by *Zfp42´*s emergence which instead has an entirely different activity program in mouse development.

1.1 Zfp42 emerged in Fat1´s structurally conserved TAD in placental mammals

The focus of this study is on the *Fat1* locus in which *Zfp42* emerged in placental mammals (Kim et al. 2007; Masui et al. 2008; Sadeqzadeh et al. 2014). To see how structurally and functionally conserved the *Fat1/Zfp42* landscape is, we first mapped its evolutionary

history. Examining the landscape by HiC in mouse E11.5 embryonic limb buds revealed a ~3.8 Mb large CTCF-delimited TAD that contains multiple genes (Fig. 13). Specifically, the *Fat1* gene encodes a large atypical cadherin and is positioned at the telomeric edge of the domain (Peng et al. 2021). Between *Fat1* and the TAD´s boundary is the melatonin receptor encoding gene *Mtnr1a* (Fig. 13). A ~300kb central region in the *Fat1* TAD, termed *Zfp42R*, contains three eutherian- specific genes, the zinc finger protein encoding gene *Zfp42* together with the E3 ubiquitin ligases encoding genes, *Triml1* and *Triml2* (Fig. 13). Adjacent to *Zfp42R* is a rodent-specific cluster of *Adam* disintegrin metalloproteinases (*AdamR*). Despite *AdamR*, the large *Zfp42R* flanking regions contain no further genes and were termed Desert 1 and 2 (D1 and D2) (Fig. 13). In sum, the *Zfp42/Fat1* landscape is a large TAD accommodating multiple genes.

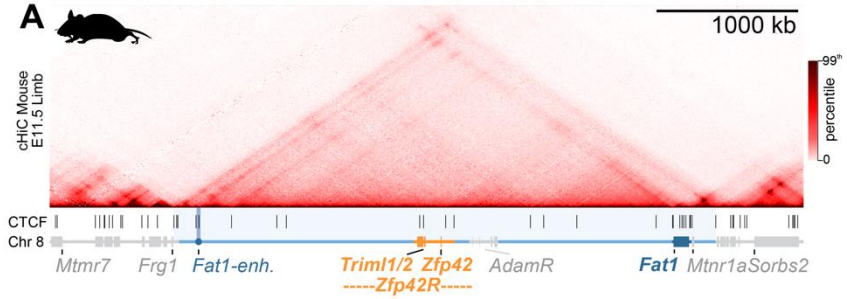


Figure 13 The *Fat1/Zfp42* TAD in mouse limbs
cHi-C from E11.5 mouse limb with corresponding CTCF ChIP-seq peaks below (see table 3). The TAD is highlighted in light blue and genes are shown as boxes: *Fat1*(dark blue), *Mtnr1a* and genes outside the TAD (grey), *Triml1*, *Triml2*, *Zfp42* (*Zfp42R*, orange) and sub-*Adam* region (*AdamR*, grey) containing *Adam26a*, *26b* and *34*.

Next, we investigated how structurally stable the locus was in evolution and if the emergence of the eutherian specific *Zfp42R* genes influenced its structure. For this, we performed HiC in the embryonic limbs of opossum and chicken (Fig. 14 A). Additionally, this structural-mapping was extended by re-mapping publicly available HiC data of differentiated tissues from species spanning the vertebrate family tree (Fig. 14 A). Based on this comparison, we found that only the *Fat1* gene was consistently present at the edge of its TAD in all vertebrates (Fig. 14). Moreover, the *Fat1* TAD is largely maintained as a unit despite frequent flanking synteny breaks and some minor structural alterations during evolution. Specifically, breaks of synteny outside the *Fat1* TAD relocated *Mtnr1a* in zebrafish, *Frg1* in opossum, *Mtmr7* in human, and *Sorbs2* in pig (Fig. 14 A). Interestingly, *Mtnr1a* joined the *Fat1* TAD only in the mammalian linage and, prior to this, occupied its own separate small TAD in chicken and xenopus (Fig. 14 A). However, the most striking evolutionary change occurred in eutherians in the TAD´s center where the *Zfp42R* genes emerged (Fig. 14 A-B). In particular, the pluripotency marker *Zfp42* is known to originate

from a retroposition-driven duplication of the TF *Yin Yang* (Kim et al. 2007; Masui et al. 2008). Similarly, *Triml1* and *Triml2* derive from a gene duplication event in eutherian with *Triml2* being functionally suggested to take part in the evolution of placentation (Zhang et al. 2020b). Finally, single exon-coding *AdamR* genes were also inserted via retroposition (Long et al. 2012). However, the latter genes are not further studied in this work as they are entirely rodent-specific (Fig. 14). Collectively, this evolutionary comparison showed that *Fat1* co-evolved with its structurally conserved TAD in ancestral vertebrates.

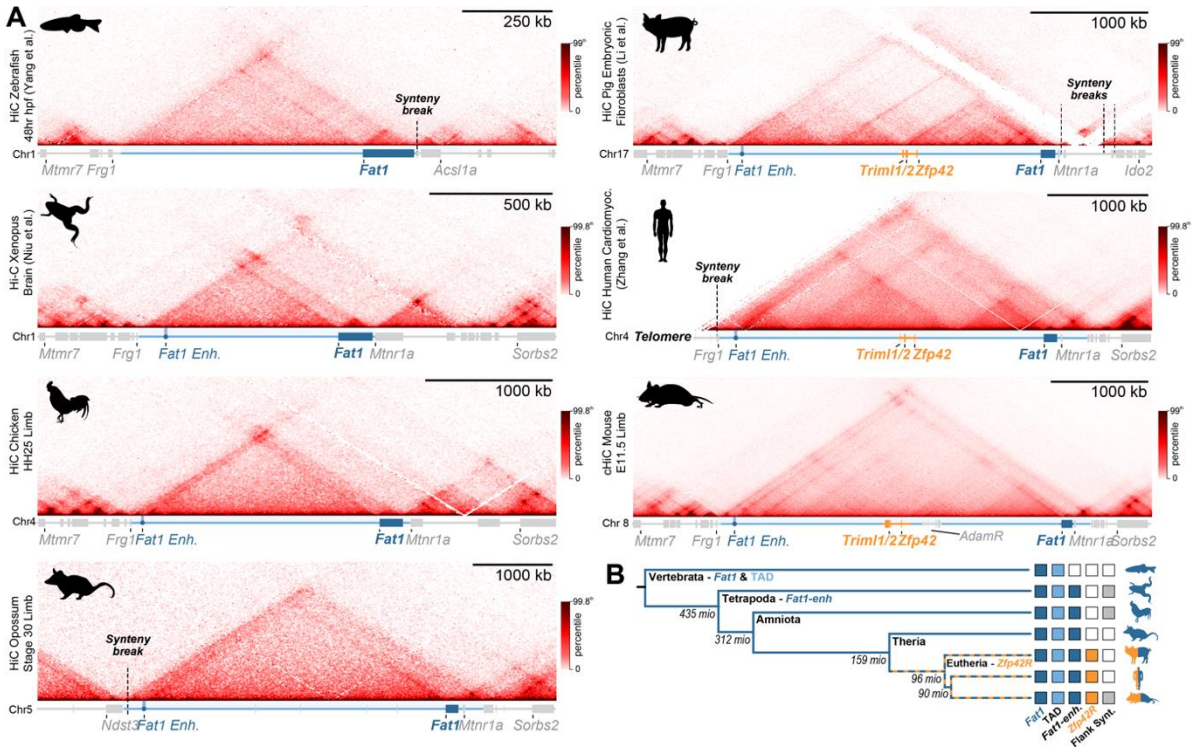


Figure 14 *Zfp42* emerged within *Fat1´*s TAD universally in all placental mammals
A. HiC view at the *Fat1* locus across the vertebrate family tree: zebrafish (Yang et al. 2020), Xenopus (Niu et al. 2021), chicken, opossum, pig (Li et al. 2020), human (Zhang et al. 2019) and mouse. Prior to *Zfp42R* emergence *Fat1* gene (dark blue) solely occupies a conserved TAD (light blue) with synteny beaks (dashed lines) outside of the domain in all vertebrates. The limb *Fat1-enh* emerged in tetrapod's (dark blue circle) and *Mtnr1a* became part of the *Fat1* TAD in the mammalian lineage. *Triml1*, *Triml2* and *Zfp42* emerged in eutherian placental mammals including pigs and humans **B.** Phylogenetic tree showing the emergence of *Zfp42R* in placental mammals. Presence of *Fat1,* the TAD, *Fat1-enh, Zfp42R* and flanking synteny is shown for species indicated on the right. Absence of a feature is indicated with an uncolored box.

1.2 Fat1´s conserved limb expression is driven by the TAD´s enhancer landscape

As the *Fat1* TAD is structurally conserved in evolution, we next examined if it is also functionally conserved. For this, putative enhancers were mapped by ATAC-seq in morphologically stage-matched limbs of chicken, opossum and mouse embryos. This revealed functional conserved elements but also evolutionary turnover of putative enhancers within the landscape. Specifically, the functional mapping of putative enhancers

identified 25, 62 and 49 elements in chicken, opossum and mouse limbs, respectively (Fig. 15 A-C). Most of these elements are found within *Fat1´*s introns and at the opposite side of the TAD in D1 (Fig. 15 A-C). With a pairwise sequence alignment approach using multiple "bridging species, ATAC-seq peaks were projected and classified according to their functional conservation and position (*analysis was done by T.Z., for details see methods 15*) (Baranasic et al. 2021). This revealed that 24-76% of the ATAC peaks showed no sequence conservation and were entirely species-specific suggesting that enhancer turnover occurred at the locus over time (NC, grey) (Fig. 15 D). Furthermore, 12-49% of the ATAC peaks lacked an ATAC signal in the comparison species, indicating function was not maintained despite direct/indirect sequence conservation (green). Finally, 5-27% of ATAC peaks were functionally conserved between each species (red) (Fig. 15 D). This latter class universally included the *Fat1* promoter itself and a conserved element termed *Fat1-enh*. (Fig. 15 A-D). Thus, the *Fat1* TAD´s enhancer composition in the limbs of all three species is partially conserved but also displays enhancer turnover.

To further demonstrate that the enhancer landscape does indeed contain the regulatory information driving *Fat1* expression, we validated *Fat1-enh*. in mice through an enhancer-reporter-assay. For this, *Fat1-enh*. was integrated together with a *ß-globin* minimal promoter driving a *LacZ* gene at the safe harbor locus *Col1a1*. This allowed us to qualitatively characterize *Fat1-enh*.´s *in vivo* activity and demonstrated its capacity to drive a *Fat1*-like expression in the proximal limb and neural tube of E11.5 embryos (Fig. 15 E). Thus, *Fat1-enh*. is a highly conserved enhancer for *Fat1*. Additionally, to directly test its role in regulating *Fat1*, we further deleted *Fat1-enh.* in E11.5 embryos and examined their limbs with RNA-seq. Unexpectedly, *Fat1-enh.* deletion showed no impact on *Fat1* expression, suggesting the element acts as a "shadow enhancer" in a redundant manner with other elements, as observed at other loci in mouse (Osterwalder et al. 2018) (Fig. 15 F).

To test regulatory redundancy, we next deleted entire enhancer clusters found in the domain. Specifically, using CRISPR/Cas9 we eliminated either Zfp42R, desserts D1 and D2 solely or combined (ΔZfp42R, ΔD1, ΔD2, ΔD1+2) in ES cells that were then used to generate transgenic embryos through tetraploid aggregation (Fig. 11). We analyzed how the loss of these sub-regions impacted *Fat1* expression in E11.5 mutant limb buds (Fig. 15 G). This revealed that *Fat1* expression is significantly impacted in ΔD1 and ΔD1+2 limbs while the loss of D2 and Zfp42R has no significant effect on *Fat1* expression. Thus, a redundant enhancer landscape drives the *Fat1* expression in the TAD, with most enhancers clustering in D1. Combined, this demonstrates *Fat1* occupies a structurally and

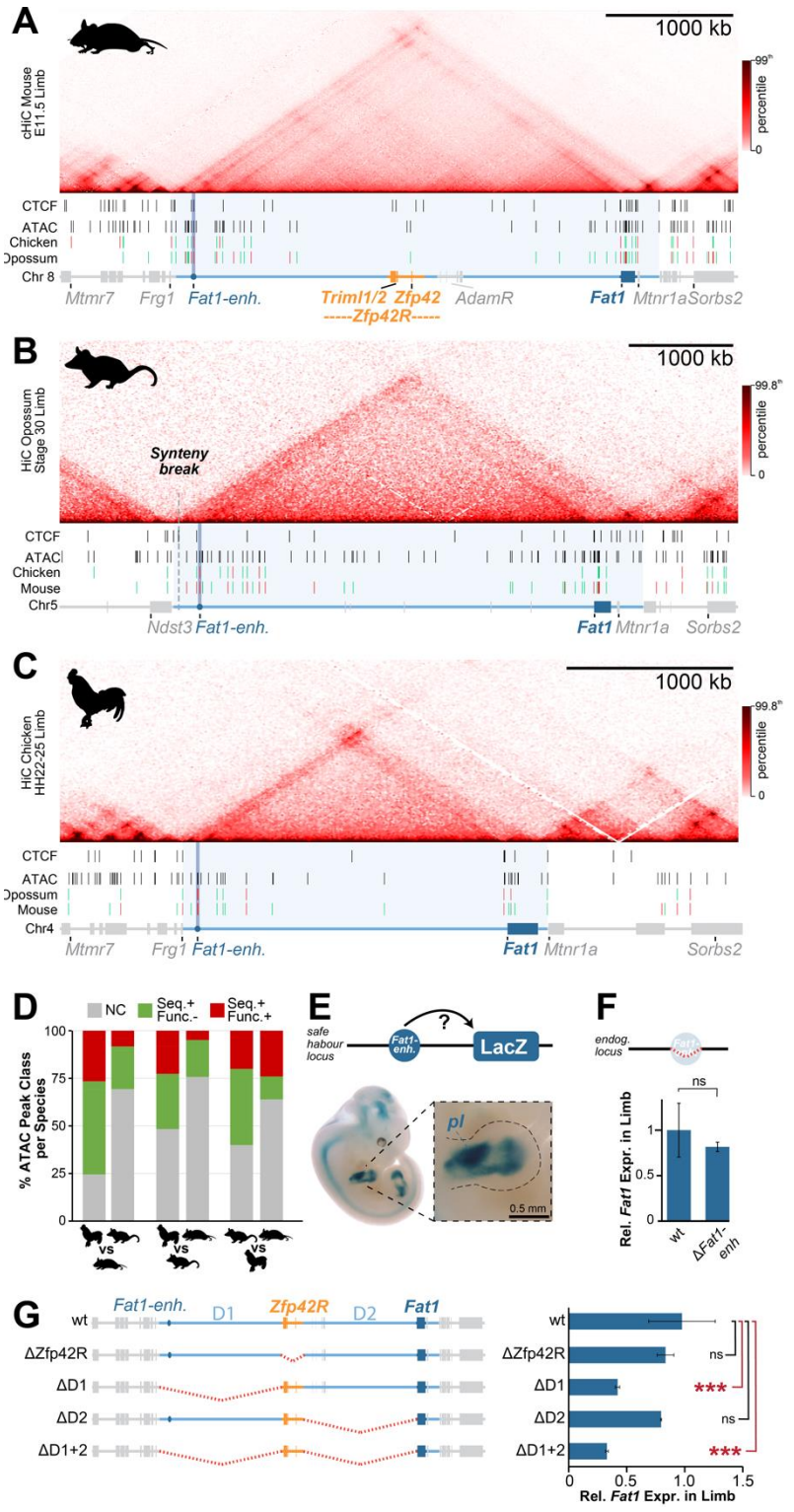


Figure 15 The *Fat1* ´s TAD is largely structurally and in parts functionally conserved
A-C. Hi-C or cHi-C from mouse (A), opossum (B), and chicken (C) in embryonic limb buds with corresponding ATAC-seq and CTCF ChIP-seq peaks below. See table 3. The TAD is highlighted in light blue and genes are shown as boxes: *Fat1*(dark blue), *Mtnr1a* and genes outside the TAD (grey) (A-C); *Triml1*, *Triml2* and *Zfp42* (Zfp42R, orange) (A). ATAC peaks are colored according to their conservation. Called peaks are compared to the indicated species and classified as non-conserved (grey: NC), sequence conserved (green: seq+, func.-) or functionally conserved (red: seq.+, func.+) *(Evolutionary conservation analysis was done by T.Z.)*. **D**. Summarized quantification of ATAC-peak classification. **E-F.** The sequence conserved enhancer element *Fat1-enh*. (dark blue circle highlighted in A-C) was cloned and integrated at *Col1a1* safe harbor locus

for a *LacZ* reporter assay (E), and eliminated at its endogenous (endog.) locus followed by RNA-seq expression analysis (F). n = 4, ns = non-significant. **G**. Schematic of the generated deletions (left) and their impact on *Fat1* expression by RNA-seq (right). n = 2-4 samples per condition, ns = non-significant, *** p<0.001.

partially functionally conserved TAD landscape that has governed it´s regulation for over 300 million years.

1.3 The Fat1 TAD drives conserved limb expression

Due to its structural and function conservation, we suggested that *Fat1*'s landscape drives a conserved *Fat1* expression. To test this, we performed species-specific whole amount *in situ* hybridization (WISH) for *Fat1* in chicken, opossum and mouse embryos. This revealed a largely conserved *Fat1* expression pattern in limbs of all three species (Fig. 16). Specifically, *Fat1* WISH detects elaborate activity patterns specifically in the forebrain, ear, snout, mammary glands, proximal and distal limbs of E11.5 mouse embryos (Fig. 16). (Ciani et al. 2003; Helmbacher 2018). Though the *Fat1* pattern seems similar in chicken and opossum embryos, a direct one-to-one comparison of each morphological feature and their staining is difficult due to their divergent embryonic morphology (Fig. 16). Nevertheless, related *Fat1* expression patterns between the species are detectable. For instance, *Fat1* staining is observed in the embryonic ear of mice and in the brachial arches, from which the ear derives, are *Fat1*-positive in chicken and opossum embryos (Fig. 16). Despite this, the morphologically stage-matched forelimbs of chicken, opossum and mouse embryos allow direct comparison and indeed show a remarkably conserved *Fat1* expression pattern (Fig. 16). Thus, the *Fat1* landscape drives a limb expression pattern that is conserved across 300 million years of evolutionary time.

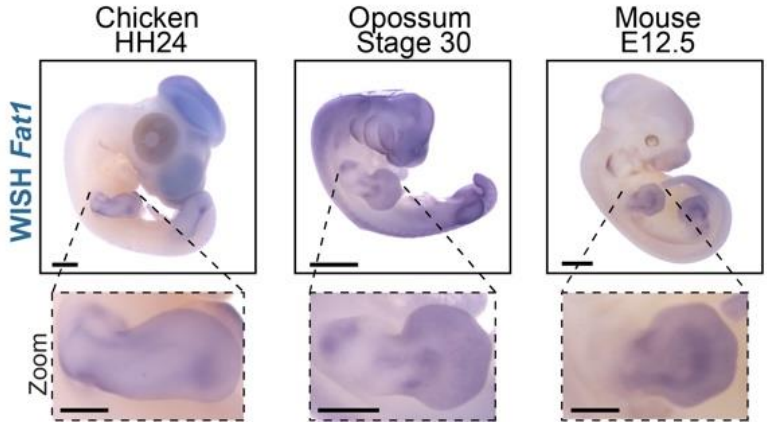


Figure 16 Conserved *Fat1* expression pattern in embryonic limbs
Species specific *Fat1* WISH in chicken (left), opossum (middle) and mouse (right) embryos with zoom of morphologically stage-matched embryonic fore limbs below. Scale bar: 1 mm

1.4 Fat1 and Zfp42R genes are independently and divergently expressed

Based on models of how TADs function, one would predict that the *Zfp42R* genes and *Mtnr1a* have a *Fat1*-like expression as they are all exposed to the same enhancers (Dixon et al. 2016). To test if the *Fat1*/*Zfp42* TAD acts as a conserved co-regulatory unit, we comprehensively investigated the genes´ developmental expression programs by WISH, publicly available Fantom5 CAGE and single-cell RNA-seq (scRNA-seq) datasets (Fig. 17) (Pijuan-Sala et al. 2019; Marsh and Blelloch 2020; Cao et al. 2019) *(scRNA-seq replotting was done by C.A.P.M.; CAGE data analysis was done by D.H*.).

This analysis surprisingly revealed that the TAD´s genes are not co-regulated in most tissues which is visibly most evident by WISH. While *Fat1* displays an elaborate expression pattern in E11.5 embryos, *Triml1/2* and *Zfp42* WISH produced no detectable staining, as reported previously (Fig. 17 A) (Kim et al. 2011; Zhang et al. 2020a). Likewise, CAGE and scRNA-seq find no gene co-regulation (Fig. 17 B-F). Specifically, co-transcription for *Zfp42R* genes and *Fat1* was detected in mouse embryonic stem cells (ESCs), placental trophoblasts stem cells (TSC), and the E6.5-8.5 extraembryonic ecto- and endoderm (Fig. 17 D-F). However, *Zfp42R* genes are repressed in essentially all later embryonic cell-types following gastrulation despite continued *Fat1* transcription (Fig. 17 D-F). Furthermore, no detectable *Mtnr1a* expression was observed in examined CAGE and scRNA-seq data, matching past reports of its restriction to the Suprachiasmatic nucleus and pars tuberalis (Fig. 17 B and D-F) (Klosen et al. 2019). As these tissues are experimentally inaccessible, *Mntr1a* was excluded from all further analyses.

Collectively, though co-transcribed in some early developmental stages and extraembryonic tissues, *Zfp42R* genes and *Fat1* display a largely independent expression within their shared TAD during later mouse development. This demonstrates that the previously monogenic *Fat1* TAD integrated novel and divergently expressed genes without disrupting its pre-existing *Fat1* expression program at least in the embryonic limbs. Next, we wanted to understand what mechanisms enabled these genes to be divergently expressed. In the following chapter, we thus focused on mapping the chromatin structures in ESCs and embryonic limbs where *Fat1* and *Zfp42R* genes have divergent expression.

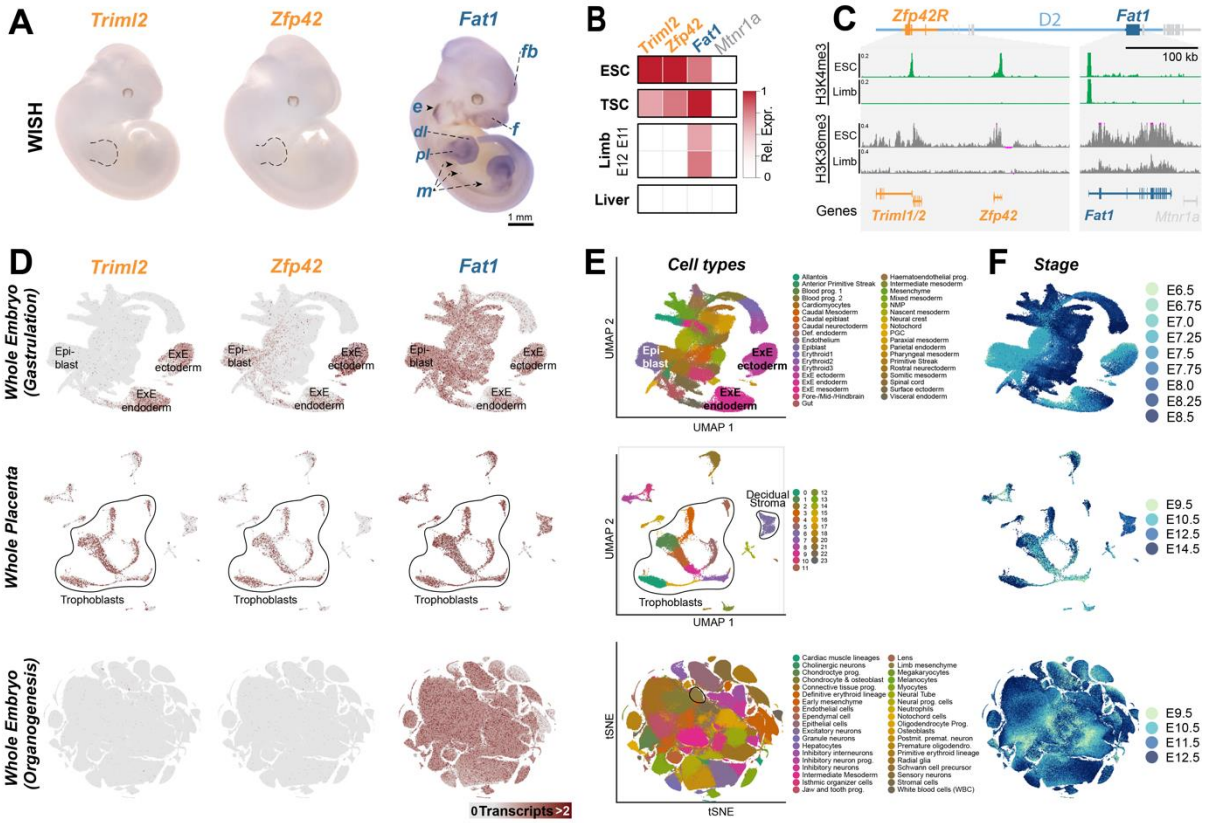


Figure 17 *Zfp42* displays a *Fat1*-independent activity program
A. WISH of *Triml2*, *Zfp42* and *Fat1* in E11.5 mouse embryos. *Fat1* expression is indicated in the developing ear (e), forebrain (fb), face (f), mammary glands (m), proximal (pl) and distal limb (dl). **B.** Relative expression of *Triml2*, *Zfp42*, *Fat1* and *Mtnr1a* in indicated tissues from Fantom5 Cap analysis gene expression (CAGE) datasets. Expression of all genes is absent in liver. *Mtnr1a* shows no expression in all investigated tissues. Trophoblast stem cells (TSC), embryonic stem cells (ESC). See table 3 (*CAGE data analysis was done by D.H.)* **C.** In mice, *Triml1* and *Triml2* have a share bidirectional promoter as indicated by a single peak of transcription associated histone modifications H3K4me3 and H3K36me3 in ESCs. In ESCs *Triml2*, *Zfp42*, *Fat1* are active together and in embryonic limb tissue only *Fat1* is active. Remapped data see table 3. **D-F.** Re-plotted published single cell RNA-seq data from E6.5-E8.5 gastrulating embryos; top (Pijuan-Sala et al. 2019), E9.5-E14.5 placentas; middle (Marsh and Blelloch 2020) and E9.5-E12.5 embryos during organogenesis; bottom (Cao et al. 2019). Visualization through UMAP or tSNE embeddings according to cell types (D) and embryonic stage (E), respectively. Expression of *Triml2*, *Zfp42*, and *Fat1* plotted as UMI counts in the range 0 to >2 (F).: *Zfp42R genes* are active together during early mouse development and trophectoderm derived placental tissues: Trophoblast stem cells (TSC), Epiblast (EPI), extraembryonic ecto- (EXE) and endoderm (EXN). *Fat1* is active in the same tissues as *Zfp42R* and in nearly all tissues of later mouse developmental stage. (*scRNA-seq replotting was done by C.A.P.M.).*

2. Tissue-specific structural reconfiguration of the *Fat1/Zfp42* landscape

Originally, it was thought that TADs are stable across different tissues during development and mostly conserved across evolution (Dixon et al. 2012; Vietri Rudan et al. 2015; Harmston et al. 2017; Fraser et al. 2015). However, recently, TADs have been reported as flexible and adaptable structures (Acemel et al. 2016; Eres et al. 2019; Torosin et al. 2020; Luo et al. 2021; Bonev et al. 2017). In particular, TADs have been observed to be altered as an evolutionary adaptation to implement neofunctionalization events of genes

(Acemel et al. 2016; Andrey et al. 2013). Moreover, several reports showed tissue-specific restructuring of TADs coinciding with gene expression changes (Bonev and Cavalli 2016; Winick-Ng et al. 2021b). Therefore, in this second section, we investigated if the emergence of *Zfp42* altered the chromatin structure of the ancient *Fat1* landscape. To map locus structure, we combined genomic approaches, super-resolution microscopy and *in silico* polymer modeling. With this, we demonstrated that, despite the landscape's conservation, massive 3D chromatin restructuring between ESCs and embryonic limb buds coincides with the genes' activities.

2.1 Independent Zfp42R genes and Fat1 activity in ESCs correlates with locus restructuring

By examining the locus' structure, we tested if its architecture is influenced by *Zfp42R'*s emergence and if this might be linked to *Zfp42'*s and *Fat1'*s divergent regulation. For this, we first mapped chromatin interactions with HiC and NE association by DamID-seq. Surprisingly, this revealed that the locus structure is extensively reconfigured between ESCs and embryonic limb buds, despite being conserved over 300 million years of evolution (Fig. 18). Particularly, in limbs, the intact TAD can support active and inactive regions at the same time. DamID-seq demonstrates that the majority of the *Fat1/Zfp42* TAD is attached at the NE (Fig. 18 A). Specifically, the inactive *Zfp42R* is incorporated together with its ~1.5 Mb flanking D1 and D2 in a large repressed B compartment LAD (Fig. 18 A). In contrast, the active *Fat1* and its *Fat1-enh*. is locally detached from the NE in a non-LAD (Fig. 18 A). This detachment coincides with H3K27ac ChIP-seq peaks within the *Fat1* gene body and active enhancers at the centromeric TAD side in limb (Fig. 18 A). Thus, the limb TAD successfully hosts chromatin regions of different activity states.

However, surprisingly, the TAD structure entirely changes in ESCs. Here, extended epigenetic profiling reveals a complete redistribution of H3K27ac-marks. Two distinct H3K27ac-marked enhancer clusters are present within *Zfp42R* and the *Fat1* gene body (Bauer et al. 2021) (Fig. 18 B). This altered enhancer landscape correlates with a radically changed chromatin structure observed by HiC and NE association (Fig. 18 B). In ESCs, the now active *Zfp42R* and *Fat1* lose interactions with D1 and D2 accompanied with *Zfp42R´*s NE release (Fig. 18 B). As a result, the limb TAD becomes partitioned into four discrete smaller domains in ESCs where *Zfp42R* and *Fat1* are physically separated from each together along with their own local enhancer clusters (Fig. 18 B). Significantly, this locus partitioning is also observed in other public HiC in cell types where *Zfp42R* and *Fat1* are simultaneously active (Du et al. 2017; Zhang et al. 2019). Likewise, public HiC in cell types where only *Fat1* is active display the evolutionarily conserved intact TAD structure

as found in limbs (Fig. 18 A) (Takebayashi et al. 2012). Hence, although evolutionarily stable, the *Fat1*/*Zfp42* TAD can flexibly adapt its structure in tissues when both genes are simultaneously active. We thus were excited to postulate that this directly links the altered structure to the genes' activities.

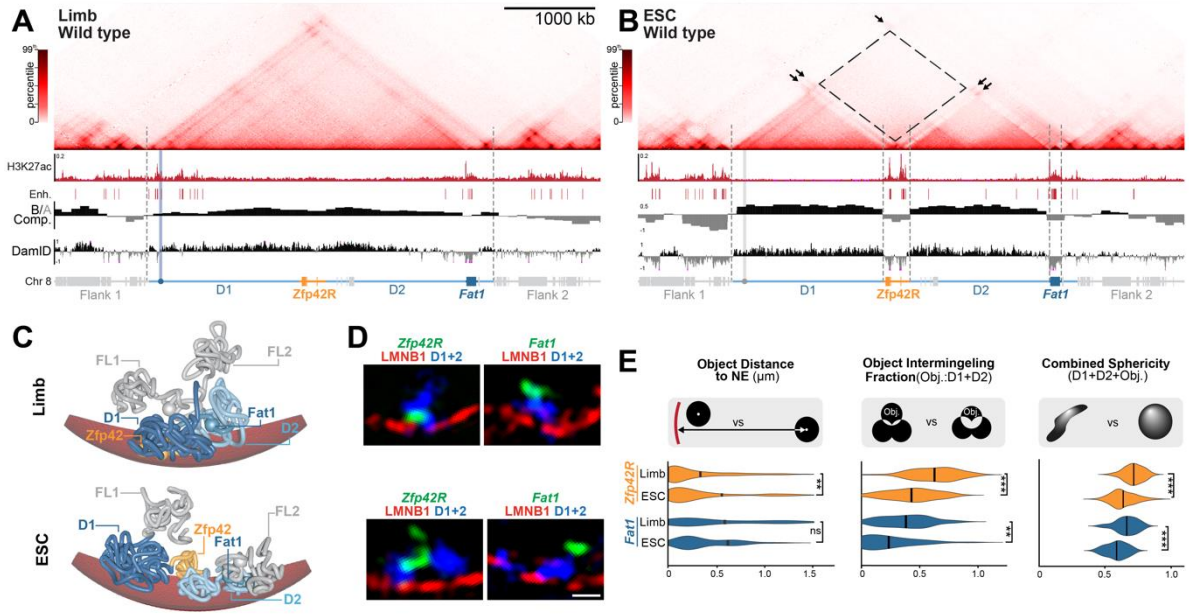


Figure 18 The *Fat1/Zfp42* TAD in limb massively reconfigures in ESCs
A-B. cHi-C from E11.5 mouse limb buds (A) and ESCs (B) with H3K27ac ChIP-seq, called putative enhancers, called B/A- compartments and DamID-seq. See table 3. Black arrows indicate interactions between active compartments and dotted rectangle highlights area of lost interaction between inactive Dessert 1 and 2 (D1, D2). **C.** Representative structure from modified strings-and-binders polymer model with simulated attachment to the NE (brown) in limb buds (top) and ESCs (bottom). Genomic adjacent regions are Flank1 and 2 (FL1 and FL2, grey), D1 and D2 (dark and light blue), Zfp42R (orange), *Zfp42* and *Fat1* are highlighted as balls **D.** Representative Z-slice of Lamin B1 immunostaining (red) with Oligopaint-stained D1+D2 (blue) and *Zfp42R* or *Fat1* (green) in limb buds (top) and ESCs (bottom). Scale bar: 500 nm. **E**. FISH measurements in wildtype E11.5 limb buds and ESCs. Following measurements are shown: Object centroid distance to the NE (left), intermingling fraction of objects with D1+D2 (middle) and combined sphericity of objects with D1+D2 (right). *** p<0.001 and ** p<0.01 from Welch's t-test comparisons between indicated samples. n = 28-138 alleles of at least two biological replicates.

2.2 3D-SIM visualizes the true physical chromatin structure at single alleles supported by mSBS modelling

Examining ESCs suggests 3D restructuring as an evolutionary adaptation that allowed the locus to successfully incorporate *Zfp42R*. However, because HiC and DamID results are performed separately on bulk cell populations, it is not clear if the altered NE-attachment and chromatin structure indeed coincide at the same single alleles. To overcome the limitations of such genomic approaches, we additionally examined the locus structure with 3D-structured illumination microscopy and *in silico* polymer simulations. With this, we

showed that the initial observed structural features in the bulk experiments indeed do occur simultaneously at the single alleles.

For the direct visualization of structures, we applied oligopaint FISH for chromatin staining combined with lamin b1 immunoabelling for NE staining in 3D-structured illumination microscopy (3D-SIM) (Fig. 18 D-E) (Gustafsson et al. 2008; Beliveau et al. 2012; Beliveau et al. 2015; Beliveau et al. 2018; Szabo et al. 2018; Szabo et al. 2020; Schermelleh et al. 2008). In addition, we generated an ensemble of simulated polymer structures *in silico* using a modified strings & binders switch model (mSBS) (Fig. 18 C) *(in silico modelling was done by A.M.C, S.B. and A.E., for details see methods 12)* (Barbieri et al. 2012; Chiariello et al. 2016; Nicodemi and Prisco 2009; Bianco et al. 2018). In brief, this modeling approach simulates chromatin as a self-avoiding polymer string consisting of n beads that serve as binding site for diffusing molecules -*binders*- (Esposito et al. 2021; Barbieri et al. 2012; Nicodemi and Prisco 2009). Interactions between bound binders then, over time, allow 3D polymer structures to evolve (Bianco et al. 2018). In all cases, binding affinities to one another is determined from experimental Hi-C data (Esposito et al. 2021; Barbieri et al. 2012; Nicodemi and Prisco 2009). Finally, the interaction of these 3D structures with the NE is then simulated with affinities determined from matched DamID-seq data (Fig. 18 A).

From these two approaches we extracted measurements of NE-distance, intermingling fraction and combined sphericity of computationally simulated and microscopically observed objects: Zfp42R, Fat1, D1, D2 and flanking regions 1 and 2 (Fl1 and Fl2) (Fig. 18 and 19). Significantly, the polymer-NE interaction was newly integrated into the modeling approach and so required refinement. Simulations of the structure's NE attachment were performed using three different interaction energies (0.4, 1.2 and 3.0 kTb) and the results were observed in reconstructed cHiC maps (r-cHiC) based on 25-88 simulated polymer structures (Fig. 19 B-C). From these results, we selected the medium 1.2 kTb energy for our analyses as it most closely matched the NE-proximity data acquired by FISH (Fig. 19 E). Specifically, the low 0.4 kTb interactions resulted in insufficient NE association while the high 3.0 kTb NE interactions deformed loci into flattened pancake -like structures that lacked internal polymer interactions. This lack of internal interactions in comparison to the unattached controls is strongly visible in the subtraction r-cHiC (Fig 19 C). Thus, it seems that *in vivo* the interactions between chromatin itself must be higher than their affinity to the NE to prevent such flatten structures which we do not observe microscopically.

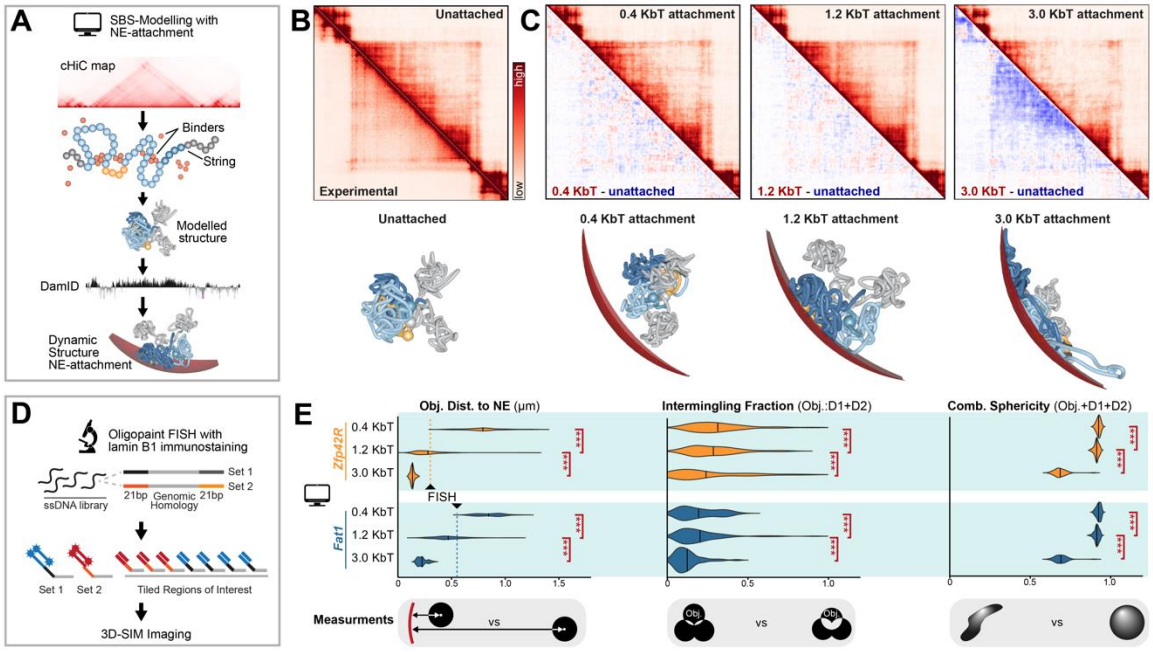


Figure 19 Principle and optimization of mSBS Modelling and Oligopaint FISH strategy
A. Schematic representation of the modified strings&binders switch (mSBS) model. Experimental cHiC data were used to define PRISM-assigned chromatin binders on a polymer string. The polymer is then structured through simulated DNA interactions created by the self-association between matching binders (Barbieri et al. 2012; Nicodemi and Prisco 2009). With this diverse polymer structures were generated and afterwards their interaction with a simulated NE was modelled based on matching DamID-seq. **B and C.** Reconstituted contact maps from simulated limb polymer structures before (B) and after (C) interaction with the NE tested with simulated energies of 0.4, 1.2 and 3.0 kTb. Subtraction maps and representative polymer structures are shown below. n = 25-88 simulations. **D.** 3D-SIM imaging strategy using Oligopaint FISH. Single-stranded Oligopaints were used to visualize the chromatin region of interest. Oligopaint probes consist of 32-mer genomic sequence with flanking non-genomic regions containing primers for amplification and labelling with fluorophores (Beliveau et al. 2015) **E.** Quantification of NE-distance (left), intermingling fraction (middle) and combined sphericity (right) for *Zfp42R* and *Fat1*, respectively. Quantifications from simulated structures are shown for NE-attachment energies 0.4, 1.2 and 3.0 kTb NE-attachment. NE-proximities from FISH measurements are indicated as dashed lines. *(in silico modelling was done by A.M.C, S.B. and A.E.)*

After refining the polymer simulations, all measurements of observed and simulated objects were compared one-to-one and showed -except for specific NE-proximity measurements (see results 4.1)- the same statistically significant trends in wildtype limbs and ESCs (Fig. 20 A-D). Hence, unless stated, the trends described below are shared between observed and simulated objects. As expected, simulations and imaging showed the same radical structural change between limb buds and ESCs as seen by HiC and DamID. The limb structure becomes indeed physically disassembled into smaller domains in ESCs. As such, in ESCs, D1 and D2 significantly lose interaction between themselves and moreover, their combined intermingling with *Zfp42R* or *Fat1* is significantly decreased (Fig. 18 E). Likewise, the D1+D2-combined sphericity with *Zfp42R* or *Fat1* is significantly lower in ESCs (Fig. 18 E). Thus, collectively this indicates that all objects exist as physically separated structures in ESCs. Furthermore, *Zfp42R´*s NE-distance significantly changes.

Zfp42R is preferentially positioned within 300 nm proximal to the NE in limbs and significantly increases its NE distances in ESCs thereby matching the DamID results (Fig. 18 A-B and E).

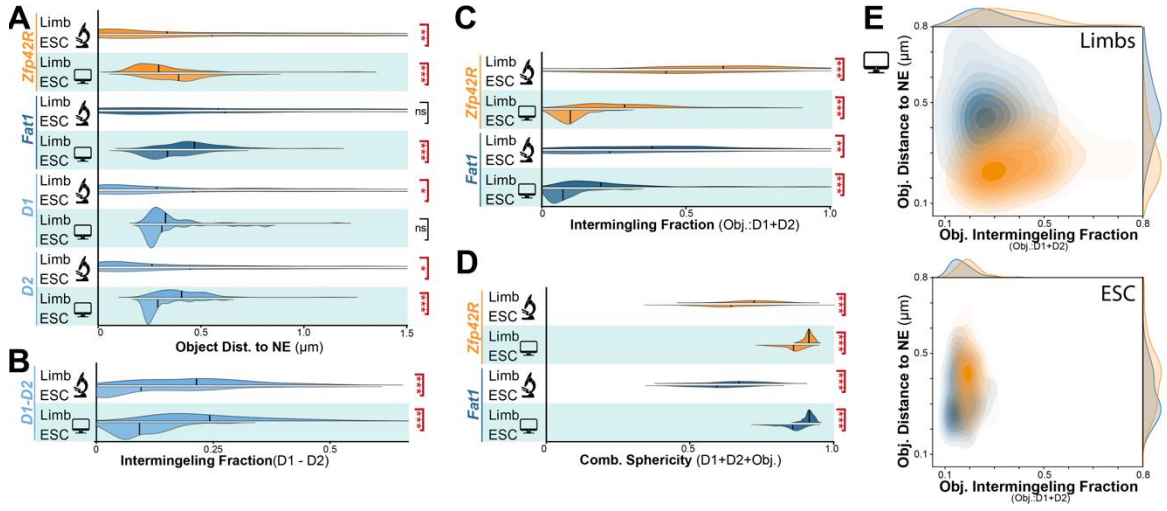


Figure 20 Comparison of simulated and observed structures
A-D. Quantification of NE-distance (B), intermingling fraction (B+C) and combined sphericity (D) for indicated objects. All simulated structures have been generated with a 1.2 kTb NE-attachment energy. **E**. Density plots of NE-distance and simultaneous D1+D2 intermingling for Zfp42R (orange) and Fat1 (blue), respectively. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05 from Welch's t-test comparisons between indicated samples. FISH n=28-138 alleles of at least two biological replicates. Modelling n= 87-105 simulated structures. *(in silico modelling was done by A.M.C, S.B. and A.E.)*

Despite this, the computational simulations convey additional information. The simultaneous readout of NE-distance and D1+D2 intermingling fraction with *Zfp42R* or *Fat1* from the same polymer structures indicates different structural dynamics (Fig. 20 E). Specifically, in ESCs, the data points of *Zfp42R* and *Fat1* show low horizontal distribution for the intermingling with D1+D2 intermingling but broader variability in the NE-distances (Fig. 20 E). In contrast, the density blots in limbs display a broader signal distribution, presumably as all chromatin is frequently mixed within the TAD (Fig. 20 E). Nonetheless, *Zfp42* is preferentially positioned at the NE while active *Fat1* transcription seems to occur preferentially NE-distally (Fig. 20 E). Thus, these results collectively highlight the variability of structures and the likely preferred intra-TAD position of genes in limbs. In summary, *in silico* modeling and imaging confirmed that the distinct chromatin structures and NE-associations detected by bulk approaches directly co-exist together at single loci.

Altogether, we showed that the evolutionarily conserved locus has a remarkable tissue -specific structural flexibility which suggests that the TAD has indeed adapted to the insertion of *Zfp42R* genes. On one hand, it shows that the limb TAD can support simultaneously active *Fat1* and inactive *Zfp42* regulation. However, on the other hand, in

ESCs the TAD is partitioned according to the chromatins underlying activity state into smaller domains. Therefore, we next addressed if these structural features drive the independent regulation of *Zfp42R* and *Fat1* genes.

3. *Zfp42R's* and *Fat1's* simultaneous activity in ESCs

This third section investigates the evolutionary adaptation of locus restructuring by examining how it influences gene regulation. For this, we performed various targeted deletions to find that the landscape's structural remodeling enables the independent regulation of *Zfp42* and *Fat1*. Next, we tested two potential forces that could drive this structural change, loop extrusion and compartmentalization. Loop extrusion drives the formation of TADs through an interplay between cohesin and the architectural insulator CTCF (Heger et al. 2012; Fudenberg et al. 2016). In contrast, chromatin's tendency to segregate based on its epigenetic identity forms compartments (Lieberman-Aiden et al. 2009; Rao et al. 2014). We used induced protein degradation and targeted genomic deletion approaches to manipulate cohesin loop extrusion and chromatin's epigenetic signature. With this we demonstrate that structural adaptation in ESCs is surprisingly not driven by CTCF/cohesin but instead by the underlying epigenetic state of chromatin.

3.1 Small separated chromatin domains drive independent Zfp42 and Fat1 expression

We first addressed how the landscape´s physical partitioning influences gene regulation. With a series of deletions combined with RNA-seq we tested which enhancers are used by *Zfp42* and *Fat1*. This demonstrated that the restructuring delimits enhancer usage to the separated smaller domain, thereby enabling a simultaneous but functionally independent regulation of *Zfp42* and *Fat1* in ESCs.

Specifically, in the ESC-specific structure *Zfp42* and *Fat1* are found in separated smaller domains that each overlap with a distinct enhancer cluster that suggests their functional independency. To test this, we used ESCs mutants, generated in section one, that either lacked the evolutionary newer *Zfp42R* or the more ancient D1 and D2 (ΔZfp42R, ΔD1, ΔD2, ΔD1+2) (Fig. 21 B). Subsequent RNA-seq showed that neither the single nor combined deletion of D1 and D2 significantly affected *Zfp42* expression in ESCs (Fig. 21 C). Similarly, *Fat1* transcription is not affected in ΔD1, ΔD2, ΔD1+2 but excitingly also not in ΔZfp42R (Fig. 21 B). Only *Triml2* expression was reduced in ΔD2 but oddly not ΔD1+2 (Fig. 21 B). Thus, combined this shows that eliminating ~80% of landscape and only leaving the *Zfp42*- and *Fat1*-containing domains intact has no impact on their expression in ΔD1+2 ESCs (Fig. 21 B). Hence, though sharing the same landscape chromatin

remodeling restricts the enhancer availability to the smaller separated domains driving independent *Zfp42* and *Fat1* activity.

In summary, evolution adapted the landscape´s structure to re-assign enhancer-promoter communications when needed. Therefore, we asked next what forces drive this chromatin reconfiguration.

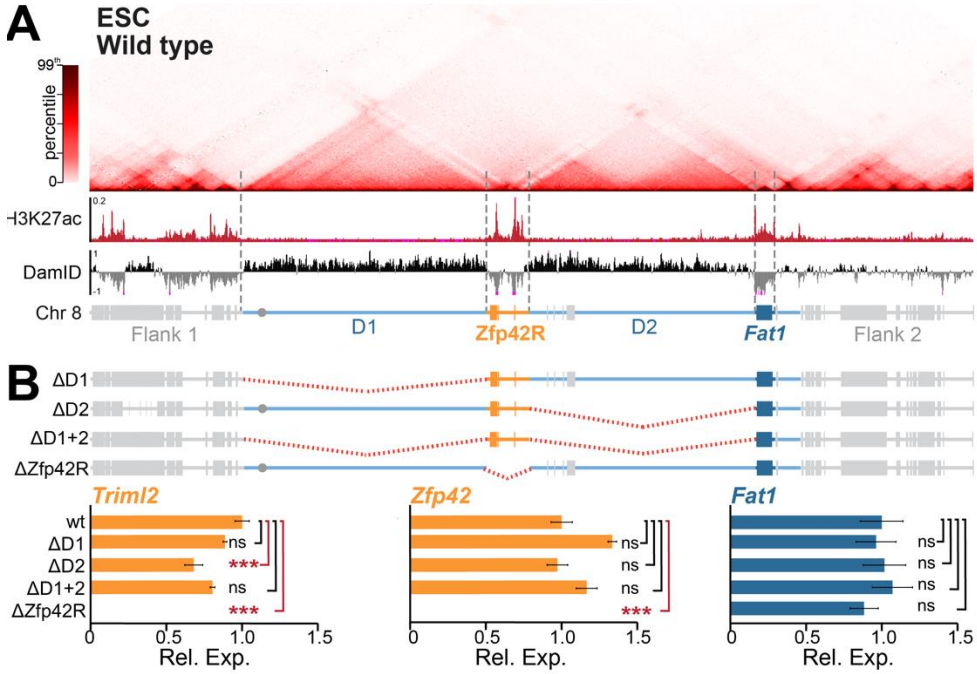


Figure 21 Structural reconfiguration enables independent *Zfp42* and *Fat1* regulation
A. cHi-C from ESCs with H3K27ac ChIP-seq and DamID-seq below. **B.** Schematic representation of CRISPR-mediated deletion: ΔD1, ΔD2, ΔD1+2 and ΔZfp42R (top)**.** Relative expression of *Triml2* (left) *Zfp42* (middle) and *Fat1*(right) from RNA-seq in wt and mutant ESCs (bottom). n = 2-4 biological replicates per condition. ns p>0.05 from DEseq for differential gene expression (Love et al. 2014).

3.2 Fat1/Zfp42 locus restructuring is independent of Rad21/CTCF

Aiming to understand which forces drive the 3D-restructuring in ESC, we first tested if loop extrusion is necessary or not. *Zfp42R* gains binding of cohesin's Rad21 subunit and CTCF specifically in ESCs, as previously reported (Fig. 22 A) (Bonev et al. 2017). Therefore, we hypothesized that restricting cohesin-loop-extrusion to *Zfp42R* inside the TAD would allow its escape from the domain followed by subsequent locus partitioning. Accordingly, we would expect unconstrained loop extrusion (dCTCF) and its elimination (dRad21) would prevent the locus partitioning.

To test this, we eliminated the cohesin subunit Rad21 or the cohesin-blocking factor CTCF and examined the consequences for the locus structure by cHiC and FISH. For this, we used available auxin-inducible degron systems enabling global depletion of GFP-tagged CTCF or Rad21 in ESCs (Liu et al. 2021b; Nora et al. 2017). To ensure efficient

CTCF/Rad21 depletion GFP-negative cells were isolated by FACs before processing for cHiC (Fig. 22 B). Additionally, Rad21-depleted ESCs were sorted according to G1-cell cycle phase by DAPI staining. This was required as cohesin loss impairs sister chromatid cohesion during mitosis which causes cells to accumulate in the G2/M phase upon longer auxin treatment (Liu et al. 2021b) (Fig. 22 C). Further, samples for imaging could not be FACs sorted due to technical reasons. In this case, the auxin treatment of Rad21-AID-GFP cells was limited to 2 hwhere cell cycle phases were only mildly affected and Rad21 efficiently degraded (Fig. 22 C).

Surprisingly, this demonstrated locus partition persists in absence of CTCF and cohesin (Fig. 22 D-E). Specifically, cHiC of dCTCF ESCs revealed that insulation and TADs flanking the locus are diminished as expected based on previous reports while the *Fat1*/*Zfp42* locus itself is largely unchanged (Fig. 22 D) (Nora et al. 2017; Wutz et al. 2017; Szabo et al. 2020). This structural maintenance is confirmed by FISH displaying no significant alterations in *Zfp42R´*s NE-distance nor *Zfp42R´*s intermingling and combined sphericity with D1+D2 in dCTCF (Fig. 22 F). Likewise, cohesin is also not required with surrounding TADs eliminated following Rad21 depletion while the *Fat1*/*Zfp42* locus is preserved. In fact, Rad21 depletion further intensifies and reinforces partitioning into the four discrete smaller domains matching previous reports where compartments are strengthened in absence of loop extrusion (Fig. 22 E) (Rao et al. 2017). Specifically, FISH demonstrates *Zfp42R´s* significantly reduced intermingling and combined sphericity with D1+D2 while not affecting its NE-proximity in dRad21 (Fig. 22 F).

Combined, these depletion experiments demonstrate that *Zfp42R´*s escape and the locus restructuring capacity are not dependent on loop extrusion. Instead, eliminating loop extrusion intensified partitioning according to chromatin´s epigenetic state suggesting compartmentalization drives partitioning (Wutz et al. 2017; Rao et al. 2017; Rhodes et al. 2020; Thiecke et al. 2020; Aljahani et al. 2022).

3.3 Fat1/Zfp42 locus restructuring is predominantly driven by compartmentalization

Previously, eliminating cohesin was shown to partition TADs into smaller compartments according to epigenetic state, similar to that observed at *Fat1/Zfp42* in ESCs (Wutz et al. 2017; Rao et al. 2017; Rhodes et al. 2020; Thiecke et al. 2020; Aljahani et al. 2022; Goel et al. 2022; Hsieh et al. 2021). Thus, we hypothesized that the locus's structure is driven by compartmentalization and so would be dependent on the epigenetic signature of *Zfp42R* (Lieberman-Aiden et al. 2009).

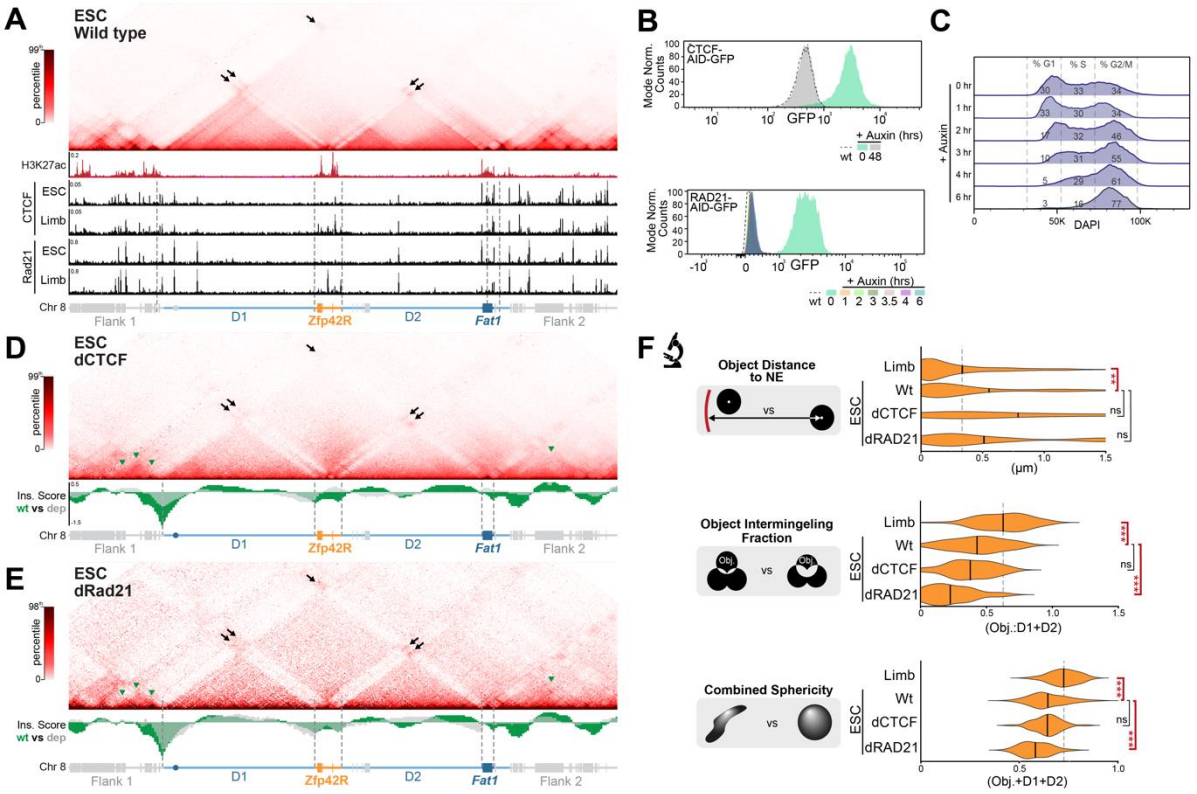


Figure 22 Chromatin restructuring is independent of CTCF and Rad21
A. cHi-C from ESCs with corresponding H3K27ac. Arrows indicate interactions between H3K27ac marked active regions. Below are corresponding CTCF and Rad21 ChIP-seq in wildtype ESCs and limb. See table 3. **B-C.** Fluorescence activated cell sorting (FACs) of dCTCF and dRad21 ESCs: Loss of GFP signal upon auxin treatment in CTCF-AID-GFP (top) and Rad21-AID-GFP (bottom) ESCs (B) Quantification of cell cycle phases in Rad21-AID-GFP ESCs. Histogram of DAPI signal shows increasing auxin treatment times causes accumulation of cells in G2/M phase (C). **D-E.** c-HiC from dCTCF (D) and dRad21 (E) ESCs with Insulation Score (Ins. Score) of wt (green) and corresponding depletion mutant (grey) below. Interactions between H3K27ac marked active regions are highlighted by black arrows and TAD disruptions in mutants are highlighted by green arrows **F.** Quantification of FISH measurements for Zfp42R's NE-distance, Zfp42R's intermingling fraction with D1+D2, and their combined Sphericity (Zfp42R+D1+D2) in wt limbs and wt, dCTCF, and dRad21 ESCs.

To test this, we thus progressively eliminated *Zfp42R´*s active marks and examined the effect on the locus structure by cHiC and DamID-seq. Sequentially removing H3K27ac peaks at *Zfp42* (ΔZfp42) and *Triml1/2* (ΔZfp42+Triml) led to a progressive loss of structural separation of *Zfp42R* from flanking D1 and D2 by cHiC (Fig. 23 B). Matching this, LAD DamID signal spreads into *Zfp42R* and gradually invades the normally non-LAD in ESCs (Fig. 23 C-D). Importantly, *Zfp42R* together with D1 and D2 regains interactions to partially restore a limb-like TAD (Fig. 23 D). As such, Hi-C interactions between D1 and D2 are progressively gained in in ΔZfp42 and ΔZfp42+Triml ESCs while *Fat1*:D1 contacts are unchanged (Fig. 23 E). Nevertheless, *Fat1* remains separated in its own domain due to its continued active signature that that avoids interactions with the inactive landscape.

To conclude, the *Fat1/Zfp42* locus restructuring in ESCs is predominantly driven by the epigenetic signature of underlying chromatin. Thus, this excitingly demonstrates that

the activity of chromatin itself can massively influence its structure. Moreover, this reveals that the antagonistic balance between loop extrusion and compartmentalization can serve as a powerful substrate for evolutionary adaptation. In this case, altering this balance allows two genes to be simultaneously active but through the use of entirely independent enhancers.

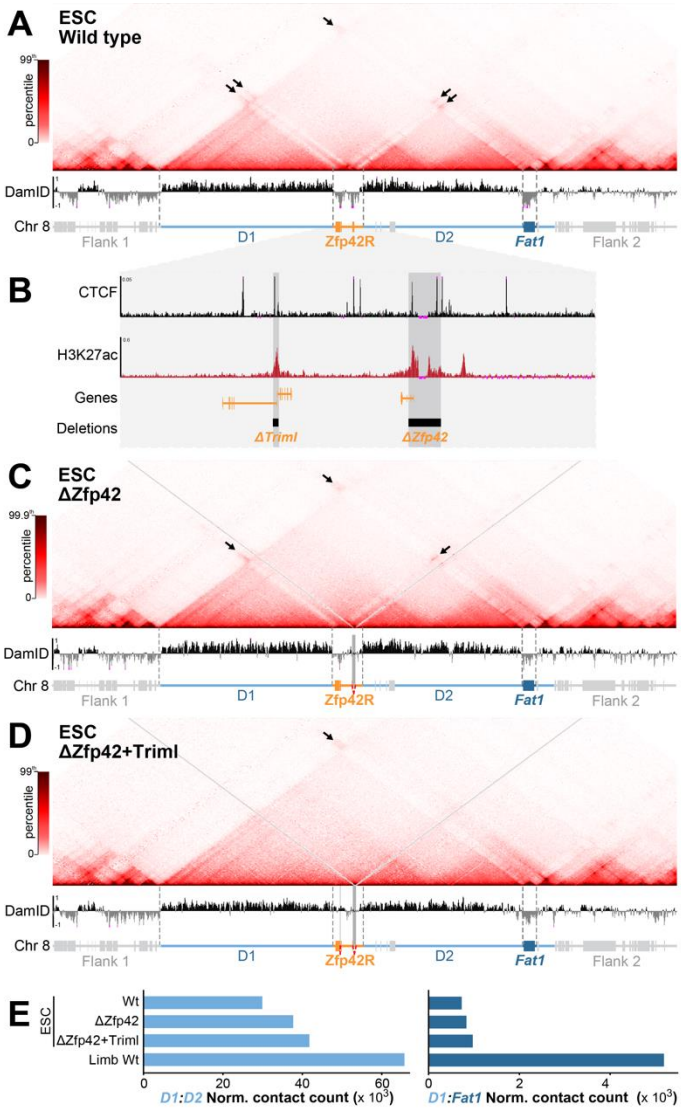


Figure 23 Chromatin restructuring is driven by the epigenetic state of chromatin itself
A. cHi-C from wt ESCs with H3K27ac ChIP-seq and DamID-seq below. **B.** Zoom of *Zfp42R* with highlighted H3K27ac regions deleted in ΔZfp42 and ΔZfp42+Triml. **C-D.** cHi- from ΔZfp42 (C) and ΔZfp42+Triml (D) ESCs with corresponding DamID-seq below. **E.** Quantification of cHi-C interactions for D1:D2 (left) and D1:*Fat1* (right) in indicated samples. *(Analysis of HiC interactions was done by R.S.)*

4. *Zfp42´*s inactivity within *Fat1´*s regulatory active TAD in Limbs

This fourth section focuses on the limb situation. Here, *Fat1* and *Zfp42* share an intact TAD together with active *Fat1*-enhancers. Unlike *Fat1*, *Zfp42* surprisingly does not respond to these enhancers and remains inactive. Hence, we asked how this is possible

and considered different potential mechanisms that evolution could have employed to silence *Zfp42*. Specifically, *Zfp42* is positioned at the NE in a LAD which is generally believed to create a repressive environment for genes (van Steensel and Belmont 2017; Guelen et al. 2008; Reddy et al. 2008; Finlan et al. 2008). Therefore, we tested with sensor experiments and genome engineering if this repressive environment influences *Zfp42´*s regulation and found that it neither directly nor indirectly does so. Next, we addressed the compatibility of ancient *Fat1*-enhancers with the newer *Zfp42* promoter. In Drosophila, strict enhancer-promoter specificity has been demonstrated, and similar principles have also been suggested for mammals (Li and Noll 1994; Merli et al. 1996; Kwon et al. 2009; Martinez-Ara et al. 2022). However, with promoter-exchange experiments, we demonstrate that strict incompatibility with *Fat1*-enhancers is not responsible for *Zfp42´*s inactivity. Last, *Zfp42* activity was reported to anticorrelate with the DNA methylation of its promoter (Borgel et al. 2010; Singer et al. 2014). Therefore, we eliminated DNA methylation and found that *Zfp42* becomes active and responsive to *Fat1*-enhancers. Thus, *Zfp42´*s promoter is repressed through highly context-dependent DNA methylation in limbs.

4.1 Progressive LAD removal causes Zfp42´s NE-release but not its activation

Our previous comprehensive structural mapping in limbs revealed that *Zfp42,* and not *Fat1,* is embedded in a large LAD while *Fat1* is not (Fig. 18). As LADs are thought to be repressive features, we predicted *Zfp42´*s NE-proximity facilitates its silencing (van Steensel and Belmont 2017; Guelen et al. 2008; Reddy et al. 2008; Finlan et al. 2008). Thus, we hypothesized that step-wise LAD removal will cause *Zfp42´*s NE release and consequently its ectopic activation. To test this, we examined structural and functional consequences in mutant limbs lacking NE-attached regions (ΔD1, ΔD2 or ΔD1+2, generated in 1.2) by cHiC, DamID-seq, mSBS modeling, imaging and RNA-seq (Fig15).

Partial removal of the LAD environment did not significantly alter the locus structure. Specifically, DamID-seq of single D1 or D2 deletions in E11.5 limbs showed no obvious changes to NE-attachment, including no spreading or weakening of LAD signals at the ectopically created nonLAD-LAD borders (Fig. 24 A-D). Likewise, ΔD1 and ΔD2 c-HiC closely resembled the wildtype interactions including internal intra-TAD structures despite the reduced TAD size (Fig. 24 B-C). Combined, *Zfp42R* and *Fat1* continue to share the TAD and remain in their LAD and non-LAD positions, respectively. Excitingly, further reducing the LAD environment in ΔD1+2 mutants leaves a small fraction of the TAD in c-HiC, which indeed detaches from the NE by FISH (Fig. 24 D and E). Specifically, imaging shows that *Zfp42R* significantly increases its NE-distance from 300 nm in wildtype limbs

to 600 nm in ΔD1+2 limbs, thereby reaching a distance comparable to wildtype ESCs (mean 550 nm) (Fig. 24 E). Along with its NE-release, *Zfp42R* significantly gains contacts with active flanking chromatin Fl1 in ΔD1+2 (Fig. 24 F). Importantly, the Zfp42-Fl1 distance is unchanged in ΔD1 where no such alterations were detected. As such, increased

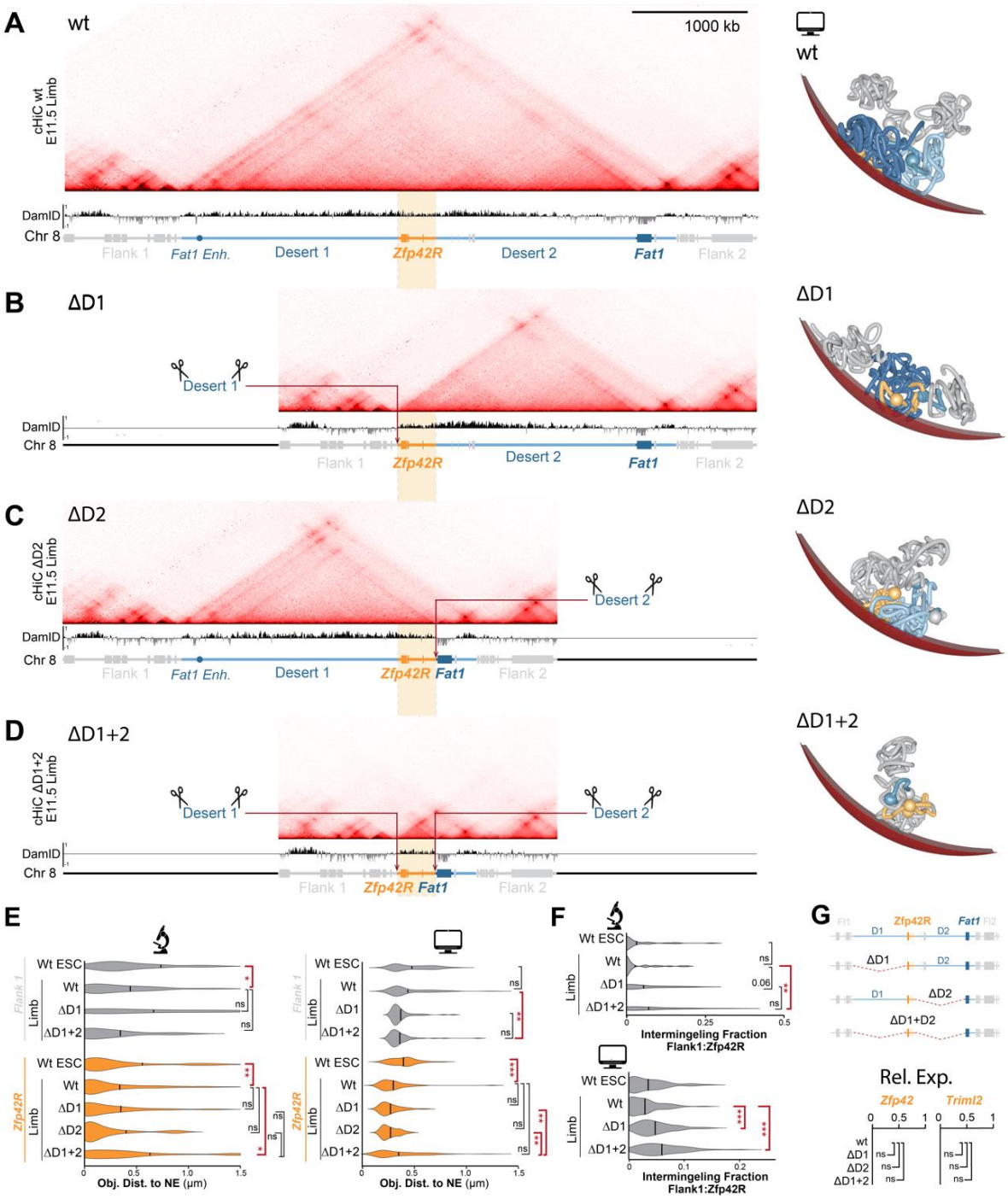


Figure 24 *Zfp42´*s NE association is uncoupled from its transcription in limbs
A-D. cHiC with corresponding DamID-seq, genomic tracks (left) and representative mSBS-simulated structure (right) from wt (A), ΔD1 (B), ΔD2 (C) and ΔΔD1+2 (D) E11.5 limbs. **E-F.** Measurements from FISH microscopy and polymer simulations. (*in silico simulations were done by A.M.C., S.B. and A.E).* Quantification of NE-distance for Flank1 and *Zfp42R* (G). Quantification of *Zfp42R* intermingling fraction with Flank1 (H). *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05 from Welch's t-test comparisons between indicated samples. FISH n=28-138 alleles of at least two biological replicates. Modelling n= 71-106 simulated structures. **G.** Schematic of CRISPR mutants

progressively eliminating NE-association around *Zfp42R* (top) and RNA-seq for *Zfp42* and *Triml* in ΔD1, ΔD2, ΔD1+2 E11.5 limbs (bottom). n = 2-4 biological replicates per condition. Differential gene expression was determined by DEseq (Love et al. 2014): ***p < 0.001, ns = non-significant hen p>0.05

Fl1 is a product of the simultaneous elimination of D1+D2 together, rather than a distant effect of reduced linear *Zfp42R* and Fl1 separation (Fig. 24 F).

Interestingly, we noted, that in contrast to imaging, DamID (and DamID-based mSBS modelling) interestingly does not detect the NE release in ΔD1+2 mutants, as its signal is averaged over a time period of 48 hours (Fig. 24 E). As such, DamID shows that over time *Zfp42R* is still able to contact the NE in limbs, unlike in ESCs, consequently suggesting divergent NE-association dynamics in mutant limbs.

Collectively, this shows that the D1 and D2 LADs surrounding *Zfp42R* act redundantly to maintain its NE attachment. However, the loss of both LADs drives *Zfp42R´*s NE release and consequently its exposure to active chromatin. Therefore, we finally examined if *Zfp42R´*s NE release in fact triggers its activation. Nevertheless, RNA-seq revealed that neither partial nor complete LAD removal caused *Zfp42R genes´* ectopic activation (Fig. 24 G). Thus, *Zfp42R´s* NE release in ΔD1+2 limbs has no functional impact on its activity demonstrating that NE association can be uncoupled from transcription at the *Fat1/Zfp42* locus.

4.2 NE attachment does not prevent Fat1-enhancer communication with Zfp42

NE-attachment, though not directly driving repression, might still represent a barrier for *Zfp42´*s communication with *Fat1* enhancers. As such, we hypothesized that *Fat1*-enhancers transmit their activities to *Fat1* but are incapable to reach *Zfp42R* genes due to their positioning at the NE in the domain´s heterochromatic core.

To test this, we mapped the availability of *Fat1*-enhancer activities throughout the TAD's LAD and non-LAD regions. For this, we integrated sensors consisting of a *LacZ* gene under the control of the *ß-globin* minimal promoter (Symmons et al. 2014). Critically, this minimal promoter lacks the ability to transcribe alone and instead requires stimulation by enhancer activities available at the integration site. These sensors were inserted at 7 positions within the TAD, including sites 20kb down- and 3kb up-stream of *Zfp42*. Sensor transcriptional activity was then detected by lacZ staining in embryos. Surprisingly, all sensors, regardless of their LAD or non-LAD position, produced a *Fat1*-like LacZ staining detectable in the ear, face and proximal limbs of E12.5 – E14.0 embryos (Fig. 25 A-C). However, minor positional differences were observed where, for instance, only the sensor closest to *Fat1* showed mammary gland-staining (Fig. 25 C). However, importantly a sensor 20 kb downstream of *Zfp42* in the ΔD1+2 mutant shows no activity, indicating the

signal is dependent on *Fat1*-enhancers found in D1 and D2 (Fig. 25 D). Moreover, an additional sensor integrated into the neighboring TAD shows distinct genital staining and, completely lacks *Fat1*-like staining as it is shielded from *Fat1*-enhancers through the intervening TAD boundary (Fig. 25 C). Taken together, *Fat1*-enhancer activities are surprisingly transmitted across heterogenous chromatin landscape, including extensive intervening LADs.

In other words, this indicates that NE association does not inevitably prevent transmission of enhancer activities or the activation of genes at the *Fat1/Zfp42* TAD. Ultimately, this suggests that other mechanisms are in place that render *Zfp42* insensitive to *Fat1* enhancers.

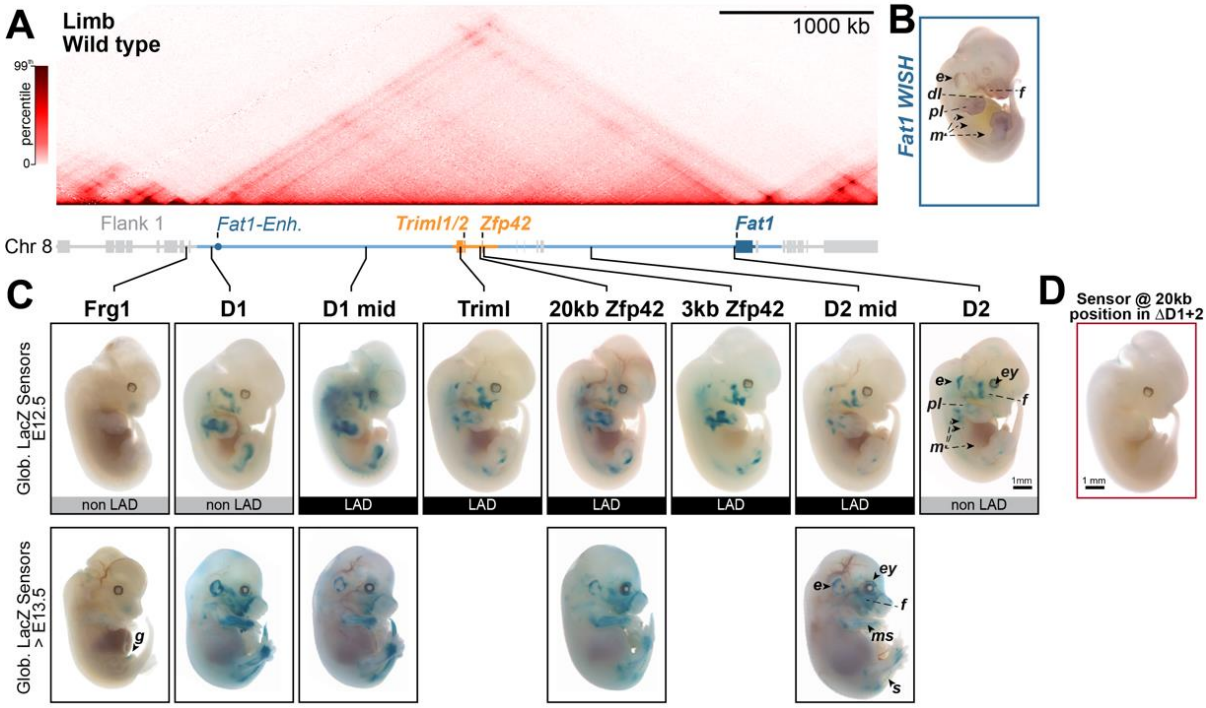


Figure 25 *Fat1*-enhancer activities are transmitted throughout non-LADs and LADs
A. HiC from wt limbs with gene track indicating sensor knock-in positions below. **B.** *Fat1* WISH in E12.5 embryos as reference for LacZ stainings. **C.** Mapping regulatory information with LacZ sensors distributed within and outside of the *Fat1*/*Zfp42* TAD. Representative images of LacZ-stained E12.5 (top) and >E13.5 embryos (bottom) show that all sensors within the TAD recapitulate a *Fat1*-like pattern independent of their position in a non-LAD or LAD. LacZ staining was detected in: ear (e), eye, proximal limb (pl), mammary glands (m), somite (s) and face (f). The sensor outside the TAD (*Frg1*) shows a different LacZ staining present in genitals (g). **D**. A sensor integrated at 20kb downstream of the endogenous *Zfp42* (20kb Zfp42) in the ΔD1+2 mutant lacks all signals. n = 2–10 embryos per position.

4.3 Zfp42´s promoter is functionally compatible with Fat1-enhancers

We next tested if strict *Zfp42* promoter incompatibility with *Fat1*-enhancer allows it to evade the TAD's regulatory content. To do so, we modified one of the LacZ sensors from section 4.2 by exchanging the ß-globin promoter with the *Triml1/2*, *Zfp42* or *Fat1* promoters (Fig.

26 A). Importantly, none of these promoters showed autonomous transcription at the enhancer-free *Rosa26* locus, as indicated by the absence of LacZ staining in all embryos (Fig. 26 C). Therefore, any activities observed when integrating these promoters in the *Fat1/Zfp42* TAD, would reflect their activation through compatible *Fat1* enhancers. Surprisingly, all promoters drove *Fat1*-like lacZ expression in embryos when integrated 20kb downstream of the endogenous *Zfp42* (Fig. 26 D). Specifically, *Triml1/2* and *Zfp42* promoters drove a *Fat1*-like lacZ-staining in the ear, face and proximal limb of E12.5 embryos, demonstrating their capacity to react to *Fat1*-enhancer activities (Fig. 26 D). Thus, to summarize, remarkably all tested promoters of the *Fat1/Zfp42* TAD are in principle compatible with *Fat1*-enhancers.

Though largely displaying similar enhancer responsiveness, differences in staining were observed among tested promoters. For example, the *Fat1* promoter drove additional expression domains in the apical ectodermal ridge and forebrain (Fig. 26 D). As lacZ staining is a qualitative assay, we next directly quantified the promoter´s transcriptional output by qPCR in E12.5 limbs (Fig. 26 E). We measured comparable transcription levels for the *Zfp42* and *Fat1* promoters while the *Triml1/2* and ß-Glob. promoters drove ~30 % and ~70% less transcription (Fig. 26 E). Thus, there is indeed a considerable range of transcriptional responsiveness to the same enhancers. Nevertheless, these variations cannot account for *Zfp42´*s inactivity. Instead, *Zfp42R* promoters must be subjected to a highly context dependent silencing mechanism at their endogenous positions.

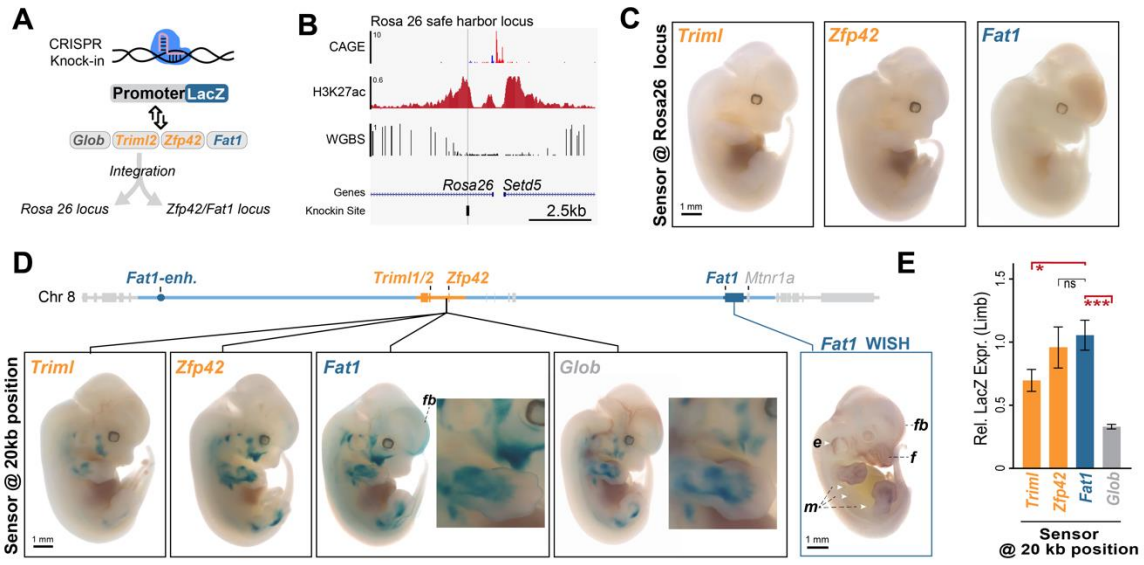


Figure 26 *Zfp42´*s inactivity is not explained by its incompatibility with *Fat1*-enhancers
A. Test strategy for enhancer-promoter compatibility. The globin (Glob) minimal promoter of a sensor is exchanged with either the *Triml2, Zfp42* or *Fat1* promoter. Knock-in of these sensors at the *Rosa26* locus tests the promoter´s autonomous capability to drive transcription while integration at *Fat1/Zfp42* TAD tests the promoter´s compatibility with *Fat1*-enhancers. **B.** Knock-in site at the *Rosa26* safe harbor locus with CAGE, H3K27ac ChIP-seq and WGBS. See table 3 **C-D.**

Representative images of LacZ stained E12.5 embryos with sensors driven by indicated promoters at *Rosa26* (C) or integrated 20 kb downstream of *Zfp42* (D) locus. **D**. All promoter versions drive *Fat1*-like LacZ-activity detectable in the ear (e), mammary glands (m), face (f), forebrain (fb), proximal limb (pl), and apical ectodermal ridge (aer) mostly overlapping with *Fat1* activity detected in WISH (right). n = 4–10 embryos. Scale bar: 1 mm. **E.** qRT-PCR quantification of LacZ transcription in E12.5 limbs from sensors shown in D. n = 3–8 biological replicates per sample. ***p < 0.001, *p < 0.05, and ns = non-significant from Welch's t test comparisons

4.4 Loss of highly context dependent DNA methylation ectopically activates Zfp42

To identify a context-dependent mechanism, we searched for epigenetic modifications at the *Zfp42R* promoters. We re-analyzed publicly available ChIP-seq data but found no evidence of H3K9me3 nor H3K27me3 modifications which would indicate silencing by classical heterochromatin or polycomb, respectively (Fig. 27 A). Considering alternative mechanisms, a report of *Zfp42´*s repression during early development coinciding with increased promoter DNA methylation caught our attention (Borgel et al. 2010). Thus, we performed whole genome bisulfite sequencing (WGBS) in ESCs, where both genes are active, and in limbs where *Fat1* is active without *Zfp42 (WGBS was done by A.M. and S.H.)*. WGBS showed that *Zfp42R* gene promoters are unmethylated in ESCs, and significantly gain methylation in limbs matching their change in activity (Fig. 27 C). Likewise, *Fat1´*s activity in both tissues corresponds with its consistently unmethylated promoter (Fig. 27 C). Hence, we predict DNA methylation at *Zfp42´*s promoter is responsible for its inactivity in limbs. Consequently, we eliminated *de novo* DNA methylation by knocking out the methyltransferases 3A and/or 3B in ESCs. From these cells, we derived E11.5 embryos by tetraploid aggregation, followed by WGBS and RNA-seq to examine knock-out effects.

WGBS revealed ~70% loss of methylation at the *Zfp42* promoter in *Dnmt3b*$^{-/-}$ limbs while methylation at the *Trim1*/*2* promotor was largely unchanged (Fig. 27 B-C). Excitingly, loss of methylation in *Dnmt3b*$^{-/-}$ limbs indeed caused a 6-fold upregulation of *Zfp42*, while *Triml2* expression was unaffected as shown by RNA-seq (Fig. 27 D). In contrast, although methylation status of *Fat1´*s promoter remained unchanged, the loss of DNMT3B caused a 50% reduced *Fat1* expression (Fig. 27 B-D). *Fat1* downregulation might be explained by a new competitive situation in the landscape, where the now active *Zfp42*`s has access to the same enhancers as *Fat1*. Regardless, collectively this shows that DNA methylation in fact contributes to *Zfp42´s* inactivity and unresponsiveness to *Fat1*-enhancers at its endogenous position.

Since the *Zfp42´*s promoter drove *Fat1*-like activity when repositioned by only 20 kb, as observed in the promoter-exchange experiments, we consequently predicted the repositioned promoter to be unmethylated. To test this, we performed bisulfite sequencing

cloning to quantify methylation at 12-17 CpGs at the transplanted *Zfp42* promoter (*Bisulfite sequencing cloning was done by J.G.*). As anticipated, tested CpGs at their endogenous positions displayed ~50% methylation but upon repositioning the methylation is strongly reduced to only ~10% (Fig. 27 E). It is important to note that the seemingly low methylation of 50% is the maximum level gained at these sites in wildtype limbs, which were chosen for technical reasons in PCR bisulfite sequencing. Combined, this shows that DNA-methylation drives *Zfp42´*s silencing in a highly context dependent manner.

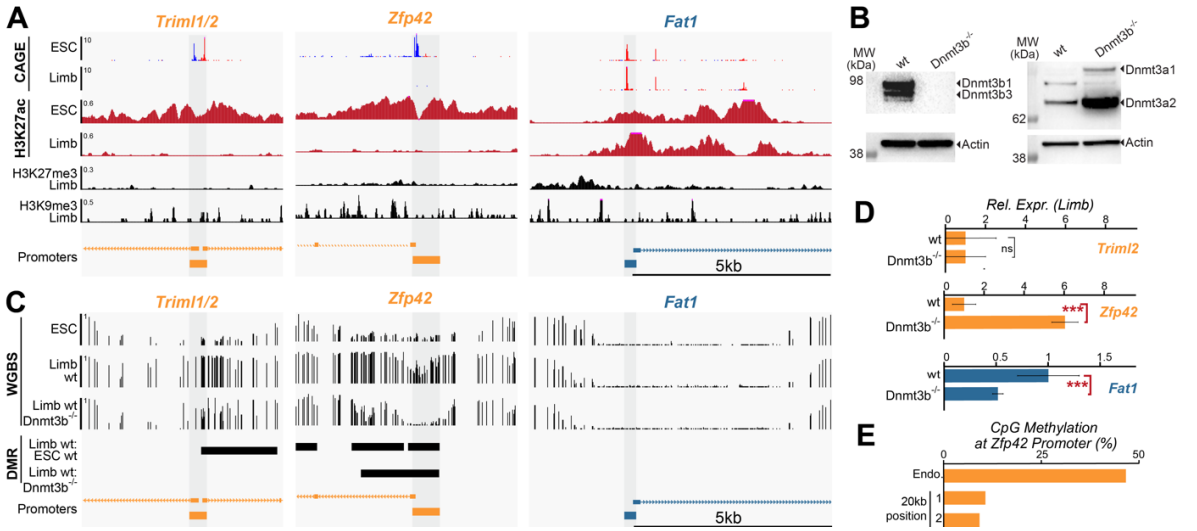


Figure 27 Context-dependent DNA-methylation drives *Zfp42* repression
A. CAGE, H3K27ac, H3K27me3 and H3K9me3 in ESCs and/or Limbs with cloned gene promoters as orange bars below. See table 3. **B.** *Dnmt3b* knock-out was generated by using guides flanking exon 3 and 20 and protein loss was confirming by western blot. DNMT3A increases upon DNMT3B loss. **C**. WGBS in wt ESC, wt limbs, and *Dnmt3b*$^{-/-}$ limb. Called differentially methylated regions (DMR) between wt Limb:ESC and wt:*Dnmt3b*$^{-/-}$ Limb are indicated as black bars. Cloned promoters are indicated as orange bars below. *(WGBS were done by A.M. and S.H.)* **D**. Relative expression of *Timl2*, *Zfp42* and *Fat1* by RNA-seq in wt and *Dnmt3b*$^{-/-}$ limbs. n = 2-4 biological replicates per condition. ns = non-significant with p>0.05 and for p < 0.001 from DEseq for differential gene expression (Love et al. 2014). **E**. Quantification of the CpG Methylation at the *Zfp42* promoter. *(Bisulfite sequencing cloning was done by J.G.)*

Nonetheless, *Zfp42´*s ectopic activity in *Dnmt3b*$^{-/-}$ limbs was not particularly strong. The six-fold upregulation in *Dnmt3b*$^{-/-}$ is a comparison to *Zfp42´*s inactive status in wildtype limbs, and as such represents only a tiny fraction of its maximum expression possible in ESCs. Accordingly, *Zfp42´*s promoter methylation was not entirely eliminated in *Dnmt3b*$^{-/-}$ limbs. Consequently, we created a double knock-out mutant of the partially redundant acting DNMT3B and DNMT3A to further reduce methylation. However, we failed to generate E11.5 mutant mouse embryos, as their development was arrested shortly after gastrulation and they died prior to E11.5 as described previously (Okano et al. 1999).

In summary, these results demonstrate that depending on *Zfp42´*s surrounding sequence-context DNA methylation marks and contributes to its repression. Thus, DNA

methylation is one mechanism that evolution employed to ensure *Zfp42´*s repression thereby enabling independent regulation of *Zfp42* and *Fat1* in the same TAD.

4.5 Zfp42´s promoter methylation ensures its unresponsiveness towards Fat1-enhancers

Finally, we tested if *Zfp42´*s ectopic activity in *Dnmt3b*$^{-/-}$ limbs is truly a product of *Zfp42* communicating with *Fat1*-enhancers. Indeed, it is also possible that the ectopic activity derives from enhancer-independent ”leaky” transcription due to the methylation loss. Therefore, we predicted that *Zfp42´*s ectopic expression as product of its communication with *Fat1*-enhancers i) could be visible in cHiC ii) should occur in an *Fat1*-like pattern, and iii) should be abolished when surrounding *Fat1*-enhancers are eliminated.

To test our first prediction, we performed cHiC in *Dnmt3b*$^{-/-}$ mutatns, where we find ectopic *Zfp42* activation is not associated with novel limb enhancer contacts or altered TAD structure. Testing the second prediction, we visualized endogenous *Zfp42* expression in wildtype and *Dnmt3b*-/- embryos by replacing its coding sequence with *LacZ* to generate embryos carrying LacZ-tagged *Zfp42.* As anticipated, no LacZ staining was detected in embryos with a wildtype background (Fig. 28 B). However, ectopic *Fat1*-like activity from *Zfp42* was observed in *Dnmt3b*$^{-/-}$ embryos (Fig. 28 B). Specifically, although staining is less intense than observed in promoter-reposition experiments, it is detected in mesenchymal limb cells and branchial arches overlapping with the *Fat1-enhancer* activity pattern detected by sensors (Fig. 25). Thus, *Zfp42* ectopic activity follows a *Fat1* activity pattern as result of its interaction with *Fat1*-enhancers upon reduced methylation in *Dnmt3b*$^{-/-}$ embryos.

To confirm this, we eliminated *Fat1*-enhancers in *Dnmt3b*$^{-/-}$ embryos, predicting this should eliminate ectopic *Zfp42* activity. Specifically we recreated the DNMT3B knockout in the ΔD1+2 mutant to create *Dnmt3b*$^{-/-}$:ΔD1+2 embryos that lack DNA methylation and all *Fat1*-enhancers. In fact, removing *Fat1*-enhancers in *Dnmt3b*$^{-/-}$:ΔD1+2 E11.5 limbs eliminated *Zfp42* mis-expression (14C). Thus, DNA methylation suppressive *Zfp42* enhancer responsiveness specifically, and not autonomous promoter activity.

In summary, evolution enables *Zfp42* to evade the regulatory activity of its surroundings though highly context-dependent DNMT3B-mediated DNA methylation.

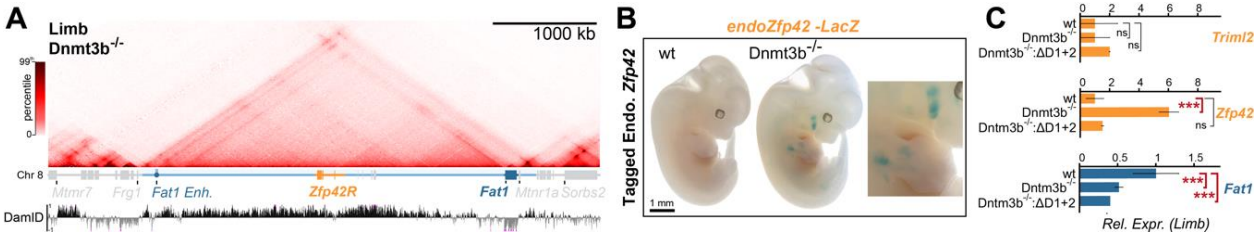


Figure 28 *Zfp42* is responsive to *Fat1*-enhancers in *Dnmt3b*$^{-/-}$ embryos
A. cHiC in *Dnmt3b*$^{-/-}$ limb with corresponding DamID-seq below. **B.** Representative images of E12.5 lacZ stained embryos carrying *Zfp42-LacZ* in wt and *Dnmt3b*$^{-/-}$ background. Scale bar: 1 mm **C.** Relative expression of *Triml2*, *Zfp42* and *Fat1* in wt, *Dnmt3b*$^{-/-}$ and *Dnmt3b-/-*:ΔD1+2 E11.5 mutant limbs by RNA-seq. n = 2-4 biological replicates per condition. ns = non-significant with p>0.05 and *** for p < 0.001 from DEseq for differential gene expression (Love et al. 2014).

DISCUSSION

Genomic rearrangements during evolution frequently modify gene regulation and thus are considered a central mechanism for acquiring novel traits. Yet, in humans, such rearrangements can disrupt the interconnected mechanisms controlling gene activities and cause diseases. Therefore, a longstanding question in biology is what mechanisms allow genomes to “safely” incorporate genomic changes and novel functions during evolution. Here, we addressed this question, by studying how the potentially harmful emergence of *Zfp42R* genes into *Fat1´*s ancient landscape occurred without perturbing *Fat1´*s functions. After reconstructing the locus´s evolutionary history, we searched for mechanisms that adapted this landscape to integrate the novel genes. For this, we functionally examined how TADs, LADs, promoter context, and DNA methylation interact *in vivo.* This led to the identification of two mechanisms, *3D-restructuring* and *context-dependent promoter repression,* acting in different tissues to achieve independent *Zfp42R* genes and *Fat1* regulation. As such, we demonstrate that even a single regulatory landscape can harbor enormous regulatory complexity to control multiple independently-regulated genes in parallel.

The following section discusses the significance of the genome´s regulatory flexibility for disease and evolution and how this work expands our understanding of gene regulation. Firstly, we showed that presumably stable TADs can be massively restructured independently of CTCF/cohesin (Bonev et al. 2017). Furthermore, we presented the novel finding that supposedly repressive LADs are not inaccessible environments for enhancer-promoter communication nor necessary for gene repression (Finlan et al. 2008; Reddy et al. 2008). Moreover, our enhancer-promoter-compatibility results will be discussed in context of the ongoing controversy about its relevance for mammalian gene regulation (Martinez-Ara et al. 2022; Bergman et al. 2022). Finally, new insights into DNA methylation acting in a highly-context dependent manner to limit promoter responsiveness will be discussed.

1. TADs are a fertile target to modify gene expression

Two assumptions regarding TADs are prominent in the field: namely (1) they are stable between cell types and species and (2) act as co-regulation units for genes. However, from an evolutionary point of view, this seems highly impracticable as it would force all genes to have the same expression pattern, even if their biological functions must avoid this. Likewise, emerging novel genes would always need to adopt the pre-existing expression program of their surroundings. With this work, we challenged these assumptions with our

results showing that even largely evolutionary stable TADs are structurally flexible and host multiple differentially regulated genes. In other words, TAD regulatory landscapes represent a fertile substrate for evolutionary innovations.

2. TADs are flexible entities

Due to TAD´s importance for enhancer-promoter communication, they are generally believed to be stable across cell types and in evolution to sustain gene regulation. Indeed, several groups reported TADs to be stable when analyzing a subset of TADs, or providing evidence though correlations with functional genomic data (Dixon et al. 2012; Harmston et al. 2017; Renschler et al. 2019; Huang et al. 2017). Supporting this, a recent deep learning approach reconstructed the genome organization of archaic hominins based on DNA sequence, and found it to be largely constrained when compared to modern humans (McArthur et al. 2022). Accordingly, TADs are reported to be more likely reshuffled as an entire functional unit rather than being broken during evolution, suggestive of purifying selection (Krefting et al. 2018; Vietri Rudan et al. 2015; Liao et al. 2021; Lazar et al. 2018). Strong evidence for this stability, exists especially for TADs harboring genes critical for developmental processes (Woltering et al. 2014; Gómez-Marín et al. 2015; Muro et al. 2019). Here, we likewise found the *Fat1/Zfp42* TAD structurally largely maintained and conserved from mouse back to zebrafish (Fig. 14). Finally, the observation that disrupting TADs can drive gene mis-expression resulting in developmental disorders and cancer further suggests TADs as a critical component to gene regulation (Anania and Lupiáñez 2020).

Nevertheless, though largely conserved, we found the *Fat1*/*Zfp42* TAD also experienced changes during evolution (Fig. 14-15). For example, prior to *Mtnr1a´s* incorporation into the *Fat1/Zfp42* TAD in placental mammals, it was separated in its own TAD in xenopus and chicken (Fig. 14). In agreement, several groups reported that TADs are not conserved and do change due to genomic rearrangements or boundary gain/loss (Acemel et al. 2016; Eres et al. 2019; Luo et al. 2021; Liao et al. 2021). As such, an evolutionary study between two drosophila species which diverged from a common ancestor ~15 million years ago displays only 25% orthologous TADs (Torosin et al. 2020). Moreover, a single TAD at the *HoxD* locus in amphioxus got separated into two TADs in vertebrates (Acemel et al. 2016). This evolutionary structural separation enables *Hox* genes to switch between enhancers of both TADs dependent on the tissue and developmental stages thereby expanding their patterning function to novel organs (Acemel et al. 2016; Andrey et al. 2013). Likewise, TAD-rearrangements in humans relative to macaque monkeys are associated with altered gene expressions, which seems related to

our greater neurological complexity (Luo et al. 2021). Collectively, these studies suggest that TADs are changeable and that this might provide selective advantages for novel evolutionary traits.

As TADs have been shown to change in evolution, they are also reconfigured during development. For example, a tissue-specific chromatin configuration at the *Pitx1* locus allows the physical accessibility of the *Pen*-enhancer to *Pitx1* only in hindlimbs (Kragesteen et al. 2018). Ectopically creating this enhancer access for *Pitx1* in forelimbs results in pathogenic *Pitx1* expression and a partial-arm-to-leg transformation (Liebenberg syndrome) (Kragesteen et al. 2018). Moreover, a comparison of 37 human cell types revealed that less than 10% of all TAD boundaries are shared among them and 18-34%, dependent on the HiC-resolution used, are entirely cell type-specific (McArthur and Capra 2021). In line with other genome-wide reports, this suggest that restructuring events are prevalent equipping TADs with a considerably flexibility to adapt dependent on the cellular context (Bonev et al. 2017; Isoda et al. 2017; Winick-Ng et al. 2021a). Likewise, we demonstrated the massive reconfiguration of the *Fat1/Zfp42* landscape which enabled a new assignment of enhancer-promoter contacts specifically in ESCs (Fig. 18 and 21). Thus, collectively, TADs should not be viewed as rigid blocks that can be assembled like Lego. Rather, they are dynamic structures changeable in evolution and across tissues.

The debate on whether or not TADs are stable likely arises from multiple sources. First, accurately defining a TAD is neither trivial nor standardized. Diverse algorithms developed for TAD calling result in substantially varying average TAD sizes and total numbers of TADs which can differ almost by an order of magnitude (Forcato et al. 2017; Zufferey et al. 2018) (reviewed in (de Wit 2020)). Additionally, variability between studies arises from diverse experimental protocols used and the algorithm´s dependencies on sequencing depth & coverage, and available genome assemblies (de Wit 2020). Second, aside from inconsistent approaches challenging data interpretation, the field needs to define what evidence is required for a certain degree of TAD conservation genome-wide (Eres and Gilad 2021). Of note, one of the first and likely most cited studies for TAD conservation is Dixon at al. reporting highly conserved TADs based on 54% of human boundaries being shared with mouse, and 76% vice versa (Dixon et al. 2012). Moreover, quantifying the shared overlap of all boundaries (rather than unilateral comparisons) results in 31% conservation (Dixon et al. 2012; Eres and Gilad 2021). Defining the scope of “TAD conservation” will be challenging but should certainly consider that TADs harboring developmental genes display a different behavior compared to others (Muro et al. 2019; Wu et al. 2021).

Collectively, there is yet no satisfying answer to what extent TADs are generally conserved at a genome-wide scale. Nevertheless, some TADs are conserved entities to sustain expression programs while others are frequently subjected to modification when necessary.

3. TADs are not units of co-regulation

Though frequently viewed as units of coregulation, we demonstrated that the *Fat1/Zfp42* TAD can host multiple divergently regulated genes. This contrasted past findings suggesting that TADs allow enhancer activities to be transmitted to all genes in the same domain. Specifically, sensor insertions at multiple developmental loci, including here at the *Fat1/Zfp42* locus, show enhancer activities are shared throughout TADs (Symmons et al. 2014; Despang et al. 2019; Marinic et al. 2013) (Fig. 25). Consistently, a recent genome-wide study analyzed genome topology across hundreds of individuals and reported that genetic variations of CREs affected only genes sharing the same TAD (Delaneau et al. 2019). Finally, multiple groups demonstrated coordinated gene activities within TADs (Nora et al. 2012; Le Dily et al. 2014; Neems et al. 2016; Rennie et al. 2018). This holds especially true, when examining TADs which contain genes with greater functional similarity. For example, a study in breast cancer cells described coordinated gene activation/repression upon hormone treatment (Le Dily et al. 2014). Collectively, this gave rise to the assumption that TADs act as units for gene-coregulation (Dixon et al. 2016).

However, considering a general, genome-wide gene co-regulation within TADs conveys a different view. Specifically, in genome-wide analysis, not only a correlation between gene co-expression and TAD co-occupancy was found, but also that multiple genes sharing a TAD often display non-coordinated activities (Ringel et al. 2022; Long et al. 2022). Accordingly, this was also observed at individual loci including the *Hox* and *ß-globin* TADs supporting sequential, rather than simultaneous gene activations (Andrey et al. 2013; Palstra et al. 2003; Soshnikova and Duboule 2009; Huang et al. 2017). Moreover, genomic rearrangements creating new fused TADs and novel enhancer-promoter combinations do not always cause mis-expression and disease (Laugsch et al. 2019; Ghavi-Helm 2020). Thus, this collectively suggests that the primary function of TADs, just as at *Fat1/Zfp42* TAD, is not to provide co-regulation. Rather, TADs reflect one mechanism to help restrict enhancer activities to target genes but it is clear other additional mechanisms must operate to achieve divergent enhancer-promoter communications within them.

4. Chromatin activity can drive chromatin structure

We show that a largely evolutionary stable TAD can be massively restructured by chromatin activity rather than cohesin and CTCF. Previous reports showed that new enhancer-promoter assignments are achieved by TFs, CTCF and cohesin (Bonev et al. 2017; Deng et al. 2014; Isoda et al. 2017). However, despite specific enrichment of CTCF and cohesin at *Zfp42R* in ESC, we show that they are dispensable for the *Fat1*/*Zfp42*-locus partitioning (Fig. 22) (Bonev et al. 2017). Instead, enhancer availability is restricted through the spatial separation of chromatin (compartmentalization) according to its activity into smaller domains (Fig. 21). Thus, the locus structure is defined by chromatin´s underlying epigenetic signature and eliminating the active epigenetic marks at *Zfp42R* leads to loss of separation and partial re-assembly of the TAD (Fig. 23). Excitingly, similar deletion experiments at the *Dppa2*/*4* locus also reported a collapse of a compartment-separation (Sima et al. 2019). Thus, chromatin activity may possess a wider ability to shape its own structure which potentially is promoted by a feedback-loop where the structure is reinforced through the underlying epigenetic state. Moreover, more generally this suggests that compartments operate at TAD levels to functionally control genes and thus, they should not be viewed as a simple, higher order collective of TADs (Hu et al. 2018).

TADs and compartments act antagonistically and both originate from independent forces (Schwarzer et al. 2017; Mirny et al. 2019). While loop extrusion is well studied, the factors driving "compartmentalization" are difficult to examine due to the intrinsic relationship between chromatin state and structure (Hildebrand and Dekker 2020). Nevertheless, a recent study demonstrated that BRD2, which binds acetylated histones and interacts with CTCF, promotes the compartmentalization of active chromatin (Xie et al. 2022). Moreover, other BET family members display similar functions to BRD2. As such, the yeast BET homolog Bdf1 prevents heterochromatin spreading by creating a boundary between eu- and heterochromatin (Ladurner et al. 2003). Likewise, the Drosophila BET homoglog Fs(1)h is also found at insulator sites (Kellner et al. 2013). Collectively, BET family members might be actively involved in the formation of compartments across evolution. As such, compartmentalization at some loci could be the product of factors actively counteracting and overwriting loop extrusion rather than reflecting the physical property of chromatin to separate due to activity. In this view, changing the balance of forces driving chromatin structure might have occurred at the *Fat1/Zfp42* in ESCs as adaptation to *Zfp42´*s activity. Hence, it would be exciting to investigated if the observed locus separation in mouse ESCs likewise exists in opossum stem cells where *Zfp42* is not present.

5. NE association neither prevents enhancer-promoter communication nor is it required for gene repression

We found that LADs neither prevent the communication of genes with enhancers nor do they necessitate gene silencing. This contradicts prevailing models where LADs are silencing environments (van Steensel and Belmont 2017). Specifically, early tethering experiments showed that gene activities are reduced when artificially attached to the NE (Reddy et al. 2008; Finlan et al. 2008). Likewise, in a recent high-throughput approach, promoters were transplanted into LAD and non-LAD environments which demonstrated that LADs intrinsically influence gene transcription *in vitro* (Leemans et al. 2019). Finally, the loss of proteins tethering LADs triggers the NE-release of genes and their activation during *in vitro* differentiation (Robson et al. 2016; Poleshko et al. 2017). Consequently, the current dominant thinking is that LADs act as a repressive regulatory layer for gene control.

Yet, we find LADs are dispensable for endogenous gene regulation, at least at *Zfp42R*. Specifically, we find that *Zfp42R´*s NE-association is redundantly maintained by D1 and D2 LADs showing a cooperative behavior, as previously reported (Fig. 24 A-C and E) (Kind et al. 2015). However, elimination of D1 and D2 LADs together caused NE-release of *Zfp42R* genes, which then intermingled with flanking active chromatin (Fig. 24 D-E). Despite this, NE-release did not cause *Zfp42* activation in embryos (Fig. 24 G). Significantly, we noted that, in contrast to imaging, DamID (and DamID-based mSBS modelling) did not detect the NE-release in ΔΔD1+2 mutants (Fig. 24 D). However, as DamID signal is averaged over 48 hours, this might indicate that *Zfp42R* can still transiently contact the NE in ΔΔD1+2 mutant limbs. As such, we cannot rule out that this continued limited attachment may be sufficient to sustain silencing.

Despite this, two additional *in vivo* experiments demonstrate that LADs are not intrinsically linked to gene silencing. Specifically, we found that promoters embedded in LADs can still be activated by *Fat1*-enhancers, indicating NE-attachment does not present an impenetrable barrier (Fig. 25). Hence, LADs are likely easily locally restructured to *Fat1*-enhancers to emerge from heterochromatin and readily activate lacZ reporter genes despite large intervening NE-attachment. Supporting this, artificially activating genes within LADs drives local chromatin release from the NE *in vitro* (Brueckner et al. 2020; Therizols et al. 2014). Combined, this demonstrates that LADs are in principle permissive for gene regulation, and not entirely unfavorable to the emergence of new genes in evolution.

6. Enhancer-promoter compatibility does not account for strong differences in mammalian gene expression

Although existing in Drosophila, we found no strict enhancer-promoter compatibility at the *Fat1/Zfp42* locus in mice. Currently, there is an on-going controversy regarding the relevance of enhancer-promoter specificity for mammalian gene regulation (van Arensbergen et al. 2014; Galouzis and Furlong 2022). Certainly, such a mechanism would present an easy solution to allow independent gene activities within the same TAD. However, here we showed that *Fat1´*s and *Zfp42R* gene´s promoters can be activated by *Fat1*-enhancers, agreeing with previous mammalian studies where ectopic integrated promoters were compatible with surrounding enhancers (Fig. 26) (Marinic et al. 2013; Shima et al. 2016; Symmons et al. 2014). Likewise, high-throughput STARR-seq experiments in human cells demonstrated that nearly all promoters can be activated by all enhancers to a similar extent (Bergman et al. 2021). Collectively, this supports the lack of strict enhancer-promoter compatibility in mammals in agreement with another group (Hong and Cohen 2022).

Nevertheless, we and others have also observed differences in promoter's responsiveness to enhancers (Martinez-Ara et al. 2022; Pachano et al. 2022). Specifically, quantifying the transcriptional responsiveness to *Fat1*-enhancers revealed ~3-fold differences between the activity levels of promoters. Moreover, *Fat1´*s promoter drove additional novel expression patterns in embryos, suggesting some degree of specificity exists (Fig. 26 E). Previously, support for selective compatibility mostly came from studies in Drosophila where some promoters are only compatible with certain enhancers (Butler and Kadonaga 2001; Li and Noll 1994; Merli et al. 1996). Moreover, Drosophila developmental and housekeeping promoters generally display preferences towards different classes of enhancer (Arnold et al. 2017; Zabidi et al. 2015). Similarly, in mammals, a recent high-throughput combinatorial reporter assay reported strong promoter specificities with more than half of the active enhancers displaying significant promoter selectivity (Martinez-Ara et al. 2021). Thus, though not as well understood as in Drosophila, variable promoter responsiveness exists in mammals and may contribute to gene regulation.

The mechanisms driving this specificity remain unknown. However, such variation could be driven by the degree of promoter accessibility, a core-promoter´s sequence composition, the recruitment of different transcriptional co-factors and the mechanism of enhancer transcription initiation (Haberle et al. 2019; Serebreni et al. 2022; Galouzis and Furlong 2022). Alternatively, although not accounting here as all promoters were tested at the same position, genomic distance between enhancers and promoters within TADs can

also influence a gene´s transcriptional output (Zuin et al. 2021). As an example, a gene that is five times further away and two times less sensitive to an enhancer could be unresponsive to enhancers compared to another more proximal promoter. Therefore, at other loci, observed differences in promoter responsiveness might be a product of distance rather than specificity.

Moreover, it should be noted that what is true for promoters is also true for enhancers: *they are not all the same* (Arnold et al. 2017). In this study, all examined promoters were tested in their native environment which introduced additional complexity to answer why they vary in their transcriptional output. In contrast to *in vitro* assays where isolated enhancer-promoter pairs are investigated, here tested promoters can communicate with an entire *Fat1*-enhancer cluster. As such, the varying degree of expression between the promoters could be also explained by the enhancer's selectivity towards promoters, and thus their different additive or synergistic contribution to the overall transitional output. In other words, different responsiveness of the tested promoters might not be apparent at other loci and thus the source of difference should not be attributed to the promoter alone.

Combined, although the molecular basis of different promoter sensitivities remains to be determined, a strict enhancer-promoter compatibility cannot explain *Zfp42´*s inactivity in limbs. Nonetheless, sensitivity as a mechanism allows the genome a greater regulatory complexity and as such might be an evolutionary tool to regulate multiple expression programs in parallel without conflicts.

7. Context dependent DNA-Methylation blocks enhancer-promoter communication

We found that context-dependent promoter repression through DNA methylation prevents enhancer-promoter communication. DNA methylation is a major epigenetic mark strongly linked to gene silencing, and was previously reported to coincide with the *Zfp42* inactivity (Borgel et al. 2010). Here we extended the mechanism of how DNA methylation controls genes: by preventing inappropriate enhancer communication rather than repressing a promoter's intrinsic transcriptional activity. As such, we showed that eliminating DNMT3B-driven methylation desensitizes the *Zfp42´*s promoter to *Fat1*-enhancers, thereby driving its ectopic activity in embryos (Fig. 27 C-D). Nonetheless, this ectopic activation at its endogenous position was mild, which might be explained by compensatory or additional unknown silencing mechanisms. Alternatively, our inability to further reduce methylation levels at *Zfp42´*s endogenous site may also be to blame (Fig. 27 C).

Excitingly, our data suggest that a CpG-poor promoter sequence alone is insufficient to be targeted for methylation, and instead requires flanking sequence context

(Deaton and Bird 2011; Long et al. 2016a; Pachano et al. 2022). As such, when only slightly repositioned, *Zfp42´*s promoter drives robust transcription in inappropriate tissues with potential pathogenic consequences (Fig. 26 D). Supporting this, in DNMT3B KO embryos, *Fat1* expression is strongly impaired which has been frequently reported with cancer (Wang et al. 2022; Peng et al. 2021). *Fat1´*s downregulation might be due to *Zfp42´*s activity and competing communication with the same enhancers as *Fat1,* although alternative explanations are possible (Fig. 27 D) (Oudelaar et al. 2019). Collectively, we demonstrate that *Zfp42´*s inactivity is explained by DNA methylation-driven repression that is surprisingly (1) highly dependent on sequence context, and (2) renders the promoter unresponsiveness towards enhancers. As such, this repressive mechanism extends the current understanding of DNA methylation beyond transcriptional inactivation of loci though the absence/presence of DNMTs together with distinct chromatin modifications (Ooi et al. 2007; Weinberg et al. 2021; Zhang et al. 2010)

Significantly, employing DNA-methylation as mechanism to prevent enhancer-promoter communication might explain mild effects on gene expression observed genome-wide following DNA methylation loss in early embryos (Grosswendt et al. 2020; Ringel et al. 2022; Yagi et al. 2020). In such a scenario, lost DNA methylation would only rarely have consequences in cases when de-methylated promoters are exposed to inappropriate active enhancers in the same TAD. Interestingly, a recent study took the opposite approach and induced wide-spread methylation at thousands of promoters in cells and similarly found that transcriptional responses to DNA methylation is remarkably contextual (de Mendoza et al. 2022). Thus, collectively, DNA-methylation is another mechanism that supports the simultaneous and independent gene regulation within multi-gene TADs by refining a promoter´s responsiveness.

8. Context matters

In this section, we want to emphasize the importance of studying gene regulation at endogenous loci and, ideally, *in vivo*. Even though correlative genome-wide studies and *in vitro* functional assays have revealed many features influencing gene activities, they cannot yet recapitulate the native environment of endogenous loci. Consequently, even with the knowledge gained in recent decades, we are still unable to predict how genes are endogenously regulated, like at the *Fat1/Zfp42* landscape. Accordingly, likely efforts to answer the question of enhancer-promoter specificity with only episomal *in vitro* assays alone will fail, as they lack the impact of a promoter's native context. Likewise, understanding how enhancers are function cannot be answered with only *in vitro* assays, as they cannot capture the multifaceted, redundant, and complementary functions of

enhancer in native genomic environment (Brosh et al. 2022). Together, this demonstrates the need to combine ectopic *in vitro* assays to systematically test distinct features and *in vivo* dissections at endogenous loci to understand gene regulation fundamentally.

Collectively, our genomes today are a product of evolution. As different genes frequently require divergent and independent expression programs, creates the emergence of new genes into pre-existing landscapes inevitably gene regulatory conflicts. As genomes likely cannot control where a genomic modification occurs within them, it should be expected that over time many different context-dependent solutions were evolved to control a promoter's activity. In this way, the genome acquired an enormous regulatory complexity where a promoter´s response to enhancers is influenced though diverse pre-existing and evolving mechanisms that are interconnected at any given genomic landscape. Consequently, I believe that only few mechanisms are universally transferable to all enhancer-promoter communications in general. Many further studies are thus required to deeply dissect single loci until more accurate predictions for gene expressions are possible.

9. Significance of structural and regulatory flexibility

Here, we showed that an evolutionary landscape can flexibly adapt its structure to regulate different genes independently from one another. Moreover, we describe DNA methylation as a context-dependent mechanism that selectively represses a promoter´s responsiveness to surrounding enhancer activities within the same TAD. This structural and regulatory complexity of a single locus is likely highly pervasive genome-wide, and has significant consequences for understanding evolution and disease (Ringel et al. 2022). As such, this might explain a number of observations.

First, the elimination of cohesin-loop extrusion forming TADs results in only minor gene expression effects (Nora et al. 2017; Rao et al. 2017). Second, not all TAD rearrangements lead to inappropriate enhancer-promoter communications (Ghavi-Helm et al. 2019; Schöpflin et al. 2022; Laugsch et al. 2019). For example, translocations in individuals affecting 34 genes did not cause differential gene activities (Dong et al. 2018). Another, quite extreme example is chromothripsis where chromosomes shatter and are reassembled with massive genomic rearrangement (Ostapińska et al. 2022). Despite these extensive changes, patients with congenital chromothripsis can be largely healthy, exhibiting only minor pathogenic phenotypes (Schöpflin et al. 2022). Though survivor bias may preclude more dramatic lethal effects, it is surprising that such massive chromatin changes have so little effects.

Hence, this demonstrates on the one hand the robustness of our genomes and on the other hand raises the question if all TADs are equally important (López-Rivera et al. 2020; Despang et al. 2019). Unquestionably, TADs are important for a subset of loci. This certainly includes mono-genic developmental TADs which are under evolutionary positive selection and whose disruptions are more likely disease-causing (Muro et al. 2019; Wu et al. 2021). However, at other loci, TADs might benefit gene regulation but are often not absolutely necessary. In these cases, it is likely often that yet to be identified mechanisms exist to control gene regulation, similarly to the ones descripted here. Thus, collectively, in order to fundamentally understand gene regulation and to predict the consequences of genomic rearrangements we need to identify such additional mechanisms.

APPENDIX

1. List of abbreviations

bp	Basepair
BSA	Bovine serum albumin
c-HiC	Capture HiC
CTCF	CCCTC-binding factor
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
cDNA	Complementary DNA
°C	Degrees celsius
del	Deletion
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
d	Depleted
DEPC	Diethylpyrocarbonate
DIG	Digoxygen
DMSO	Dimethylsulfoxide
DMEM	Dulbecco’s modified eagle’s medium
E	Embryonic stage
ESC	Embryonic stem cell
ENCODE	Encyclopedia of DNA elements
E.coli E	Escherichia coli
EtOH	Ethanol
EDTA	Ethylenediaminetetraacetic acid
Fat1	FAT Atypical Cadherin 1
FCS	Fetal calf serum
FL	Forelimb
g	Gram
HL	Hindlimb
h	Hour
kb	Kilobases
KI	KnockIn
KO	KnockOut
LAD	Lamina Associated Domain

LAGeSO	Landesamt für Gesundheit und Soziales
LIF	Leukemia Inhibiting Factor
l	Liter
MPIMG	Max Planck Institute for Molecular Genetics
Mb	Megabase
mRNA	Messenger RNA
MeOH	Methanol
µ	Micro
m	Milli (prefix)
min	Minute
mol	Moles
mESC	Mouse embryonic stem cell
mm	Mus musculus
n	Nano (prefix)
nonLAD	non Lamina Associated Domain
NE	Nuclear Envelop
PFA	Paraformaldehyd
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PIC	Preinitiation complex
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
RT	Room temperature
rpm	Rounds per minute
SSC	Saline Sodium Citrate
sec	Second
sgRNA	Single guide RNA
SV	Structural Variation
taq	Thermus aquatius
3D	Three-dimensional
TAD	Topologically Associating Domain
TF	Transcription factor
TSS	Transcriptional start site
Triml	Tripartite Motif Family Like
U	Units
UCSC	University of California, Santa Cruz

Vol	Volume
w/v	Weight per volume
WISH	whole amount in-situ hybridization
WISH	Whole Mount In-Situ Hybridization
wt	Wildtype
Zfp42	Zinc Finger Protein 42

2. List of figures

3. List of tables

Recombinant DNA	
pLGW-Dam-V5-Lamin B1 (Mm)	Steensel Lab
pLGW-V5-Dam	Steensel Lab
pMD2.G	Bird Lab
psPAX2	Bird Lab
BAC for *Fat1R*	RP23-451E23 CHORI/BACPAC
pX459 pSpCas9(BB)-2A-Puro vector	Addgene
Fat1 promoter 302bp	Genewiz synthesized
Zfp42 promoter 602bp	PCR amplified
Triml promoter 427bp	Genewiz synthesized
Fat1 enhancer	PCR amplified

Table 1 Recombinant DNA

Software and Algorithms	
CRISPR design	https://www.benchling.com
R	https://www.r-project.org
MACS2.0	https://github.com/taoliu/MACS
Bowtie2	Langmead and Salzberg, 2012
Samtools	http://samtools.sourceforge.net
HiCUP v0.5.9	Wingett et al., 2015
Juicer	Durand et al., 2016
HiGlass	Kerpedjiev et al., 2018
UCSC genome browser	https://genome.ucsc.edu
WashU browser	https://epigenomegateway.wustl.edu
HMMt	https://github.com/gui11aume/HMMt
IGV browser	https://www.broadinstitute.org/igv/
Prism	https://www.graphpad.com/scientific-software/prism/
Adobe Photoshop and Illustrator	https://www.adobe.com/

Table 2 Software and algorithms

Data	Publication	Identifier
ChIP-seq for CTCF, Rad21 and H3K9me3 in mouse ESCs and E11.5 limb buds	Kraft et al. 2019	GEO: GSE116794
ChIP-seq for H3K4me1, H3K4me3, H3K27ac and H3K27me3 in E11.5 limb buds	Andrey et al. 2017	GEO: GSE84795
ATAC-seq in mouse ESCs	Bauer et al. 2021	GEO: GSE157448
Fantom5 CAGE Expression datasets	Lizio et al. 2015	https://fantom.gsc.riken.jp/5/data/
Hi-C in 48 hr hpf Zebrafish	Yang et al. 2020	GEO: GSE134055
Hi-C in xenopus brain	Niu et al. 2021	SRA: PRJNA606649
Hi-C in pig embryonic fibroblasts	Li et al. 2020	GEO: GSE153452
ChIP-seq for H3K36me3 in mouse ESCs	Encode	GEO: GSE31039
ChIP-seq for CTCF in chicken	Jerkovic et al. 2017	GEO: GSE86089
scRNA-seq in gastrulating E6.5-8.5 mouse embryos	Pijuan-Sala et al. 2019	ArrayExpress: E-MTAB-6967
scRNA-seq in E9.5-E12.5 mouse embryos	Cao et al. 2019	GEO: GSE119945
scRNA-seq in E9.5-E14.5 mouse placentas	Marsh et al. 2020	GEO: GSE152248

Table 3 Re-analyzed public data

Mutant ESC Line	*Parental ESC Line*	*Type*	*sgRNA/Homology Construct*		*Relavent Target Coordinates (mm10)*	*sgRNA sequence (5'-->3')*
ΔD1	Wt G4 ESC	SV Deletion	57_sgRNA_D1_Cen	58_sgRNA_D1_Tel	chr8:41457521-41457541 + chr8:43108934-43108954	GTGCCATTTAGACATAACGG GATGTCTCACAAGTAACCAG
ΔD2	Wt G4 ESC	SV Deletion	59_sgRNA_D2_Cen	60_sgRNA_D2_Tel	chr8:43402062-43402082 + chr8:44932010-44932030	gatataaggagggtccaagg tcataggcccctccacaccg
ΔD1+2	ΔD2	SV Deletion	57_sgRNA_D1_Cen	58_sgRNA_D1_Tel	chr8:41457521-41457541 + chr8:43108934-43108954	GTGCCATTTAGACATAACGG GATGTCTCACAAGTAACCAG
ΔZfp42R	Wt G4 ESC	SV Deletion	58_sgRNA_D1_Tel	59_sgRNA_D2_Cen	chr8:43108934-43108954 + chr8:43402062-43402082	GATGTCTCACAAGTAACCAG gatataaggagggtccaagg
FRT_Hsp68_Fat1enh	C2 ESC	Enhancer reporter	210_FRT_Fat1_Enh	pCAGGS-flpE-puro (addgene)	Col1a1	N/A
Fat1enh_KO	Wt G4 ESC	Knockout	281_sgRNA_ Fat1enh_cen	282_sgRNA_ Fat1enh_tel	chr8:41591098 -41591117 + chr8:41595146-41595165	AAGACGGACTGATCTTCCAG ACGATTTTAGGTGACCTCTG
D1_Glob_LacZ	Wt G4 ESC	Sensor	164_Hom_D1_BGlob_ LacZ	079_sgRNA_KI_D1	chr8:41547011-41547031	CCTTAAGTAAGTGTACAGCG
D1-mid_Glob_LacZ	Wt G4 ESC	Sensor	139_Hom_D1mid _BGlob_LacZ	167_sgRNA_ KI_D1mid	chr8:42548275-42548294	AAGACCATGGAAACCATGCT
Zfp42Ra_Glob_LacZ	Wt G4 ESC	Sensor	140_Hom_Zfp42Ra_B Glob_LacZ	105_sg_KI_Ins_TrimI	chr8:43159304-43159323	GGGGAGGTGAAAAATCTTGG
Zfp42Rb_Glob_LacZ	Wt G4 ESC	Sensor	137_Hom_Zfp42Rb_B Glob_LacZ	076_sg_KI_ Ins_ZFP42	chr8:43287046-43287066	TGGAAGATACCTATACACCA
Zfp42Rb_TrimIprom_LacZ	Wt G4 ESC	Sensor	218_Hom_Zfp42Rb _TrimIcore_LacZ	076_sg_KI_ Ins_ZFP42	chr8:43287046-43287066	TGGAAGATACCTATACACCA
Zfp42Rb_Zfpmidiprom_LacZ	Wt G4 ESC	Sensor	221_Hom_Zfp42Rb_Zf ppromidi_LacZ	076_sg_KI_ Ins_ZFP42	chr8:43287046-43287066	TGGAAGATACCTATACACCA
Zfp42Rb_Fat1prom_LacZ	Wt G4 ESC	Sensor	222_Hom_Zfp42Rb _Fat1core_LacZ	076_sg_KI_ Ins_ZFP42	chr8:43287046-43287066	TGGAAGATACCTATACACCA
Zfp42Rc_Glob_LacZ	Wt G4 ESC	Sensor	165_Hom_Zfp42Rc _BGlob_LacZ	151_sgRNA_KI_ZFP3	chr8:43309852-43309875	ACTGTTGGGATACAGGTCAA
D2-mid_Glob_LacZ	Wt G4 ESC	Sensor	138_Hom_D2mid_Bglo b _LacZ	78_sgRNA_KI_D2mid	chr8:44001720-44001740	CAAAGCCCTTGAACATACGG
D2_Glob_LacZ	Wt G4 ESC	Sensor	166_Hom_D2_BGlob-LacZ	168_sgRNA_KI_D2	chr8:44934807-44934826	AAGCTTGCAGAGACTGATCG
Frg1_Glob_LacZ	Wt G4 ESC	Sensor	280_Hom_Frg1 _sensor _LacZ	279_sgRNA_KI_ Frg1sensor	chr8:41385464-41385486	CTAGACGCAGAAGACATGGG
R26_Glob_LacZ	Wt G4 ESC	Sensor	152_Hom_Rosa26_ins _BGlob_LacZ	EN479 (addgene)	chr6:113076008 113076027	TGGGCGGGAGTCTTCTGGGC
R26_TrimIprom_LacZ	Wt G4 ESC	Sensor	214_Hom_Rosa26_ins _TrimIcore_LacZ	EN479 (addgene)	chr6:113076008 113076027	TGGGCGGGAGTCTTCTGGGC
R26_Zfpmidiprom_LacZ	Wt G4 ESC	Sensor	217_Hom_Rosa26_ins _ZFPpromidi_LacZ	EN479 (addgene)	chr6:113076008 113076027	TGGGCGGGAGTCTTCTGGGC
R26_Fat1prom_LacZ	Wt G4 ESC	Sensor	226_Hom_Rosa26 _Fat1core_LacZ	EN479 (addgene)	chr6:113076008 113076027	TGGGCGGGAGTCTTCTGGGC
ΔDNMT3B	Wt G4 ESC	Knockout	300_sgRNA_ALM _Dnmt3b_ex3_g1	301_sgRNA_ALM _Dnmt3b_gRNA	chr2:153661525-153661544 + chr2:153661525-153661544	GCTGAATTATACCCGCCCCA GGGAGCATCCTTCGTGTCTG
ΔD1+2+DNMT3B	ΔD1+2	Knockout	300_sgRNA_ALM _Dnmt3b_ex3_g1	301_sgRNA_ALM _Dnmt3b_gRNA	chr2:153661525-153661544 + chr2:153661525-153661544	GCTGAATTATACCCGCCCCA GGGAGCATCCTTCGTGTCTG
ΔZfp42	Wt G4 ESC	Knockout	267_sgRNA_ Zfp42prom _cen	268_sgRNA_ Zfp42prom_tel	chr8:43302254-43302273 + chr8:43332328-43332347	TTAATGTTCAAATTCCACGC CCCTTTGCTTTCACCGCGT
ΔZfp42TrimI	ΔZfp42	Knockout	269_sgRNA_ Trimprom _cen	270_sgRNA _Trimprom_tel	chr8:43176771-43176790 + chr8:43181930-43181949	CCCGAAACTGCTTAACTAAC GCACACTTAATCCAAGTAC
endoZfp42_LacZ	Wt G4 ESC	Gene tagging	287_Hom_Endotag_ Zfp42_lacZ	289_sgRNA_endo _tag_Zfp42	chr8: 43296535-43296554	TTTGGGTTATTATCTAAGGC
ΔDNMT3B_endoZfp42_LacZ	ΔDNMT3B	Gene tagging	287_Hom_Endotag_ Zfp42_lacZ	289_sgRNA_endo _tag_Zfp42	chr8: 43296535-43296554	TTTGGGTTATTATCTAAGGC

Table 4 Generated cell lines
CRISPR/Cas9 generated cell lines: Name of cell lines, parental cell line, type of mutation, sgRNAs and homology constructs with corresponding coordinates.

Mutant ESC Line	*Parental ESC Line*	*Type*	*Genotyping PCR Primer (Primers for deletion (del) /site specific integration (ssi) /specific KI-construct (kic))*	*Genotyping PCR Primer to detect wt signal surrounding sgRNA site*	*Copynumber Genotyping qPCR Primer*
ΔD1	Wt G4 ESC	SV Deletion	*1+4 (del)*	1+2 (wt) ,3+4 (wt)	9+10,11+12,13+14,15+16
ΔD2	Wt G4 ESC	SV Deletion	*5+8 (del)*	5+6 (wt) ,7+8 (wt)	17+18,19+20,21+22
ΔD1+2	ΔD2	SV Deletion	*1+4 (del), 5+8 (del)*	3+4 (wt), 5+6 (wt), 5+6 (wt) ,7+8 (wt)	9+10,11+12,13+14,15+16, 17+18,19+20,21+22
ΔZfp42R	Wt G4 ESC	SV Deletion	*3+6 (del)*	3+4 (wt), 5+6 (wt)	23+24,25+26,27+28
FRT_Hsp68_Fat1enh	C2 ESC	Enhancer reporter	69+70 (ssi), 71+72 (ssi)		29+30
Fat1enh_KO	Wt G4 ESC	Knockout	83+84 (del), 83+85 (del)	86+87 (wt)	
D1_Glob_LacZ	Wt G4 ESC	Sensor	37+38 (ssi), 29+30 (kic)	39+40 (wt)	29+30
D1-mid_Glob_LacZ	Wt G4 ESC	Sensor	33+34 (ssi), 29+30 (kic)	35+36 (wt)	29+30
Zfp42Ra_Glob_LacZ	Wt G4 ESC	Sensor	65+66 (ssi), 29+30 (kic)	67+68 (wt)	29+30
Zfp42Rb_Glob_LacZ	Wt G4 ESC	Sensor	29+30 (kic), 38+41 (kic), 38+42 (ssi), 79+80 (ssi)	31+32 (wt)	29+30
Zfp42Rb_Trimlprom_LacZ	Wt G4 ESC	Sensor	47+48, 29+30(kic), 38+41 (kic), 38+42 (ssi), 79+80(ssi)	31+32 (wt)	29+30
Zfp42Rb_Zfpmidiprom_LacZ	Wt G4 ESC	Sensor	45+46 (ssi), 29+30(kic), 38+41 (kic), 38+42 (ssi), 79+80(ssi)	31+32 (wt)	29+30
Zfp42Rb_Fat1prom_LacZ	Wt G4 ESC	Sensor	29+30 (kic), 38+41 (kic), 38+42 (ssi), 79+80(ssi)	31+32 (wt)	29+30
Zfp42Rc_Glob_LacZ	Wt G4 ESC	Sensor	49+38 (ssi), 29+30 (kic)	51+52 (wt)	29+30
D2-mid_Glob_LacZ	Wt G4 ESC	Sensor	38+51 (ssi), 29+30 (kic)	53+54 (wt)	29+30
D2_Glob_LacZ	Wt G4 ESC	Sensor	34+55 (ssi), 29+30 (kic)	61+62 (wt)	29+30
Frg1_Glob_LacZ	Wt G4 ESC	Sensor	75+76(ssi),77+78(ssi)		29+30
R26_Glob_LacZ	Wt G4 ESC	Sensor	57+58 (ssi), 29+30 (kic)	63+64 (wt)	29+30
R26_Trimlprom_LacZ	Wt G4 ESC	Sensor	57+58 (ssi), 29+30 (kic)	63+64 (wt)	29+30
R26_Zfpmidiprom_LacZ	Wt G4 ESC	Sensor	57+58 (ssi), 45+56 (ssi), 29+30 (kic)	63+64 (wt)	29+30
R26_Fat1prom_LacZ	Wt G4 ESC	Sensor	57+58 (ssi), 29+30 (kic)	63+64 (wt)	29+30
ΔDNMT3B	Wt G4 ESC	Knockout	59+60 (del), 73+74 (del)	N/A	N/A
ΔD1+2+DNMT3B	ΔD1+2	Knockout	59+60 (del), 73+74 (del)	N/A	N/A
ΔZfp42	Wt G4 ESC	Knockout	81+82 (del)	90+91 (wt)	90+91
ΔZfp42Triml	ΔZfp42	Knockout	81+82 (del), 88+89 (del)	90+91 (wt), 92+93 (wt)	90+91, 92+93
endoZfp42_LacZ	Wt G4 ESC	Gene tagging	94+66 (ssi), 94+29 (ssi), 95+43 (ssi)	94+44 (wt)	29+30
ΔDNMT3B_endoZfp42_LacZ	ΔDNMT3B	Gene tagging	94+66 (ssi), 94+29 (ssi), 95+43 (ssi)	94+44 (wt)	29+30

Table 5 Genotyping primer

Assignment of primers for genotyping of cell lines. Name of generated cell lines, parental cell lines, type of mutation and assigned primers used for genotyping. Primer combinations detected a deletion (del), site specific knock-in (sik), site specific integration (ssi) or presence of a wildtype allele (wt). qPCR primers were used for copy number analysis. Sequences to the indicated primer numbers can be found in table 6.

Primer number	Name	Sequence (5'-->3')	Binding site
1	genotyping_1F	TACAGCTGCCTCTTGTTGGT	binds downstream of sgRNA cutting side 57_sgRNA_D1_Cen
2	genotyping_2R	AGCAACTGTCAGTCCCTCTC	binds upstream of sgRNA cutting side 57_sgRNA_D1_Cen
3	genotyping_3F	TGCCGAGCTATTAGAAGGGG	binds downstream of sgRNA cutting side 58_sgRNA_D1_Tel
4	genotyping_4R	TCTGCTCGTTCACCTTACCT	binds upstream of sgRNA cutting side 58_sgRNA_D1_Tel
5	genotyping_5F	GAGACCCACCTGAAGACCAA	binds downstream of sgRNA cutting side 59_sgRNA_D2_Cen
6	genotyping_6R	CCGTCTCCCCAACACCTAAT	binds upstream of sgRNA cutting side 59_sgRNA_D2_Cen
7	genotyping_7F	TAGAATTCTGCAGTCGGGGC	binds downstream of sgRNA cutting side 60_sgRNA_D2_Tel
8	genotyping_8R	ACTGTTTTGGGAGACTGTGGA	binds upstream of sgRNA cutting side 60_sgRNA_D2_Tel
9	genotyping_9F	CCCCATTCTCAAGGTTCTCC	binds within Dessert 1
10	genotyping_10R	TGAGTTTCGCAGATGATCCA	binds within Dessert 1
11	genotyping_11F	GATTCATGGCCTCGTCTGTC	binds within Dessert 1
12	genotyping_12R	CGGGTGAAACACACAATCAA	binds within Dessert 1
13	genotyping_13F	TCAGGACCATTTTCCCATCTT	binds within Dessert 1
14	genotyping_14R	TTGCTTGCCCATCTGCTAAT	binds within Dessert 1
15	genotyping_15F	TTCACCTTGGGGTCTTTTGA	binds within Dessert 1
16	genotyping_16R	CTACCACAGCCTGGCTTGAC	binds within Dessert 1
17	genotyping_17F	TTCCCCTTCTGGATTGAGGT	binds within Dessert 2
18	genotyping_18R	CAAAATGAAACGGGCACTGT	binds within Dessert 2
19	genotyping_19F	AAATTGGCAAAGGGGACTACA	binds within Dessert 2
20	genotyping_20R	CCAACTCTGCCCACGATAAA	binds within Dessert 2
21	genotyping_21F	CCAAAGGCATCCACAGACAT	binds within Dessert 2
22	genotyping_22R	AAGGGCAGTGGAGGAAGTGT	binds within Dessert 2
23	genotyping_23F	TCCCCTGCAAGAGGGTTATT	binds within Zfp42R
24	genotyping_24R	GTCAGGCTCTGCCTGTTGTC	binds within Zfp42R
25	genotyping_25F	TCTTCCTCCCCACACAACAG	binds within Zfp42R
26	genotyping_26R	TCGGCAGAAGCAACTTAGGA	binds within Zfp42R
27	genotyping_27F	TGACCTGATTTATCAAGGCCATA	binds within Zfp42R
28	genotyping_28R	TGTATTGAGAGAGGCTGCTTTTG	binds within Zfp42R
29	genotyping_29F	TTCAACATCAGCCGCTACAG	binds within LacZ
30	genotyping_30R	CGTCGATATTCAGCCATGTG	binds within LacZ
31	genotyping_31F	TCCGTCCTTTTCAATGACCT	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
32	genotyping_32R	gccacgattgaaaccacttt	binds upstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
33	genotyping_33F	TGTGAGCAACTTTCTTAGCGGAAA	binds downstream of sgRNA cutting side 167_sgRNA_KI_D1mid
34	genotyping_34R	acctgcaggctagaagcaaa	binds withinbeta-globin promoter
35	genotyping_35F	GAACCATTTGTTAATAAGTAAAGCAG/	binds downstream of sgRNA cutting side 167_sgRNA_KI_D1mid
36	genotyping_36R	CAGGAAATTGAGCTCCTTTGA	binds upstream of sgRNA cutting side 167_sgRNA_KI_D1mid
37	genotyping_37F	CGGGTTAAGGGAGAGAGATAGGA	binds downstream of sgRNA cutting side 079_sgRNA_KI_D1
38	genotyping_38R	TCTTTCTAGAAGATGAAGGGCGA	binds binds ~50bp spacer sequence of knock-in construct
39	genotyping_39F	CCCAGGAGCTGACTATTTGG	binds downstream of sgRNA cutting side 079_sgRNA_KI_D1
40	genotyping_40R	ACCTGGAACGATTAGGCAGA	binds upstream of sgRNA cutting side 079_sgRNA_KI_D1
41	genotyping_41F	CCCCACCCATTTGTTTATTGAA	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
42	genotyping_42F	CGTCCTTTTCAATGACCTGA	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
43	genotyping_96R	CCGTAATGGGATAGGTcACG	binds withinLacZ
44	genotyping_97R	GTATAGCCCTGGCTGTCCTG	binds within genomic region close at KI sidefor endoZfp42_LacZ
45	genotyping_45R	cactctagaccccagcctct	binds upstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
46	genotyping_46F	ATGATCTCACGGACGCATGG	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
47	genotyping_47F	CCATTTGGGGCCTGATGGAT	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
48	genotyping_48R	ACAGTGGAAATCTAGGCGGC	binds upstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
49	genotyping_49F	TAAGCAGCAACAGCAAAACAAAA	binds downstream of sgRNA cutting side 151_sgRNA_KI_ZFP3
50	genotyping_50F	cataagcacaggggtgtcgt	binds downstream of sgRNA cutting side 78_sgRNA_KI_D2mid
51	genotyping_51F	CAGAAAGAGAGAGGAAGAAGTAGG	binds downstream of sgRNA cutting side 151_sgRNA_KI_ZFP3
52	genotyping_52R	CACACCTGCCACCTTACTCA	binds downstream of sgRNA cutting side 151_sgRNA_KI_ZFP3
53	genotyping_53F	CATAAATTCTGTACCAGCTTTACC C	binds downstream of sgRNA cutting side 78_sgRNA_KI_D2mid
54	genotyping_54R	GAAAGGGGAAAAGAACCTGAA	binds upstream of sgRNA cutting side 78_sgRNA_KI_D2mid
55	genotyping_55F	GCCTCAGGAGAGTTTGCATC	binds downstream of sgRNA cutting side 168_sgRNA_KI_D2
56	genotyping_56F	AGGTCCTGAAGAAGCTTGGC	binds within genomic regionat Rosa26
57	genotyping_57F	TGCCATGAGTCAAGCCAGTC	binds within genomic regionat Rosa26
58	genotyping_58F	GGTACCCTCGTGATCTGCAA	binds binds ~50bp spacer sequence of knock-in construct
59	genotyping_59F	GTTCGACTTGGTGATTGGTGGAAG	binds downstream of sgRNA cutting side 300_sgRNA_ALM_Dnmt3b_ex3_g1
60	genotyping_60R	GCCACACCCTCCTCTGAGC	binds upstream of sgRNA cutting side 300_sgRNA_ALM_Dnmt3b_ex3_g1
61	genotyping_61F	TCCAGAAGTTGGAGGACGAC	binds downstream of sgRNA cutting side 168_sgRNA_KI_D2
62	genotyping_62R	ACTCCTCGCAACGAGTCTGT	binds upstream of sgRNA cutting side 168_sgRNA_KI_D2
63	genotyping_63F	CAGGACAACGCCCACACA	binds downstream of sgRNA cutting side 152_Hom_Rosa26_ins_BGlob_LacZ
64	genotyping_64R	CCTCCTGGCTTCTGAGGAC	binds upstream of sgRNA cutting side 152_Hom_Rosa26_ins_BGlob_LacZ
65	genotyping_65F	AGTTCCCTGGGGGACAATAC	binds downstream of sgRNA cutting side 105_sg_KI_Ins_Triml
66	genotyping_66R	cggtcgctaccattaccagt	binds within LacZ
67	genotyping_67F	CCTGACCAGGATTTTGAGATG	binds downstream of sgRNA cutting side 105_sg_KI_Ins_Triml
68	genotyping_68R	GCTTACATCCTGGGAAGGTG	binds upstream of sgRNA cutting side 105_sg_KI_Ins_Triml
69	genotyping_69F	tgctcgcacgtacttcattc	binds within genomic region at Col1a
70	genotyping_70R	AAATGTGCCCTGTGAACCTC	binds within cloned Fat1-enh sequence
71	genotyping_71F	tgctcgcacgtacttcattc	binds within genomic region at Col1a
72	genotyping_72R	TATTGCACCTTTCCCACTCC	binds within cloned Fat1-enh sequence
73	genotyping_73F	GATCATGGACATGTGAGCCTGAGAC	binds downstream of sgRNA cutting side 301_sgRNA_ALM_Dnmt3b_gRNA
74	genotyping_74F	CAAGTGGTCTCAGAACTGCTAAGAG	binds upstream of sgRNA cutting side 301_sgRNA_ALM_Dnmt3b_gRNA
75	genotyping_75	TAAGTCGGCCTGTCTTGGTC	binds upstream of sgRNA cutting side 279_sgRNA_KI_Frg1sensor

Primer number	Name	Sequence (5'-->3')	Binding site
76	genotyping_76	AGAAAGACTGGAGTTGCAGA	binds binds ~50bp spacer sequence of knock-in construct
77	genotyping_77	AAGGTCAAGGCAGGCACTTA	binds downstream of sgRNA cutting side 279_sgRNA_KI_Frg1sensor
78	genotyping_78	AAGAGGGGGAAGGGATT	binds binds ~50bp spacer sequence of knock-in construct
79	genotyping_79F	AAACTTGCATTGGCTGTCCT	binds downstream of sgRNA cutting side 076_sg_KI_Ins_ZFP42
80	genotyping_80R	gacagtatcggcctcaggaa	binds withinLacZ
81	genotyping_81F	CAATGCAACTCCAGTGAGGA	binds downstream of sgRNA cutting side 267_sgRNA_Zfp42prom_cen
82	genotyping_82R	CATACCGTCATGCCCAACAC	binds upstream of sgRNA cutting side 268_sgRNA_Zfp42prom_tel
83	genotyping_83F	GCCTCCTACTGCTCCTTGTG	binds downstream of sgRNA cutting side 281_sgRNA_Fat1enh_cen
84	genotyping_84R	GGTGGGTCATGACGTCTTTT	binds upstream of sgRNA cutting side 282_sgRNA_Fat1enh_tel
85	genotyping_85R	CACACCCCAATTCACATCAA	binds upstream of sgRNA cutting side 282_sgRNA_Fat1enh_tel
86	genotyping_86F	CTGGAGTTGTGTGTGGATGG	binds within deleted region within 86Fat1enh3_wt_F
87	genotyping_87R	CCACTGGAGCGAAACAAAAT	binds within deleted region within 87Fat1enh3_wt_R
88	genotyping_88R	AAGGCAGACGGATTTCTGAC	binds downstream of sgRNA cutting side 269_sgRNA_Trimprom_cen
89	genotyping_89F	CGTGTCCCCCTTTCTAACAC	binds upstream of sgRNA cutting side 270_sgRNA_Trimprom_tel
90	genotyping_90F	TCCATGAAGGGGAGAGATTG	binds within deleted region of ΔZfp42
91	genotyping_91R	GAGCAGCAAGACCTTTTTGG	binds within deleted region of ΔZfp42
92	genotyping_92F	AAAAAGCAAGCGACTTTGGA	binds within deleted region of ΔZfp42Triml
93	genotyping_93R	GACAGCACCATGCTCTGAAA	binds within deleted region of ΔZfp42Triml
94	genotyping_94F	CTTTGCGTGGGTTAGGATGT	binds within genomic region at KI sidefor endoZfp42_LacZ
95	genotyping_95R	TCAGATGTGGAGTTGGTGGA	binds within genomic region at KI sidefor endoZfp42_LacZ
96	genotyping_96R	CCGTAATGGGATAGGTcACG	binds withinLacZ
97	genotyping_97R	GTATAGCCCTGGCTGTCCTG	binds within genomic region at KI sidefor endoZfp42_LacZ

Table 6 Primer sequences
Primer number for the assignment to cell lines of table 5. Name of the primer with corresponding sequence and information of binding site.

***Fat1* versus**	**stars**	**p-value**
Triml	*	0,0331
Zfp42	ns	0,6376
beta-Globim	***	0,0005

Table 7 qRTPCR statistics
Testing LacZ expression differences for statistical significance between Fat1-promoter driven and indicated promoters by using unpaired t-test with Welch's correction with two-tailed p-value and 95% confidence interval. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05

DE-Seq Adjusted P-Value	Limbs								ESC			
	Wt vs Wt ESCs	Wt vs ΔFat1-enh	Wt vs ΔZfp42R	Wt vs ΔD1	Wt vs ΔD2	Wt vs ΔD1+2	Wt vs ΔDNMT3B	Wt vs ΔDNMT3B-ΔD1+2	Wt vs ΔZfp42R	Wt vs ΔD1	Wt vs ΔD2	Wt vs ΔD1+2
Triml2	0.000	1.000	0.850	0.270	0.750	1.000	0.960	1.000	0.000	0.740	0.000	0.060
Zfp42	0.000	0.930	0.020	1.000	0.420	0.110	0.000	0.400	0.000	0.460	1.000	0.740
Fat1	0.490	0.580	0.790	0.000	0.560	0.000	0.000	0.000	0.900	0.910	0.920	0.770

Table 8 DE-Seq adjusted p-values for all RNA-seq samples
Testing for significant alterations in gene expression was done by using the DE-seq tool from Love et al. (Love et al. 2014). Adjusted p-value calculation in DE-Seq is based on the wald test combined with a Benjamini and Hochberg correction for multiple testing. The null hypothesis says that there is no differential expression between two sample groups (e.g. wt vs. mutant) and was rejected when p < 0.05. Thus, all comparisons with a p-adjusted value below 0.05 (yellow) report a significant change of expression for indicated genes.

Name	Sequence (5'-->3')
WISH_mouse_Fat1_F	caaactcctccccaaagtgc
WISH_mouse_Fat1_R	accacacagatccagcagaa
WISH_mouse_Zfp42-1_F	cgaggtgagttttccgaacc
WISH_mouse_Zfp42-1_R	agtttcgagctctccgtgaa
WISH_mouse_Zfp42-2_F	ggccagtccagaataccaga
WISH_mouse_Zfp42-2_R	cgtatgcaaaagtccccatc
WISH_mouse_Triml_F	ttattgggaggtggctgtgg
WISH_mouse_Triml_R	tggtgttcacaaagtctggatc
WISH_chicken_Fat1-1_F	gtgtcgtggtggttttggtt
WISH_chicken_Fat1-1_R	tcccctctctctctcttccc
WISH_chicken_Fat1-2_F	tgtcaaaccttggttgtggg
WISH_chicken_Fat1-2_R	tgtatgcttccactgcttcg
WISH_opossum_Fat1_F	GGTAGACATTTGGCCATGCT
WISH_opossum_Fat1_R	CACGATGGCGTAGGTAGGAT

Table 9 Primers used for generation of WISH probes

Primer	Sequence (5'-->3')
outer primer F - Zfp42 endog. promoter	TAAGATTTTTATTAATTTGGAAGAGTT
outer primer R - Zfp42 endog. promoter	CAAATCTATAACAAAAATACTAAAATA
inner primer F - Zfp42 endog. promoter	GGAAGAGTTAAAATTTTAAATTGTTG
inner primer R - Zfp42 endog. promoter	AATATAACTAAATCTCAAAACCAAAC
outer primer F - Zfp42 KI at 20kb	CCTATAACCAACTTTCATCAACAT
outer primer R - Zfp42 KI at 20kb	GTTTGTAATAGAGGTATTGAGATG
inner primer F - Zfp42 KI at 20kb	ACCTACAAAAACTACCACTC
inner primer R - Zfp42 KI at 20kb	ATGTGATTGAGTTTTAAGGTTAGG

Table 10 Primers used for bisulfite-cloning sequencing

ID	chr	start	end	size	probes	probes/kb
Flank 1	chr 8	40400000	41457521	1057521	7780	7.364
Desert 1	chr 8	41457521	43108935	1651414	8179	4.953
Zfp42R	chr 8	43108935	43402062	293127	1819	6.218
Desert 2	chr 8	43402062	44932011	1529949	8556	5.595
Flank 2	chr 8	44932011	46100000	1167989	11328	9.742

Table 11 Oligopaint library

Coordinates, size, number and density of oligopaint probe sets (35-41mers). Library was constructed following the procedures described by Beliveau et al. and was ordered as CustomArray in a 12KOligo pool format (Beliveau et al. 2015). Details for library design can be found here: http://genetics.med.harvard.edu/oligopaints.

Emulsion PCR with universal primers

Primer code	Primer name	Primer sequence	Primer name	Primer sequence
Univ1	BB297-FWD	GACTGGTACTCGCGTGACTTG	BB299-REV	GTAGGGACACCTCTGGACTGG

PCR1 with FWD 5' phosphorylation and REV 53mer primer

Primer code-color	Primer name	Primer sequence	Primer name	Primer sequence
A - green	BB82-FWD	/5Phos/GTATCGTGCAAGGGTGAATGC	Sec1-BB278-REV	/Sec1BS/GAGCAGTCACAGTCCAGAAGG
B - green	BB81-FWD	/5Phos/ATCCTAGCCCATACGGCAATG	Sec1-BB281-REV	/Sec1BS/GGACATGGGTCAGGTAGGTTG
B - red	BB81-FWD	/5Phos/ATCCTAGCCCATACGGCAATG	Sec6-BB281-REV	/Sec6BS/GGACATGGGTCAGGTAGGTTG
C - red	BB287-FWD	/5Phos/CGCTCGGTCTCCGTTCGTCTC	Sec6-BB288-REV	/Sec6BS/GGGCTAGGTACAGGGTTCAGC
D -green	BB293-FWD	/5Phos/CCGAGTCTAGCGTCTCCTCTG	Sec1-BB294-REV	/Sec1BS/AACAGAGCCAGCCTCTACCTG
E - green	BB298-FWD	/5Phos/CGTCAGTACAGGGTGTGATGC	Sec1-BB187-REV	/Sec1BS/TTGATCTTGACCCATCGAAGC
			Secondary Binding Sequences hooked to the REV 53mer primers	
			Sec1BS	CACCGACGTCGCATAGAACGGAAGAGCGTGTG
			Sec6BS	CACACGCTCTCCGTCTTGGCCGTGGTCGATCA

PCR2 with labeled REV 'back primer'

Primer name	Primer sequence	
BB506-Alexa488	/5Alex488N/CACCGACGTCGCATAGAACGG	Sec1BS compatible
BB511-ATTO565	/5ATTO565N/CACACGCTCTCCGTCTTGGC	Sec6BS compatible

Secondary Oligos carrying two fluorochromes

Oligo name	Oligo sequence
Sec1-Alexa488-X2	/5Alex488N/CACACGCTCTTCCGTTCTATGCGACGTCGGTGagatgttt/3AlexF488N/
Sec6-ATTO565-X2	/5ATTO565N/TGATCGACCACGGCCAAGACGGAGAGCGTGTGagatgttt/3ATTO565N/

Table 12 Primer and Oligo sequences for FISH
Sequences of universal primers, primers for two step PCRs and secondary oligos. Universal primer pairs are used in the emulsion PCR and bind a sequence flanking the sequence with genomic homology. ePCR is followed by two sequential PCRs leading to the addition of a secondary oligonucleotide (Sec1-Alexa 488 - green probes or Sec6-Atto 565 - red probes) containing two additional fluorophores for signal amplification.

mSBS/ FISH	Measurment	Object 1	Tissue	Condition	data points (n)	median	mean	Object 2	Tissue	Condition	data points (n)	median	mean	stars	p-value
FISH	Dis. to NE	Fat1	Limb	wt	28	0.2912	0.5831	Fat1	ESC	wt	34	0.2836	0.6148	ns	0.853
FISH	Dis. to NE	Zfp42R	Limb	wt	108	0.1407	0.3348	Zfp42R	ESC	wt	138	0.2358	0.5525	**	0.0047
mSBS	Dis. to NE	Zfp42R	Limb	wt 0.4 KbT	25	0.8137	0.8064	Zfp42R	Limb	wt 1.2 KbT	88	0.277	0.2948	***	<0.0001
mSBS	Dis. to NE	Zfp42R	Limb	wt 1.2 KbT	88	0.277	0.2948	Zfp42R	Limb	wt 3.0 KbT	31	0.1355	0.1355	***	<0.0001
mSBS	Dis. to NE	Zfp42R	Limb	wt 3.0KbT	31	0.1355	0.1355	Zfp42R	Limb	wt 0.4 KbT	25	0.8137	0.8064	***	<0.0001
mSBS	Dis. to NE	Fat1	Limb	wt 0.4 KbT	25	0.8574	0.8465	Fat1	Limb	wt 1.2 KbT	88	0.4431	0.4683	***	<0.0001
mSBS	Dis. to NE	Fat1	Limb	wt 1.2 KbT	88	0.4431	0.4683	Fat1	Limb	wt 3.0 KbT	31	0.2216	0.2214	***	<0.0001
mSBS	Dis. to NE	Fat1	Limb	wt 3.0KbT	31	0.2216	0.2214	Fat1	Limb	wt 0.4 KbT	25	0.8574	0.8465	***	<0.0001
FISH	Interm. Frac. with D1+D2	Fat1	Limb	wt	65	0.3906	0.3814	Fat1	ESC	wt	39	0.1892	0.2344	**	0.0012
FISH	Interm. Frac. with D1+D2	Zfp42R	Limb	wt	84	0.6497	0.6267	Zfp42R	ESC	wt	65	0.4502	0.429	***	<0.0001
mSBS	Interm. Frac. with D1+D2	Fat1	Limb	wt 0.4 KbT	48	0.1699	0.1944	Fat1	Limb	wt 1.2 KbT	88	0.1627	0.2037	ns	0.6868
mSBS	Interm. Frac. with D1+D2	Fat1	Limb	wt 1.2 KbT	88	0.1627	0.2037	Fat1	Limb	wt 3.0 KbT	54	0.1058	0.1242	***	<0.0001
mSBS	Interm. Frac. with D1+D2	Fat1	Limb	wt 3.0KbT	54	0.1058	0.1242	Fat1	Limb	wt 0.4 KbT	48	0.1699	0.1944	***	0.0006
mSBS	Interm. Frac. with D1+D2	Zfp42R	Limb	wt 0.4 KbT	48	0.2791	0.3143	Zfp42R	Limb	wt 1.2 KbT	88	0.2603	0.2862	ns	0.3416
mSBS	Interm. Frac. with D1+D2	Zfp42R	Limb	wt 1.2 KbT	88	0.2603	0.2862	Zfp42R	Limb	wt 3.0 KbT	54	0.1657	0.2408	ns	0.1734
mSBS	Interm. Frac. with D1+D2	Zfp42R	Limb	wt 3.0KbT	54	0.1657	0.2408	Zfp42R	Limb	wt 0.4 KbT	48	0.2791	0.3143	ns	0.0554
FISH	Com. Spher. with D1+D2	Fat1	Limb	wt	61	0.6757	0.6694	Fat1	ESC	wt	31	0.5989	0.5925	***	<0.0001
FISH	Com. Spher. with D1+D2	Zfp42R	Limb	wt	87	0.7195	0.7225	Zfp42R	ESC	wt	63	0.6228	0.6424	***	<0.0001
mSBS	Com. Spher. with D1+D2	Fat1	Limb	wt 0.4 KbT	48	0.9261	0.9245	Fat1	Limb	wt 1.2 KbT	88	0.9136	0.914	***	<0.0001
mSBS	Com. Spher. with D1+D2	Fat1	Limb	wt 1.2 KbT	88	0.9136	0.914	Fat1	Limb	wt 3.0 KbT	54	0.6898	0.6909	***	<0.0001
mSBS	Com. Spher. with D1+D2	Fat1	Limb	wt 3.0KbT	54	0.6898	0.6909	Fat1	Limb	wt 0.4 KbT	48	0.9261	0.9245	***	<0.0001
mSBS	Com. Spher. with D1+D2	Zfp42R	Limb	wt 0.4 KbT	48	0.9252	0.9244	Zfp42R	Limb	wt 1.2 KbT	88	0.9132	0.9129	***	<0.0001
mSBS	Com. Spher. with D1+D2	Zfp42R	Limb	wt 1.2 KbT	88	0.9132	0.9129	Zfp42R	Limb	wt 3.0 KbT	54	0.6809	0.6862	***	<0.0001
mSBS	Com. Spher. with D1+D2	Zfp42R	Limb	wt 3.0KbT	54	0.6809	0.6862	Zfp42R	Limb	wt 0.4 KbT	48	0.9252	0.9244	***	<0.0001

Table 13 FISH and mSBS statistics with different NE attachment energies
Measurements of FISH and mSBS using different NE-interaction energies of 0.4 K_BT, 1.2 K_BT and 3 K_BT, for: Distance (Dis.) to the NE, Intermingling fraction (Interm. Frac.) and Combined Sphericity (Comb. Spher.). Indicated samples were compared for statistic significant differences using unpaired t-test with Welch's correction with two-tailed p-value and 95% confidence interval. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05

mSBS/ FISH	Measurment	Object 1	Tissue	Condition	data points (n)	median	mean	Object 2	Tissue	Condition	data points (n)	median	mean	stars	p-value
FISH	Dis. to NE	Fat1	Limb	wt	28	0.2912	0.5831	Fat1	ESC	wt	34	0.2836	0.6148	ns	0.853
FISH	Dis. to NE	Zfp42R	Limb	wt	108	0.1407	0.3348	Zfp42R	ESC	wt	138	0.2358	0.5525	**	0.0047
FISH	Dis. to NE	D1	Limb	wt	60	0.1394	0.285	D1	ESC	wt	92	0.1697	0.463	*	0.0272
FISH	Dis. to NE	D2	Limb	wt	76	0.1225	0.2619	D2	ESC	wt	84	0.1569	0.4457	*	0.0205
mSBS	Dis. to NE	Fat1	Limb	wt 1.2 KbT	88	0.4431	0.4683	Fat1	ESC	wt 1.2 KbT	106	0.3181	0.3369	***	<0.0001
mSBS	Dis. to NE	Zfp42R	Limb	wt 1.2 KbT	88	0.277	0.2948	Zfp42R	ESC	wt 1.2 KbT	106	0.3993	0.3911	***	<0.0001
mSBS	Dis. to NE	D1	Limb	wt 1.2 KbT	88	0.3009	0.3276	D1	ESC	wt 1.2 KbT	106	0.2706	0.3099	ns	0.3122
mSBS	Dis. to NE	D2	Limb	wt 1.2 KbT	88	0.3846	0.4061	D2	ESC	wt 1.2 KbT	106	0.2562	0.2902	***	<0.0001
FISH	Interm. Frac. with D1+D2	Fat1	Limb	wt	65	0.3906	0.3814	Fat1	ESC	wt	39	0.1892	0.2344	**	0.0012
FISH	Interm. Frac. with D1+D2	Zfp42R	Limb	wt	84	0.6497	0.6267	Zfp42R	ESC	wt	65	0.4502	0.429	***	<0.0001
mSBS	Interm. Frac. with D1+D2	Zfp42R	Limb	wt 1.2 KbT	88	0.2603	0.2862	Zfp42R	ESC	wt 1.2 KbT	106	0.06645	0.09854	***	<0.0001
mSBS	Interm. Frac. with D1+D2	Fat1	Limb	wt 1.2 KbT	88	0.1627	0.2037	Fat1	ESC	wt 1.2 KbT	106	0.06156	0.07296	***	<0.0001
FISH	Com. Spher. with D1+D2	Fat1	Limb	wt	61	0.6757	0.6694	Fat1	ESC	wt	31	0.5989	0.5925	***	<0.0001
FISH	Com. Spher. with D1+D2	Zfp42R	Limb	wt	87	0.7195	0.7225	Zfp42R	ESC	wt	63	0.6228	0.6424	***	<0.0001
mSBS	Com. Spher. with D1D2	Fat1	Limb	wt 1.2 KbT	88	0.9136	0.914	Fat1	ESC	wt 1.2 KbT	106	0.863	0.8569	***	<0.0001
mSBS	Com. Spher. with D1D2	Zfp42R	Limb	wt 1.2 KbT	88	0.9132	0.9129	Zfp42R	ESC	wt 1.2 KbT	106	0.8634	0.8594	***	<0.0001
FISH	Interm. Frac.	D1+D2	Limb	wt	93	0.161	0.1599	D1+D2	ESC	wt	79	0.02576	0.07131	***	<0.0001
mSBS	Interm. Frac.	D1+D2	Limb	wt 1.2 KbT	88	0.1578	0.1811	D1+D2	ESC	wt 1.2 KbT	106	0.05955	0.06872	***	<0.0001
FISH	Com. Spher.	D1+D2	Limb	wt	98	0.6929	0.6779	D1+D2	ESC	wt	73	0.6427	0.6469	*	0.0187
mSBS	Com. Spher.	D1+D2	Limb	wt 1.2 KbT	88	0.9117	0.9117	D1+D2	ESC	wt 1.2 KbT	106	0.8614	0.8548	***	<0.0001

Table 14 FISH and mSBS statistics for wildtype limbs and ESCs

FISH and mSBS (1.2 K_BT) measurements for Distance (Dis.) to the NE, Intermingling fraction (Interm. Frac.) and Combined Sphericity (Comb. Spher.). Comparisons between indicated samples for statistic significant differences by using unpaired t-test with Welch's correction with two-tailed p-value and 95% confidence interval. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05.

Measurment	Object 1	Tissue	Condition	data points (n)	median	mean	Object 2	Tissue	Condition	data points (n)	median	mean	stars	p-value
Dis. to NE	Zfp42R	ESC	wt	138	0.2358	0.5525	Zfp42R	ESC	dRad21	16	0.2971	0.5107	ns	0.7772
Dis. to NE	Zfp42R	ESC	wt	138	0.2358	0.5525	Zfp42R	ESC	dCTCF	42	0.526	0.7879	ns	0.0756
Interm. Frac. with D1+D2	Zfp42R	ESC	wt	65	0.4502	0.429	Zfp42R	ESC	dRad21	35	0.193	0.2263	***	<0.0001
Interm. Frac. with D1+D2	Zfp42R	ESC	wt	65	0.4502	0.429	Zfp42R	ESC	dCTCF	49	0.3773	0.3781	ns	0.1514
Com. Spher. with D1+D2	Zfp42R	ESC	wt	63	0.6228	0.6424	Zfp42R	ESC	dRad21	27	0.5709	0.5793	***	0.0009
Com. Spher. with D1+D2	Zfp42R	ESC	wt	63	0.6228	0.6424	Zfp42R	ESC	dCTCF	47	0.6443	0.6399	ns	0.8739

Table 15 FISH statistics for CTCF and Rad21 depleted ESCs

Measurements for Distance (Dis.) to the NE, Intermingling fraction (Interm. Frac.) amd Combined Sphericity (Comb. Spher.). Comparisons between indicated samples for statistic significant differences by using unpaired t-test with Welch's correction with two-tailed p-value and 95% confidence interval. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05.

mSBS/ FISH	Measurment	Object 1	Condition	Condition	data points (n)	median	mean	Object 2	Tissue	Condition	data points (n)	median	mean	stars	p-value
mSBS	Interm. Frac.	Zfp42R+Fl1	Limb	wt	88	0.02349	0.02910	Zfp42R+Fl1	ESC	wt	106	0.02425	0.03509	ns	0.1783
mSBS	Interm. Frac.	Zfp42R+Fl1	Limb	wt	88	0.02349	0.02910	Zfp42R+Fl1	Limb	DeltaD1	71	0.03985	0.04771	***	0.0001
mSBS	Interm. Frac.	Zfp42R+Fl1	Limb	wt	88	0.02349	0.02910	Zfp42R+Fl1	Limb	DeltaD1+D2	75	0.05217	0.05972	***	0.0001
mSBS	Interm. Frac.	Zfp42R+Fl1	Limb	DeltaD1	71	0.03985	0.04771	Zfp42R+Fl1	Limb	DeltaD1+D2	75	0.05217	0.05972	*	0.0389
FISH	Interm. Frac.	Zfp42R+Fl1	Limb	wt	39	0.00191	0.02845	Zfp42R+Fl1	ESC	wt	44	0.0059	0.03318	ns	0.6906
FISH	Interm. Frac.	Zfp42R+Fl1	Limb	wt	39	0.00191	0.02845	Zfp42R+Fl1	Limb	DeltaD1	32	0.0297	0.05523	ns	0.0605
FISH	Interm. Frac.	Zfp42R+Fl1	Limb	wt	39	0.00191	0.02845	Zfp42R+Fl1	Limb	DeltaD1+D2	33	0.0577	0.07150	**	0.0097
FISH	Interm. Frac.	Zfp42R+Fl1	Limb	DeltaD1	32	0.0297	0.05523	Zfp42R+Fl1	Limb	DeltaD1+D2	33	0.0577	0.07150	ns	0.3685
mSBS	Dis. to NE	Zfp42R	ESC	wt	106	0.3993	0.3911	Zfp42R	Limb	wt	88	0.2770	0.2948	***	0.0001
mSBS	Dis. to NE	Zfp42R	Limb	wt	88	0.2770	0.2948	Zfp42R	Limb	DeltaD1	71	0.2366	0.2672	ns	0.1549
mSBS	Dis. to NE	Zfp42R	Limb	wt	88	0.2770	0.2948	Zfp42R	Limb	DeltaD2	71	0.2409	0.2637	ns	0.0823
mSBS	Dis. to NE	Zfp42R	Limb	wt	88	0.2770	0.2948	Zfp42R	Limb	DeltaD1+D2	75	0.2516	0.3452	ns	0.0741
mSBS	Dis. to NE	Zfp42R	Limb	DeltaD1	71	0.2366	0.2672	Zfp42R	Limb	DeltaD1+D2	75	0.2516	0.3452	*	0.0036
mSBS	Dis. to NE	Zfp42R	Limb	DeltaD2	71	0.2409	0.2637	Zfp42R	Limb	DeltaD1+D2	75	0.2516	0.3452	**	0.0016
mSBS	Dis. to NE	Zfp42R	Limb	DeltaD1	71	0.2366	0.2672	Zfp42R	Limb	DeltaD2	71	0.2409	0.2637	ns	0.8129
mSBS	Dis. to NE	Zfp42R	ESC	wt	106	0.3993	0.3911	Zfp42R	Limb	DeltaD1+D2	34	0.2516	0.3452	ns	0.0831
FISH	Dis. to NE	Zfp42R	ESC	wt	138	0.2358	0.5525	Zfp42R	Limb	wt	108	0.1407	0.3348	**	0.0047
FISH	Dis. to NE	Zfp42R	Limb	wt	108	0.1407	0.3348	Zfp42R	Limb	DeltaD1	28	0.1216	0.3456	ns	0.9246
FISH	Dis. to NE	Zfp42R	Limb	wt	108	0.1407	0.3348	Zfp42R	Limb	DeltaD2	32	0.1256	0.3956	ns	0.5969
FISH	Dis. to NE	Zfp42R	Limb	wt	108	0.1407	0.3348	Zfp42R	Limb	DeltaD1+D2	34	0.3652	0.6267	*	0.0295
FISH	Dis. to NE	Zfp42R	Limb	DeltaD1	28	0.1216	0.3456	Zfp42R	Limb	DeltaD1+D2	34	0.3652	0.6267	ns	0.0829
FISH	Dis. to NE	Zfp42R	Limb	DeltaD2	32	0.1256	0.3956	Zfp42R	Limb	DeltaD1+D2	34	0.3652	0.6267	ns	0.1536
FISH	Dis. to NE	Zfp42R	Limb	DeltaD1	28	0.1216	0.3456	Zfp42R	Limb	DeltaD2	32	0.1256	0.3956	ns	0.7352
FISH	Dis. to NE	Zfp42R	ESC	wt	138	0.2358	0.5525	Zfp42R	Limb	DeltaD1+D2	34	0.3652	0.6267	ns	0.5869
mSBS	Dis. to NE	Fl1	ESC	wt	106	0.4081	0.4723	Fl1	Limb	wt	88	0.3497	0.4336	ns	0.1373
mSBS	Dis. to NE	Fl1	Limb	wt	88	0.3497	0.4336	Fl1	Limb	DeltaD1	71	0.3351	0.3650	**	0.0039
mSBS	Dis. to NE	Fl1	Limb	wt	88	0.3497	0.4336	Fl1	Limb	DeltaD1+D2	75	0.3243	0.3562	**	0.0019
mSBS	Dis. to NE	Fl1	Limb	DeltaD1	71	0.3351	0.3650	Fl1	Limb	DeltaD1+D2	75	0.3243	0.3562	ns	0.6546
mSBS	Dis. to NE	Fl1	ESC	wt	106	0.4081	0.4723	Fl1	Limb	DeltaD1+D2	75	0.3243	0.3562	***	0.0001
FISH	Dis. to NE	Fl1	ESC	wt	72	0,4213	0,7313	Fl1	Limb	wt	42	0,2366	0,4363	*	0.0188
FISH	Dis. to NE	Fl1	Limb	wt	42	0,2366	0,4363	Fl1	Limb	DeltaD1	8	0,4164	0,6613	ns	0.3823
FISH	Dis. to NE	Fl1	Limb	wt	42	0,2366	0,4363	Fl1	Limb	DeltaD1+D2	22	0,2008	0,341	ns	0.4785
FISH	Dis. to NE	Fl1	Limb	DeltaD1	8	0,4164	0,6613	Fl1	Limb	DeltaD1+D2	22	0,2008	0,341	ns	0.2351
FISH	Dis. to NE	Fl1	ESC	wt	72	0,4213	0,7313	Fl1	Limb	DeltaD1+D2	22	0,2008	0,341	**	0.0083

Table 16 FISH and mSBS statistics for mutant limbs

FISH and mSBS (1.2 K_BT) measurements for Intermingling fraction (Interm. Frac.) and Distance (Dis.) to the NE. Comparisons between indicated samples for statistic significant differences by using unpaired t-test with Welch's correction with two-tailed p-value and 95% confidence interval. *** p<0.001, ** p<0.01, * p<0.05 and ns p>0.05.

	Common Name	Scientific Name	Build	Source	Function
mouse-chicken comparison	Mouse	*Mus musculus*	mm10	NCBI	Reference clade
	Daurian Ground Squirrel	*Spermophilus dauricus*	speDau1	zoonomia / ncbi	
	Human	*Homo sapiens*	hg38	NCBI	
	North Atlantic Right Whale	*Eubalaena glacialis*	eubGla1	MPI	
	African Bush Elephant	*Loxodonta africana*	loxAfr3	zoonomia / ncbi	
	Opossum	*Monodelphis domestica*	monDom5	MPI	Target clade
	Koala	*Phascolarctos cinereus*	phaCin1	dnazoo	
	Tasmanian Devil	*Sarcophilus harrisii*	sarHar1	zoonomia / ncbi	
	Platypus	*Ornithorhynchus anatinus*	ornAna4	NCBI	Outgroup
	Western Clawed Frog	*Xenopus tropicalis*	xenTro9	NCBI	
	Elephant Shark	*Callorhinchus milii*	calMil1	NCBI	
mouse-opossum comparison	Mouse	*Mus musculus*	mm10	NCBI	Reference clade
	Human	*Homo sapiens*	hg38	NCBI	
	Opossum	*Monodelphis domestica*	monDom5	MPI	
	Platypus	*Ornithorhynchus anatinus*	ornAna4	NCBI	
	Chicken	*Gallus gallus*	galGal6	NCBI	Target clade
	Adelie Penguin	*Pygoscelis adeliae*	pygAde1	GigaDB	
	Ostrich	*Strutio camelus*	strCam1	dnazoo	
	Chinese Alligator	*Alligator sinensis*	allSin1	dnazoo	
	Green Anole Lizard	*Anolis carolinensis*	anoCar2	NCBI	
	Western Clawed Frog	*Xenopus tropicalis*	xenTro9	NCBI	Outgroup
	Zebrafish	*Danio rerio*	danRer10	NCBI	
	Spotted Gar	*Lepistosteus oculatus*	lepOcu1	NCBI	
	Elephant Shark	*Callorhinchus milii*	calMil1	UCSC	
opossum-chicken comparison	Opossum	*Monodelphis domestica*	monDom5	MPI	Reference clade
	Koala	*Phascolarctos cinereus*	phaCin1	dnazoo	
	Tasmanian Devil	*Sarcophilus harrisii*	sarHar1	zoonomia/ncbi	
	Platypus	*Ornithorhynchus anatinus*	ornAna4	NCBI	
	African Bush Elephant	*Loxodonta africana*	loxAfr3	zoonomia/ncbi	
	Chicken	*Gallus gallus*	galGal6	NCBI	Target clade
	Adelie Penguin	*Pygoscelis adeliae*	pygAde1	GigaDB	
	Ostrich	*Strutio camelus*	strCam1	dnazoo	
	Chinese Alligator	*Alligator sinensis*	allSin1	dnazoo	
	Green Anole Lizard	*Anolis carolinensis*	anoCar2	NCBI	
	Western Clawed Frog	*Xenopus tropicalis*	xenTro9	NCBI	Outgroup
	Elephant Shark	*Callorhinchus milii*	calMil1	UCSC	

Table 17 Bridging Species
Sets of bridging species used for the stepped pairwise sequence alignment approach. Also see Material&Methods 16.

REFERENCES

Abdennur, N., and L. A. Mirny. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* 36 (1):311-316.

Acemel, R. D., I. Maeso, and J. L. Gomez-Skarmeta. 2017. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdisciplinary Reviews-Developmental Biology* 6 (3).

Acemel, R. D., J. J. Tena, I. Irastorza-Azcarate, F. Marlétaz, C. Gómez-Marín, E. de la Calle-Mustienes, S. Bertrand, S. G. Diaz, D. Aldea, J. M. Aury, S. Mangenot, P. W. Holland, D. P. Devos, I. Maeso, H. Escrivá, and J. L. Gómez-Skarmeta. 2016. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet* 48 (3):336-341.

Alexander, J. M., J. Guan, B. Li, L. Maliskova, M. Song, Y. Shen, B. Huang, S. Lomvardas, and O. D. Weiner. 2019. Live-cell imaging reveals enhancer-dependent. *Elife* 8.

Aljahani, A., P. Hua, M. A. Karpinska, K. Quililan, J. O. J. Davies, and A. M. Oudelaar. 2022. Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* 13 (1):2139.

Allen, M. P., and D. J. Tildesley. 2017. Computer Simulation of Liquids Oxford University Press.

Anania, C., and D. G. Lupiáñez. 2020. Order and disorder: abnormal 3D chromatin organization in human disease. *Brief Funct Genomics* 19 (2):128-138.

Anderson, E., P. S. Devenney, R. E. Hill, and L. A. Lettice. 2014. Mapping the Shh long-range regulatory domain. *Development* 141 (20):3934-3943.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507 (7493):455-461.

Andrey, G., T. Montavon, B. Mascrez, F. Gonzalez, D. Noordermeer, M. Leleu, D. Trono, F. Spitz, and D. Duboule. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340 (6137):1234167.

Arnold, C. D., M. A. Zabidi, M. Pagani, M. Rath, K. Schernhuber, T. Kazmar, and A. Stark. 2017. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat Biotechnol* 35 (2):136-144.

Artus, J., and A. K. Hadjantonakis. 2011. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol Biol* 693:37-56.

Ball, G., J. Demmerle, R. Kaufmann, I. Davis, I. M. Dobbie, and L. Schermelleh. 2015. SIMcheck: a Toolbox for Successful Super-resolution Structured Illumination Microscopy. *Sci Rep* 5:15915.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27 (2 Pt 1):299-308.

Bannister, A. J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Res* 21 (3):381-395.

Baranasic, D., M. Hörtenhuber, P. Balwierz, T. Zehnder, A. K. Mukarram, C. Nepal, C. Varnai, Y. Hadzhiev, A. Jimenez-Gonzalez, N. Li, J. Wragg, F. D'Orazio, N. Díaz, B. Hernández-Rodríguez, Z. Chen, M. Stoiber, M. Dong, I. Stevens, S. E. Ross, A. Eagle, R. Martin, P. Obasaju, S. Rastegar, A. C. McGarvey, W. Kopp, E. Chambers, D. Wang, H. R. Kim, R. D. Acemel, S. Naranjo, M. Lapinski, V. Chong, S. Mathavan, B. Peers, T. Sauka-Spengler, M. Vingron, P. Carninci, U. Ohler, S. A. Lacadie, S. Burgess, C. Winata, F. van Eeden, J. M. Vaquerizas, J. L. Gómez-Skarmeta, D. Onichtchouk, B. J. Brown, O. Bogdanovic, M. Westerfield, F. C. Wardle, C. O. Daub, B. Lenhard, and F.

Müller. 2021. Integrated annotation and analysis of genomic features reveal new types of functional elements and large-scale epigenetic phenomena in the developing zebrafish. *bioRxiv*:2021.2008.2009.454869.

Barbieri, M., M. Chotalia, J. Fraser, L. M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi. 2012. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A* 109 (40):16173-16178.

Bauer, M., E. Vidal, E. Zorita, N. Uresin, S. F. Pinter, G. J. Filion, and B. Payer. 2021. Chromosome compartments on the inactive X guide TAD formation independently of transcription during X-reactivation. *Nat Commun* 12 (1):3499.

Beagrie, R. A., and A. Pombo. 2016. Gene activation by metazoan enhancers: Diverse mechanisms stimulate distinct steps of transcription. *Bioessays* 38 (9):881-893.

Beard, C., K. Hochedlinger, K. Plath, A. Wutz, and R. Jaenisch. 2006. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44 (1):23-28.

Behringer, R., M. D. Anderson, K. V. Nagy, and A. Nagy. 1994. Manipulating the Mouse Embryo: A Laboratory Manual Cold Spring Harbor Laboratory Press.

Belaghzal, H., T. Borrman, A. D. Stephens, D. L. Lafontaine, S. V. Venev, Z. Weng, J. F. Marko, and J. Dekker. 2021. Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. *Nat Genet* 53 (3):367-378.

Beliveau, B. J., A. N. Boettiger, M. S. Avendano, R. Jungmann, R. B. McCole, E. F. Joyce, C. Kim-Kiselak, F. Bantignies, C. Y. Fonseka, J. Erceg, M. A. Hannan, H. G. Hoang, D. Colognori, J. T. Lee, W. M. Shih, P. Yin, X. Zhuang, and C. T. Wu. 2015. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat Commun* 6:7147.

Beliveau, B. J., E. F. Joyce, N. Apostolopoulos, F. Yilmaz, C. Y. Fonseka, R. B. McCole, Y. Chang, J. B. Li, T. N. Senaratne, B. R. Williams, J. M. Rouillard, and C. T. Wu. 2012. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences of the United States of America* 109 (52):21301-21306.

Beliveau, B. J., J. Y. Kishi, G. Nir, H. M. Sasaki, S. K. Saka, S. C. Nguyen, C. T. Wu, and P. Yin. 2018. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc Natl Acad Sci U S A* 115 (10):E2183-E2192.

Belmont, A. S. 2022. Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harb Perspect Biol* 14 (7).

Benabdallah, N. S., I. Williamson, R. S. Illingworth, L. Kane, S. Boyle, D. Sengupta, G. R. Grimes, P. Therizols, and W. A. Bickmore. 2019. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* 76 (3):473-484.e477.

Benko, S., J. A. Fantes, J. Amiel, D. J. Kleinjan, S. Thomas, J. Ramsay, N. Jamshidi, A. Essafi, S. Heaney, C. T. Gordon, D. McBride, C. Golzio, M. Fisher, P. Perry, V. Abadie, C. Ayuso, M. Holder-Espinasse, N. Kilpatrick, M. M. Lees, A. Picard, I. K. Temple, P. Thomas, M. P. Vazquez, M. Vekemans, H. Roest Crollius, N. D. Hastie, A. Munnich, H. C. Etchevers, A. Pelet, P. G. Farlie, D. R. Fitzpatrick, and S. Lyonnet. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics* 41 (3):359-364.

Bergman, D. T., T. R. Jones, V. Liu, J. Ray, E. Jagoda, L. Siraj, H. Y. Kang, J. Nasser, M. Kane, A. Rios, T. H. Nguyen, S. R. Grossman, C. P. Fulco, E. S. Lander, and J. M. Engreitz. 2022. Compatibility rules of human enhancer and promoter sequences. *Nature* 607 (7917):176-184.

Bergman, D. T., T. R. Jones, V. Liu, L. Siraj, H. Y. Kang, J. Nasser, M. Kane, T. H. Nguyen, S. R. Grossman, C. P. Fulco, E. S. Lander, and J. M. Engreitz. 2021. Compatibility logic of human enhancer and promoter sequences. *bioRxiv*:2021.2010.2023.462170.

Bertolino, E., and H. Singh. 2002. POU/TBP cooperativity: a mechanism for enhancer action from a distance. *Mol Cell* 10 (2):397-407.

Bianco, S., D. G. Lupianez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schopflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos, and M. Nicodemi. 2018. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet* 50 (5):662-667.

Bintu, B., L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362 (6413).

Bock, C., S. Reither, T. Mikeska, M. Paulsen, J. Walter, and T. Lengauer. 2005. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* 21 (21):4067-4068.

Boehning, M., C. Dugast-Darzacq, M. Rankovic, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, and M. Zweckstetter. 2018. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* 25 (9):833-840.

Bolt, C. C., L. Lopez-Delisle, A. Hintermann, B. Mascrez, A. Rauseo, G. Andrey, and D. Duboule. 2022. Context-dependent enhancer function revealed by targeted inter-TAD relocation. *Nat Commun* 13 (1):3488.

Bonev, B., and G. Cavalli. 2016. Organization and function of the 3D genome. *Nature reviews. Genetics* 17 (12):772.

Bonev, B., N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. L. Xu, X. D. Lv, J. P. Hugnot, A. Tanay, and G. Cavalli. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171 (3):557-+.

Borgel, J., S. Guibert, Y. Li, H. Chiba, D. Schubeler, H. Sasaki, T. Forne, and M. Weber. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42 (12):1093-1100.

Brosh, R., C. Coelho, A. M. Ribeiro-dos-Santos, G. Ellis, M. S. Hogan, H. J. Ashe, N. Somogyi, R. Ordoñez, R. D. Luther, E. Huang, J. D. Boeke, and M. T. Maurano. 2022. Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. *bioRxiv*:2022.2006.2020.495832.

Brueckner, L., P. A. Zhao, T. van Schaik, C. Leemans, J. Sima, D. Peric-Hupkes, D. M. Gilbert, and B. van Steensel. 2020. Local rewiring of genome-nuclear lamina interactions by transcription. *EMBO J* 39 (6):e103159.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10 (12):1213-1218.

Buenrostro, J. D., B. Wu, H. Y. Chang, and W. J. Greenleaf. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109:21 29 21-21 29 29.

Bulger, M., and M. Groudine. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144 (3):327-339.

Butler, J. E., and J. T. Kadonaga. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* 15 (19):2515-2519.

Calo, E., and J. Wysocka. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49 (5):825-837.

Cao, J., M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, and J. Shendure. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566 (7745):496-502.

Carter, B., and K. Zhao. 2021. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet* 22 (4):235-250.

Chakraborty, A., and F. Ay. 2019. The role of 3D genome organization in disease: From compartments to single nucleotides. *Semin Cell Dev Biol* 90:104-113.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic

reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327 (5963):302-305.
Chen, H., M. Levo, L. Barinov, M. Fujioka, J. B. Jaynes, and T. Gregor. 2018a. Dynamic interplay between enhancer-promoter topology and gene activity. *Nature Genetics* 50 (9):1296-1303.
Chen, J., H. Boyaci, and E. A. Campbell. 2021. Diverse and unified mechanisms of transcription initiation in bacteria. *Nat Rev Microbiol* 19 (2):95-109.
Chen, Y., Y. Zhang, Y. C. Wang, L. G. Zhang, E. K. Brinkman, S. A. Adams, R. Goldman, B. van Steensel, J. Ma, and A. S. Belmont. 2018b. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *Journal of Cell Biology* 217 (11):4025-4048.
Chen, Z., V. Snetkova, G. Bower, S. Jacinto, B. Clock, I. Barozzi, B. J. Mannion, A. Alcaina-Caro, J. Lopez-Rios, D. E. Dickel, A. Visel, L. A. Pennacchio, and E. Z. Kvon. 2022. Widespread Increase in Enhancer—Promoter Interactions during Developmental Enhancer Activation in Mammals. *bioRxiv*:2022.2011.2018.516017.
Chiariello, A. M., C. Annunziatella, S. Bianco, A. Esposito, and M. Nicodemi. 2016. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep* 6:29775.
Cho, W. K., J. H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and Cisse, II. 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361 (6400):412-415.
Chong, S. S., C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq, and R. Tjian. 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361 (6400).
Ciani, L., A. Patel, N. D. Allen, and C. ffrench-Constant. 2003. Mice lacking the giant protocadherin mFAT1 exhibit renal slit junction abnormalities and a partially penetrant cyclopia and anophthalmia phenotype. *Mol Cell Biol* 23 (10):3575-3582.
Conte, M., L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, and M. Nicodemi. 2020. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat Commun* 11 (1):3289.
Crawford, G. E., I. E. Holt, J. Whittle, B. D. Webb, D. Tai, S. Davis, E. H. Margulies, Y. Chen, J. A. Bernat, D. Ginsburg, D. Zhou, S. Luo, T. J. Vasicek, M. J. Daly, T. G. Wolfsberg, and F. S. Collins. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* 16 (1):123-131.
Croft, B., T. Ohnesorg, J. Hewitt, J. Bowles, A. Quinn, J. Tan, V. Corbin, E. Pelosi, J. van den Bergen, R. Sreenivasan, I. Knarston, G. Robevska, D. C. Vu, J. Hutson, V. Harley, K. Ayers, P. Koopman, and A. Sinclair. 2018. Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nat Commun* 9 (1):5319.
Daneshvar, K., M. B. Ardehali, I. A. Klein, F. K. Hsieh, A. J. Kratkiewicz, A. Mahpour, S. O. L. Cancelliere, C. Zhou, B. M. Cook, W. Li, J. V. Pondick, S. K. Gupta, S. P. Moran, R. A. Young, R. E. Kingston, and A. C. Mullen. 2020. lncRNA DIGIT and BRD3 protein form phase-separated condensates to regulate endoderm differentiation. *Nat Cell Biol* 22 (10):1211-1222.
Davidson, I. F., B. Bauer, D. Goetz, W. Tang, G. Wutz, and J. M. Peters. 2019. DNA loop extrusion by human cohesin. *Science* 366 (6471):1338-1345.
de Mendoza, A., T. V. Nguyen, E. Ford, D. Poppe, S. Buckberry, J. Pflueger, M. R. Grimmer, S. Stolzenburg, O. Bogdanovic, A. Oshlack, P. J. Farnham, P. Blancafort, and R. Lister. 2022. Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biol* 23 (1):163.
De Santa, F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C. L. Wei, and G. Natoli. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 8 (5):e1000384.
de Wit, E. 2020. TADs as the Caller Calls Them. *J Mol Biol* 432 (3):638-642.
Deaton, A. M., and A. Bird. 2011. CpG islands and the regulation of transcription. *Genes Dev* 25 (10):1010-1022.

Dekker, J., M. A. Marti-Renom, and L. A. Mirny. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics* 14 (6):390-403.
Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing chromosome conformation. *Science* 295 (5558):1306-1311.
Delaneau, O., M. Zazhytska, C. Borel, G. Giannuzzi, G. Rey, C. Howald, S. Kumar, H. Ongen, K. Popadin, D. Marbach, G. Ambrosini, D. Bielser, D. Hacker, L. Romano, P. Ribaux, M. Wiederkehr, E. Falconnet, P. Bucher, S. Bergmann, S. E. Antonarakis, A. Reymond, and E. T. Dermitzakis. 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* 364 (6439).
Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P. D. Gregory, A. Dean, and G. A. Blobel. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149 (6):1233-1244.
Deng, W., J. W. Rupon, I. Krivega, L. Breda, I. Motta, K. S. Jahn, A. Reik, P. D. Gregory, S. Rivella, A. Dean, and G. A. Blobel. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 158 (4):849-860.
Despang, A., R. Schöpflin, M. Franke, S. Ali, I. Jerkovic, C. Paliou, W.-L. Chan, B. Timmermann, L. Wittler, M. Vingron, S. Mundlos, and D. M. Ibrahim. 2019. Functional dissection of TADs reveals non-essential and instructive roles in regulating gene expression. *bioRxiv*:566562.
Dixon, J. R., D. U. Gorkin, and B. Ren. 2016. Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell* 62 (5):668-680.
Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenkov, J. R. Ecker, J. A. Thomson, and B. Ren. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518 (7539):331-336.
Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485 (7398):376-380.
Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1):15-21.
Dong, Z., H. Wang, H. Chen, H. Jiang, J. Yuan, Z. Yang, W. J. Wang, F. Xu, X. Guo, Y. Cao, Z. Zhu, C. Geng, W. C. Cheung, Y. K. Kwok, H. Yang, T. Y. Leung, C. C. Morton, S. W. Cheung, and K. W. Choy. 2018. Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. *Genet Med* 20 (7):697-707.
Du, Z., H. Zheng, B. Huang, R. Ma, J. Wu, X. Zhang, J. He, Y. Xiang, Q. Wang, Y. Li, J. Ma, K. Zhang, Y. Wang, M. Q. Zhang, J. Gao, J. R. Dixon, X. Wang, J. Zeng, and W. Xie. 2017. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547 (7662):232-235.
Durand, N. C., M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3 (1):95-98.
Eakin, G. S., and A. K. Hadjantonakis. 2006. Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Nat Protoc* 1 (3):1145-1153.
Erdel, F., and K. Rippe. 2018. Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* 114 (10):2262-2270.
Eres, I. E., and Y. Gilad. 2021. A TAD Skeptic: Is 3D Genome Topology Conserved? *Trends Genet* 37 (3):216-223.
Eres, I. E., K. Luo, C. J. Hsiao, L. E. Blake, and Y. Gilad. 2019. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet* 15 (7):e1008278.

Esposito, A., S. Bianco, L. Fiorillo, M. Conte, A. Abraham, F. Musella, M. Nicodemi, A. Prisco, and A. M. Chiariello. 2021. Polymer models are a versatile tool to study chromatin 3D organization. *Biochem Soc Trans* 49 (4):1675-1684.

Farré, D., N. Bellora, L. Mularoni, X. Messeguer, and M. M. Albà. 2007. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol* 8 (7):R140.

Fasciani, A., S. D'Annunzio, V. Poli, L. Fagnocchi, S. Beyes, D. Michelatti, F. Corazza, L. Antonelli, F. Gregoretti, G. Oliva, R. Belli, D. Peroni, E. Domenici, S. Zambrano, D. Intartaglia, C. Settembre, I. Conte, C. Testi, P. Vergyris, G. Ruocco, and A. Zippo. 2020. MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome. *Nat Genet* 52 (12):1397-1411.

Finlan, L. E., D. Sproul, I. Thomson, S. Boyle, E. Kerr, P. Perry, B. Ylstra, J. R. Chubb, and W. A. Bickmore. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics* 4 (3):e1000039.

Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suva, and B. E. Bernstein. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529 (7584):110-114.

Fong, S. L., and J. A. Capra. 2021. Modeling the Evolutionary Architectures of Transcribed Human Enhancer Sequences Reveals Distinct Origins, Functions, and Associations with Human Trait Variation. *Mol Biol Evol* 38 (9):3681-3696.

Forcato, M., C. Nicoletti, K. Pal, C. M. Livi, F. Ferrari, and S. Bicciato. 2017. Comparison of computational methods for Hi-C data analysis. *Nat Methods* 14 (7):679-685.

Franke, M., D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schopflin, K. Kraft, R. Kempfer, I. Jerkovic, W. L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz, and S. Mundlos. 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538 (7624):265-269.

Fraser, J., C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, S. Q. Xie, K. J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A. R. Forrest, C. A. Semple, J. Dostie, A. Pombo, and M. Nicodemi. 2015. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular systems biology* 11 (12):852.

Fudenberg, G., M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15 (9):2038-2049.

Furniss, D., L. A. Lettice, I. B. Taylor, P. S. Critchley, H. Giele, R. E. Hill, and A. O. Wilkie. 2008. A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum Mol Genet* 17 (16):2417-2423.

Gabriele, M., H. B. Brandão, S. Grosse-Holz, A. Jha, G. M. Dailey, C. Cattoglio, T. S. Hsieh, L. Mirny, C. Zechner, and A. S. Hansen. 2022. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* 376 (6592):496-501.

Galouzis, C. C., and E. E. M. Furlong. 2022. Regulating specificity in enhancer-promoter communication. *Curr Opin Cell Biol* 75:102065.

Gardiner-Garden, M., and M. Frommer. 1987. CpG islands in vertebrate genomes. *J Mol Biol* 196 (2):261-282.

George, S. H., M. Gertsenstein, K. Vintersten, E. Korets-Smith, J. Murphy, M. E. Stevens, J. J. Haigh, and A. Nagy. 2007. Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc Natl Acad Sci U S A* 104 (11):4455-4460.

Ghavi-Helm, Y. 2020. Functional Consequences of Chromosomal Rearrangements on Gene Expression: Not So Deleterious After All? *J Mol Biol* 432 (3):665-675.

Ghavi-Helm, Y., A. Jankowski, S. Meiers, R. R. Viales, J. O. Korbel, and E. E. M. Furlong. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 51 (8):1272-1282.

Goel, V. Y., M. K. Huseyin, and A. S. Hansen. 2022. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *bioRxiv*:2022.2007.2012.499637.

Gratz, S. J., A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue, M. M. Harrison, J. Wildonger, and K. M. O'Connor-Giles. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194 (4):1029-1035.

Green, M. R., and J. Sambrook. 2012. *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory Press.

Grosswendt, S., H. Kretzmer, Z. D. Smith, A. S. Kumar, S. Hetzel, L. Wittler, S. Klages, B. Timmermann, S. Mukherji, and A. Meissner. 2020. Epigenetic regulator function through mouse gastrulation. *Nature*.

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453 (7197):948-951.

Gustafsson, M. G., L. Shao, P. M. Carlton, C. J. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat. 2008. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys J* 94 (12):4957-4970.

Gómez-Marín, C., J. J. Tena, R. D. Acemel, M. López-Mayorga, S. Naranjo, E. de la Calle-Mustienes, I. Maeso, L. Beccari, I. Aneas, E. Vielmas, P. Bovolenta, M. A. Nobrega, J. Carvajal, and J. L. Gómez-Skarmeta. 2015. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A* 112 (24):7542-7547.

Haarhuis, J. H. I., R. H. van der Weide, V. A. Blomen, J. O. Yanez-Cuna, M. Amendola, M. S. van Ruiten, P. H. L. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R. Brummelkamp, E. de Wit, and B. D. Rowland. 2017. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* 169 (4):693-707 e614.

Haberle, V., C. D. Arnold, M. Pagani, M. Rath, K. Schernhuber, and A. Stark. 2019. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* 570 (7759):122-126.

Haberle, V., and A. Stark. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature reviews. Molecular cell biology* 19 (10):621-637.

Hansen, A. S., I. Pustova, C. Cattoglio, R. Tjian, and X. Darzacq. 2017. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6.

Harmston, N., E. Ing-Simmons, G. Tan, M. Perry, M. Merkenschlager, and B. Lenhard. 2017. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* 8 (1):441.

Heger, P., B. Marin, M. Bartkuhn, E. Schierenberg, and T. Wiehe. 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A* 109 (43):17507-17512.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459 (7243):108-112.

Helmbacher, F. 2018. Tissue-specific activities of the Fat1 cadherin cooperate to control neuromuscular morphogenesis. *PLoS Biol* 16 (5):e2004734.

Hildebrand, E. M., and J. Dekker. 2020. Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem Sci* 45 (5):385-396.

Hiragami-Hamada, K., S. Soeroes, M. Nikolov, B. Wilkins, S. Kreuz, C. Chen, I. A. De La Rosa-Velázquez, H. M. Zenn, N. Kost, W. Pohl, A. Chernev, D. Schwarzer, T. Jenuwein, M. Lorincz, B. Zimmermann, P. J. Walla, H. Neumann, T. Baubec, H. Urlaub, and W. Fischle. 2016. Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. *Nat Commun* 7:11310.

Hnisz, D., K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp. 2017. A Phase Separation Model for Transcriptional Control. *Cell* 169 (1):13-23.

Holmer, L., and H. J. Worman. 2001. Inner nuclear membrane proteins: functions and targeting. *Cell Mol Life Sci* 58 (12-13):1741-1747.
Hon, G. C., N. Rajagopal, Y. Shen, D. F. McCleary, F. Yue, M. D. Dang, and B. Ren. 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* 45 (10):1198-1206.
Hong, C. K. Y., and B. A. Cohen. 2022. Genomic environments scale the activities of diverse core promoters. *Genome Res* 32 (1):85-96.
Hsieh, T.-H. S., C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, and R. Tjian. 2021. Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. *bioRxiv*:2021.2007.2014.452365.
Hsieh, T. S., C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq. 2020. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell* 78 (3):539-553.e538.
Hsu, P. D., D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31 (9):827-832.
Hu, G., K. Cui, D. Fang, S. Hirose, X. Wang, D. Wangsa, W. Jin, T. Ried, P. Liu, J. Zhu, E. V. Rothenberg, and K. Zhao. 2018. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* 48 (2):227-242.e228.
Huang, D., H. M. Petrykowska, B. F. Miller, L. Elnitski, and I. Ovcharenko. 2019. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res* 29 (4):657-667.
Huang, P., C. A. Keller, B. Giardine, J. D. Grevet, J. O. J. Davies, J. R. Hughes, R. Kurita, Y. Nakamura, R. C. Hardison, and G. A. Blobel. 2017. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev* 31 (16):1704-1713.
Ibn-Salem, J., S. Kohler, M. I. Love, H. R. Chung, N. Huang, M. E. Hurles, M. Haendel, N. L. Washington, D. Smedley, C. J. Mungall, S. E. Lewis, C. E. Ott, S. Bauer, P. N. Schofield, S. Mundlos, M. Spielmann, and P. N. Robinson. 2014. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biology* 15 (9):423.
Isoda, T., A. J. Moore, Z. He, V. Chandra, M. Aida, M. Denholtz, J. Piet van Hamburg, K. M. Fisch, A. N. Chang, S. P. Fahl, D. L. Wiest, and C. Murre. 2017. Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* 171 (1):103-119 e118.
Isono, K., T. A. Endo, M. Ku, D. Yamada, R. Suzuki, J. Sharif, T. Ishikura, T. Toyoda, B. E. Bernstein, and H. Koseki. 2013. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* 26 (6):565-577.
Jerkovic, I., and G. Cavalli. 2021. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* 22 (8):511-528.
Jindal, G. A., and E. K. Farley. 2021. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev Cell* 56 (5):575-587.
Jing, H., C. R. Vakoc, L. Ying, S. Mandat, H. Wang, X. Zheng, and G. A. Blobel. 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol Cell* 29 (2):232-242.
Juhling, F., H. Kretzmer, S. H. Bernhart, C. Otto, P. F. Stadler, and S. Hoffmann. 2016. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res* 26 (2):256-262.
Kass, E. M., and M. Jasin. 2010. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett* 584 (17):3703-3708.
Kaya-Okur, H. S., S. J. Wu, C. A. Codomo, E. S. Pledger, T. D. Bryson, J. G. Henikoff, K. Ahmad, and S. Henikoff. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10 (1):1930.

Kechin, A., U. Boyarskikh, A. Kel, and M. Filipenko. 2017. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol* 24 (11):1138-1143.

Kellner, W. A., K. Van Bortle, L. Li, E. Ramos, N. Takenaka, and V. G. Corces. 2013. Distinct isoforms of the Drosophila Brd4 homologue are present at enhancers, promoters and insulator sites. *Nucleic Acids Res* 41 (20):9274-9283.

Kempfer, R., and A. Pombo. 2020. Methods for mapping 3D chromosome architecture. *Nat Rev Genet* 21 (4):207-226.

Kilic, S., S. Felekyan, O. Doroshenko, I. Boichenko, M. Dimura, H. Vardanyan, L. C. Bryan, G. Arya, C. A. M. Seidel, and B. Fierz. 2018. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nat Commun* 9 (1):235.

Kim, J. D., C. Faulk, and J. Kim. 2007. Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res* 35 (10):3442-3452.

Kim, J. D., H. Kim, M. B. Ekram, S. Yu, C. Faulk, and J. Kim. 2011. Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. *Hum Mol Genet* 20 (7):1353-1362.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, and M. E. Greenberg. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465 (7295):182-187.

Kim, Y., Z. Shi, H. Zhang, I. J. Finkelstein, and H. Yu. 2019. Human cohesin compacts DNA by loop extrusion. *Science* 366 (6471):1345-1349.

Kind, J., L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey, M. Bienko, Y. Zhan, B. Lajoie, C. A. de Graaf, M. Amendola, G. Fudenberg, M. Imakaev, L. A. Mirny, K. Jalink, J. Dekker, A. van Oudenaarden, and B. van Steensel. 2015. Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* 163 (1):134-147.

Klosen, P., S. Lapmanee, C. Schuster, B. Guardiola, D. Hicks, P. Pevet, and M. P. Felder-Schmittbuhl. 2019. MT1 and MT2 melatonin receptors are expressed in nonoverlapping neuronal populations. *J Pineal Res* 67 (1):e12575.

Knight, P. A., and D. Ruiz. 2012. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis* 33 (3):1029-1047.

Kraft, K., S. Geuer, A. J. Will, W. L. Chan, C. Paliou, M. Borschiwer, I. Harabula, L. Wittler, M. Franke, D. M. Ibrahim, B. K. Kragesteen, M. Spielmann, S. Mundlos, D. G. Lupianez, and G. Andrey. 2015. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell reports*.

Kragesteen, B. K., M. Spielmann, C. Paliou, V. Heinrich, R. Schopflin, A. Esposito, C. Annunziatella, S. Bianco, A. M. Chiariello, I. Jerkovic, I. Harabula, P. Guckelberger, M. Pechstein, L. Wittler, W. L. Chan, M. Franke, D. G. Lupianez, K. Kraft, B. Timmermann, M. Vingron, A. Visel, M. Nicodemi, S. Mundlos, and G. Andrey. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* 50 (10):1463-1473.

Krefting, J., M. A. Andrade-Navarro, and J. Ibn-Salem. 2018. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 16 (1):87.

Kremer, K., and G. S. Grest. 1990. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics* 92 (8):5057-5086.

Kueng, S., B. Hegemann, B. H. Peters, J. J. Lipp, A. Schleiffer, K. Mechtler, and J. M. Peters. 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127 (5):955-967.

Kumaran, R. I., and D. L. Spector. 2008. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *Journal of Cell Biology* 180 (1):51-65.

Kvon, E. Z. 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* 106 (3):185-192.

Kvon, E. Z., O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, B. J. Mannion, V. Tissieres, C. S. Pickle, I. Plajzer-Frick, E. A. Lee, M. Kato, T. H. Garvin, J. A. Akiyama, V. Afzal, J. Lopez-Rios, E. M. Rubin, D. E. Dickel, L. A. Pennacchio, and A. Visel. 2016.

Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* 167 (3):633-642 e611.
Kwon, D., D. Mucci, K. K. Langlais, J. L. Americo, S. K. DeVido, Y. Cheng, and J. A. Kassis. 2009. Enhancer-promoter communication at the Drosophila engrailed locus. *Development* 136 (18):3067-3075.
Kyrchanova, O., and P. Georgiev. 2021. Mechanisms of Enhancer-Promoter Interactions in Higher Eukaryotes. *Int J Mol Sci* 22 (2).
Ladurner, A. G., C. Inouye, R. Jain, and R. Tjian. 2003. Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. *Mol Cell* 11 (2):365-376.
Lafontaine, D. L. J., J. A. Riback, R. Bascetin, and C. P. Brangwynne. 2021. The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Biol* 22 (3):165-182.
Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9 (4):357-359.
Laugsch, M., M. Bartusel, R. Rehimi, H. Alirzayeva, A. Karaolidou, G. Crispatzu, P. Zentis, M. Nikolic, T. Bleckwehl, P. Kolovos, W. F. J. van Ijcken, T. Saric, K. Koehler, P. Frommolt, K. Lachlan, J. Baptista, and A. Rada-Iglesias. 2019. Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* 24 (5):736-752 e712.
Lazar, N. H., K. A. Nevonen, B. O'Connell, C. McCann, R. J. O'Neill, R. E. Green, T. J. Meyer, M. Okhovat, and L. Carbone. 2018. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res* 28 (7):983-997.
Le Dily, F., D. Bau, A. Pohl, G. P. Vicent, F. Serra, D. Soronellas, G. Castellano, R. H. Wright, C. Ballare, G. Filion, M. A. Marti-Renom, and M. Beato. 2014. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* 28 (19):2151-2162.
Leemans, C., M. C. H. van der Zwalm, L. Brueckner, F. Comoglio, T. van Schaik, L. Pagie, J. van Arensbergen, and B. van Steensel. 2019. Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell* 177 (4):852-864 e814.
Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. 2003a. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics* 12 (14):1725-1735.
———. 2003b. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12 (14):1725-1735.
Levine, M. 2010. Transcriptional enhancers in animal development and evolution. *Curr Biol* 20 (17):R754-763.
Li, F., D. Wang, R. Song, C. Cao, Z. Zhang, Y. Wang, X. Li, J. Huang, Q. Liu, N. Hou, B. Xu, X. Li, X. Gao, Y. Jia, J. Zhao, and Y. Wang. 2020. The asynchronous establishment of chromatin 3D architecture between in vitro fertilized and uniparental preimplantation pig embryos. *Genome Biology* 21 (1):203.
Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26 (5):589-595.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079.
Li, P., L. Marshall, G. Oh, J. L. Jakubowski, D. Groot, Y. He, T. Wang, A. Petronis, and V. Labrie. 2019. Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nat Commun* 10 (1):2246.
Li, X., and M. Noll. 1994. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. *EMBO J* 13 (2):400-406.
Liao, Y., X. Zhang, M. Chakraborty, and J. J. Emerson. 2021. Topologically associating domains and their role in the evolution of genome structure and function in. *Genome Res* 31 (3):397-410.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326 (5950):289-293.

Liu, G., Q. Lin, S. Jin, and C. Gao. 2022. The CRISPR-Cas toolbox and gene editing technologies. *Mol Cell* 82 (2):333-347.

Liu, M., Y. Lu, B. Yang, Y. Chen, J. S. D. Radda, M. Hu, S. G. Katz, and S. Wang. 2020. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat Commun* 11 (1):2907.

Liu, M., B. Yang, M. Hu, J. S. D. Radda, Y. Chen, S. Jin, Y. Cheng, and S. Wang. 2021a. Chromatin tracing and multiplexed imaging of nucleome architectures (MINA) and RNAs in single mammalian cells and tissue. *Nat Protoc* 16 (5):2667-2697.

Liu, N. Q., M. Maresca, T. van den Brand, L. Braccioli, M. Schijns, H. Teunissen, B. G. Bruneau, E. P. Nora, and E. de Wit. 2021b. WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat Genet* 53 (1):100-109.

Lobe, C. G., K. E. Koop, W. Kreppner, H. Lomeli, M. Gertsenstein, and A. Nagy. 1999. Z/AP, a double reporter for cre-mediated recombination. *Dev Biol* 208 (2):281-292.

Long, H. K., H. W. King, R. K. Patient, D. T. Odom, and R. J. Klose. 2016a. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res* 44 (14):6693-6706.

Long, H. K., S. L. Prescott, and J. Wysocka. 2016b. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167 (5):1170-1187.

Long, H. S., S. Greenaway, G. Powell, A. M. Mallon, C. M. Lindgren, and M. M. Simon. 2022. Making sense of the linear genome, gene function and TADs. *Epigenetics Chromatin* 15 (1):4.

Long, J., M. Li, Q. Ren, C. Zhang, J. Fan, Y. Duan, J. Chen, B. Li, and L. Deng. 2012. Phylogenetic and molecular evolution of the ADAM (A Disintegrin And Metalloprotease) gene family from Xenopus tropicalis, to Mus musculus, Rattus norvegicus, and Homo sapiens. *Gene* 507 (1):36-43.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15 (12):550.

Lu, H., D. Yu, A. S. Hansen, S. Ganguly, R. Liu, A. Heckert, X. Darzacq, and Q. Zhou. 2018. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 558 (7709):318-323.

Lund, E., A. R. Oldenburg, E. Delbarre, C. T. Freberg, I. Duband-Goulet, R. Eskeland, B. Buendia, and P. Collas. 2013. Lamin A/C-promoter interactions specify chromatin state-dependent transcription outcomes. *Genome Research* 23 (10):1580-1589.

Luo, C., P. Hajkova, and J. R. Ecker. 2018. Dynamic DNA methylation: In the right place at the right time. *Science* 361 (6409):1336-1340.

Luo, X., Y. Liu, D. Dang, T. Hu, Y. Hou, X. Meng, F. Zhang, T. Li, C. Wang, M. Li, H. Wu, Q. Shen, Y. Hu, X. Zeng, X. He, L. Yan, S. Zhang, C. Li, and B. Su. 2021. 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis. *Cell* 184 (3):723-740.e721.

Lupianez, D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161 (5):1012-1025.

López-Rivera, F., O. K. Foster Rhoades, B. J. Vincent, E. C. G. Pym, M. D. J. Bragdon, J. Estrada, A. H. DePace, and Z. Wunderlich. 2020. A Mutation in the. *G3 (Bethesda)* 10 (12):4473-4482.

Mach, P., P. I. Kos, Y. Zhan, J. Cramard, S. Gaudin, J. Tünnermann, E. Marchi, J. Eglinger, J. Zuin, M. Kryzhanovska, S. Smallwood, L. Gelman, G. Roth, E. P. Nora, G. Tiana, and

L. Giorgetti. 2022. Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* 54 (12):1907-1918.
Machida, S., Y. Takizawa, M. Ishimaru, Y. Sugita, S. Sekine, J. I. Nakayama, M. Wolf, and H. Kurumizaka. 2018. Structural Basis of Heterochromatin Formation by Human HP1. *Mol Cell* 69 (3):385-397.e388.
Marinic, M., T. Aktas, S. Ruf, and F. Spitz. 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* 24 (5):530-542.
Marsh, B., and R. Blelloch. 2020. Single nuclei RNA-seq of mouse placental labyrinth development. *Elife* 9.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 17 (1):3.
Martinez-Ara, M., F. Comoglio, J. van Arensbergen, and B. van Steensel. 2021. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *bioRxiv*:2021.2010.2021.465269.
———. 2022. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol Cell* 82 (13):2519-2531.e2516.
Maslova, A., and A. Krasikova. 2021. FISH Going Meso-Scale: A Microscopic Search for Chromatin Domains. *Front Cell Dev Biol* 9:753097.
Masui, S., S. Ohtsuka, R. Yagi, K. Takahashi, M. S. H. Ko, and H. Niwa. 2008. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *Bmc Developmental Biology* 8.
Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A. Stamatoyannopoulos. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337 (6099):1190-1195.
McArthur, E., and J. A. Capra. 2021. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* 108 (2):269-283.
McArthur, E., D. C. Rinker, E. N. Gilbertson, G. Fudenberg, M. Pittman, K. Keough, K. S. Pollard, and J. A. Capra. 2022. Reconstructing the 3D genome organization of Neanderthals reveals that chromatin folding shaped phenotypic and sequence divergence. *bioRxiv*:2022.2002.2007.479462.
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20 (9):1297-1303.
Merkenschlager, M., and E. P. Nora. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet* 17:17-43.
Merli, C., D. E. Bergstrom, J. A. Cygan, and R. K. Blackman. 1996. Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev* 10 (10):1260-1270.
Mirny, L. A., M. Imakaev, and N. Abdennur. 2019. Two major mechanisms of chromosome organization. *Curr Opin Cell Biol* 58:142-152.
Mishal, R., and J. P. Luna-Arias. 2022. Role of the TATA-box binding protein (TBP) and associated family members in transcription regulation. *Gene* 833:146581.
Montavon, T., N. Soshnikova, B. Mascrez, E. Joye, L. Thevenet, E. Splinter, W. de Laat, F. Spitz, and D. Duboule. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147 (5):1132-1145.
Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9 (22):6047-6068.
Muro, E. M., J. Ibn-Salem, and M. A. Andrade-Navarro. 2019. The distributions of protein coding genes within chromatin domains in relation to human disease. *Epigenetics Chromatin* 12 (1):72.

Nagano, T., Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502 (7469):59-64.

Nagano, T., Y. Lubling, E. Yaffe, S. W. Wingett, W. Dean, A. Tanay, and P. Fraser. 2015. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc* 10 (12):1986-2003.

Nagulapalli, M., S. Maji, N. Dwivedi, P. Dahiya, and J. K. Thakur. 2016. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res* 44 (4):1591-1612.

Neems, D. S., A. G. Garza-Gongora, E. D. Smith, and S. T. Kosak. 2016. Topologically associated domains enriched for lineage-specific genes reveal expression-dependent nuclear topologies during myogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 113 (12):E1691-1700.

Nguyen, H. Q., S. Chattoraj, D. Castillo, S. C. Nguyen, G. Nir, A. Lioutas, E. A. Hershberg, N. M. C. Martins, P. L. Reginato, M. Hannan, B. J. Beliveau, G. M. Church, E. R. Daugharthy, M. A. Marti-Renom, and C. T. Wu. 2020. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat Methods* 17 (8):822-832.

Nicodemi, M., and A. Prisco. 2009. Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophys J* 96 (6):2168-2177.

Nir, G., I. Farabella, C. Perez Estrada, C. G. Ebeling, B. J. Beliveau, H. M. Sasaki, S. H. Lee, S. C. Nguyen, R. B. McCole, S. Chattoraj, J. Erceg, J. AlHaj Abed, N. M. C. Martins, H. Q. Nguyen, M. A. Hannan, S. Russell, N. C. Durand, S. S. P. Rao, J. Y. Kishi, P. Soler-Vila, M. Di Pierro, J. N. Onuchic, S. P. Callahan, J. M. Schreiner, J. A. Stuckey, P. Yin, E. L. Aiden, M. A. Marti-Renom, and C. T. Wu. 2018. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet* 14 (12):e1007872.

Niu, L., W. Shen, Z. Shi, Y. Tan, N. He, J. Wan, J. Sun, Y. Zhang, Y. Huang, W. Wang, C. Fang, J. Li, P. Zheng, E. Cheung, Y. Chen, L. Li, and C. Hou. 2021. Three-dimensional folding dynamics of the Xenopus tropicalis genome. *Nature Genetics* 53 (7):1075-1087.

Nora, E. P., A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169 (5):930-944 e922.

Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, and E. Heard. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485 (7398):381-385.

Northcott, P. A., C. Lee, T. Zichner, A. M. Stutz, S. Erkek, D. Kawauchi, D. J. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. Jones, M. Kool, M. Remke, F. M. Cavalli, S. Zuyderduyn, G. D. Bader, S. VandenBerg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jager, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H. J. Warnatz, T. Risch, M. L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turanyi, P. Hauser, E. Sanden, A. Darabi, P. Siesjo, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya, and S. M. Pfister. 2014. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511 (7510):428-434.

Ogbourne, S., and T. M. Antalis. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* 331 (Pt 1) (Pt 1):1-14.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99 (3):247-257.

Olova, N., F. Krueger, S. Andrews, D. Oxley, R. V. Berrens, M. R. Branco, and W. Reik. 2018. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol* 19 (1):33.

Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng, and T. H. Bestor. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448 (7154):714-717.

Ostapińska, K., B. Styka, and M. Lejman. 2022. Insight into the Molecular Basis Underlying Chromothripsis. *Int J Mol Sci* 23 (6).

Osterwalder, M., I. Barozzi, V. Tissieres, Y. Fukuda-Yuzawa, B. J. Mannion, S. Y. Afzal, E. A. Lee, Y. Zhu, I. Plajzer-Frick, C. S. Pickle, M. Kato, T. H. Garvin, Q. T. Pham, A. N. Harrington, J. A. Akiyama, V. Afzal, J. Lopez-Rios, D. E. Dickel, A. Visel, and L. A. Pennacchio. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554 (7691):239-243.

Oudelaar, A. M., C. L. Harrold, L. L. P. Hanssen, J. M. Telenius, D. R. Higgs, and J. R. Hughes. 2019. A revised model for promoter competition based on multi-way chromatin interactions at the α-globin locus. *Nat Commun* 10 (1):5412.

Pachano, T., E. Haro, and A. Rada-Iglesias. 2022. Enhancer-gene specificity in development and disease. *Development* 149 (11).

Palstra, R. J., and F. Grosveld. 2012. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Front Genet* 3:195.

Palstra, R. J., B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics* 35 (2):190-194.

Panne, D. 2008. The enhanceosome. *Curr Opin Struct Biol* 18 (2):236-242.

Peng, Z., Y. Gong, and X. Liang. 2021. Role of FAT1 in health and disease. *Oncol Lett* 21 (5):398.

Pennacchio, L. A., W. Bickmore, A. Dean, M. A. Nobrega, and G. Bejerano. 2013. Enhancers: five essential questions. *Nat Rev Genet* 14 (4):288-295.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels, and B. van Steensel. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Molecular Cell* 38 (4):603-613.

Peric-Hupkes, D., and B. van Steensel. 2010. Role of the nuclear lamina in genome organization and gene expression. *Cold Spring Harb Symp Quant Biol* 75:517-524.

Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod, and B. van Steensel. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet* 38 (9):1005-1014.

Pijuan-Sala, B., J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. V. Tyser, D. L. L. Ho, W. Reik, S. Srinivas, B. D. Simons, J. Nichols, J. C. Marioni, and B. Gottgens. 2019. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566 (7745):490-+.

Plimpton, S. 1995. Fast parallel algorithms for short-range Molecular Dynamics. Journal of Computational Physics.

Poleshko, A., P. P. Shah, M. Gupta, A. Babu, M. P. Morley, L. J. Manderfield, J. L. Ifkovits, D. Calderon, H. Aghajanian, J. E. Sierra-Pagan, Z. Sun, Q. Wang, L. Li, N. C. Dubois, E. E. Morrisey, M. A. Lazar, C. L. Smith, J. A. Epstein, and R. Jain. 2017. Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* 171 (3):573-587 e514.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6):841-842.

Ramirez, F., D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dundar, and T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44 (W1):W160-165.

Ramisch, A., V. Heinrich, L. V. Glaser, A. Fuchs, X. Yang, P. Benner, R. Schopflin, N. Li, S. Kinkley, A. Romer-Hillmann, J. Longinotto, S. Heyne, B. Czepukojc, S. M. Kessler, A. K. Kiemer, C. Cadenas, L. Arrigoni, N. Gasparoni, T. Manke, T. Pap, J. A. Pospisilik,

J. Hengstler, J. Walter, S. H. Meijsing, H. R. Chung, and M. Vingron. 2019. CRUP: a comprehensive framework to predict condition-specific regulatory units. *Genome Biol* 20 (1):227.

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159 (7):1665-1680.

Rao, S. S. P., S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, and E. L. Aiden. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* 171 (2):305-320 e324.

Real, F. M., S. A. Haas, P. Franchini, P. Xiong, O. Simakov, H. Kuhl, R. Schopflin, D. Heller, M. H. Moeinzadeh, V. Heinrich, T. Krannich, A. Bressin, M. F. Hartmann, S. A. Wudy, D. K. N. Dechmann, A. Hurtado, F. J. Barrionuevo, M. Schindler, I. Harabula, M. Osterwalder, M. Hiller, L. Wittler, A. Visel, B. Timmermann, A. Meyer, M. Vingron, R. Jimenez, S. Mundlos, and D. G. Lupianez. 2020. The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science* 370 (6513):208-214.

Reddy, K. L., J. M. Zullo, E. Bertolino, and H. Singh. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452 (7184):243-247.

Rennie, S., M. Dalby, L. van Duin, and R. Andersson. 2018. Transcriptional decomposition reveals active chromatin architectures and cell specific regulatory interactions. *Nat Commun* 9 (1):487.

Renschler, G., G. Richard, C. I. K. Valsecchi, S. Toscano, L. Arrigoni, F. Ramírez, and A. Akhtar. 2019. Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes Dev* 33 (21-22):1591-1612.

Reynolds, N., A. O'Shaughnessy, and B. Hendrich. 2013. Transcriptional repressors: multifaceted regulators of gene expression. *Development* 140 (3):505-512.

Rhodes, J. D. P., A. Feldmann, B. Hernández-Rodríguez, N. Díaz, J. M. Brown, N. A. Fursova, N. P. Blackledge, P. Prathapan, P. Dobrinic, M. K. Huseyin, A. Szczurek, K. Kruse, K. A. Nasmyth, V. J. Buckle, J. M. Vaquerizas, and R. J. Klose. 2020. Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Rep* 30 (3):820-835.e810.

Rice, M. C., and S. J. O'Brien. 1980. Genetic variance of laboratory outbred Swiss mice. *Nature* 283 (5743):157-161.

Richter, W. F., S. Nayak, J. Iwasa, and D. J. Taatjes. 2022. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nat Rev Mol Cell Biol* 23 (11):732-749.

Ringel, A. R., Q. Szabo, A. M. Chiariello, K. Chudzik, R. Schöpflin, P. Rothe, A. L. Mattei, T. Zehnder, D. Harnett, V. Laupert, S. Bianco, S. Hetzel, J. Glaser, M. H. Q. Phan, M. Schindler, D. M. Ibrahim, C. Paliou, A. Esposito, C. A. Prada-Medina, S. A. Haas, P. Giere, M. Vingron, L. Wittler, A. Meissner, M. Nicodemi, G. Cavalli, F. Bantignies, S. Mundlos, and M. I. Robson. 2022. Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes. *Cell* 185 (20):3689-3704.e3621.

Ringel, A. R., Q. Szabo, A. M. Chiariello, K. Chudzik, R. Schöpflin, P. Rothe, A. L. Mattei, T. Zehnder, D. Harnett, V. Laupert, S. Bianco, S. Hetzel, M. Phan, M. Schindler, D. Ibrahim, C. Paliou, A. Esposito, C. A. Prada-Medina, S. Haas, P. Giere, M. Vingron, L. Wittler, A. Meissner, M. Nicodemi, G. Cavalli, F. Bantignies, S. Mundlos, and M. I. Robson. 2021. Promoter repression and 3D-restructuring resolves divergent developmental gene expression in TADs. *bioRxiv*:2021.2010.2008.463672.

Rippe, K. 2022. Liquid-Liquid Phase Separation in Chromatin. *Cold Spring Harb Perspect Biol* 14 (2).

Robson, M. I., J. I. de Las Heras, R. Czapiewski, P. Le Thanh, D. G. Booth, D. A. Kelly, S. Webb, A. R. W. Kerr, and E. C. Schirmer. 2016. Tissue-Specific Gene Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis. *Mol Cell* 62 (6):834-847.

Robson, M. I., J. I. de Las Heras, R. Czapiewski, A. Sivakumar, A. R. W. Kerr, and E. C. Schirmer. 2017. Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. *Genome Res* 27 (7):1126-1138.

Robson, M. I., A. R. Ringel, and S. Mundlos. 2019. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* 74 (6):1110-1122.

Robson, M. I., and E. C. Schirmer. 2016. The Application of DamID to Identify Peripheral Gene Sequences in Differentiated and Primary Cells. *Methods Mol Biol* 1411:359-386.

Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. Sequence complexity of disordered protein. *Proteins* 42 (1):38-48.

Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361 (6400).

Sadeqzadeh, E., C. E. de Bock, and R. F. Thorne. 2014. Sleeping giants: emerging roles for the fat cadherins in health and disease. *Med Res Rev* 34 (1):190-221.

Sanborn, A. L., S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. L. Aiden. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112 (47):E6456-6465.

Santiago-Algarra, D., L. T. M. Dao, L. Pradel, A. España, and S. Spicuglia. 2017. Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res* 6:939.

Sawh, A. N., and S. E. Mango. 2020. Multiplexed Sequential DNA FISH in. *STAR Protoc* 1 (3):100107.

Schermelleh, L., P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, H. Leonhardt, and J. W. Sedat. 2008. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* 320 (5881):1332-1336.

Schmidl, C., A. F. Rendeiro, N. C. Sheffield, and C. Bock. 2015. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* 12 (10):963-965.

Schmidt, D., M. D. Wilson, B. Ballester, P. C. Schwalie, G. D. Brown, A. Marshall, C. Kutter, S. Watt, C. P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek, and D. T. Odom. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328 (5981):1036-1040.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132 (5):887-898.

Schultz, M. D., Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, and J. R. Ecker. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523 (7559):212-216.

Schwarzer, W., N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, H. H. C, L. Mirny, and F. Spitz. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551 (7678):51-56.

Schöpflin, R., U. S. Melo, H. Moeinzadeh, D. Heller, V. Laupert, J. Hertzberg, M. Holtgrewe, N. Alavi, M. K. Klever, J. Jungnitsch, E. Comak, S. Türkmen, D. Horn, Y. Duffourd, L. Faivre, P. Callier, D. Sanlaville, O. Zuffardi, R. Tenconi, N. E. Kurtas, S. Giglio, B. Prager, A. Latos-Bielenska, I. Vogel, M. Bugge, N. Tommerup, M. Spielmann, A. Vitobello, V. M. Kalscheuer, M. Vingron, and S. Mundlos. 2022. Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. *Nat Commun* 13 (1):6470.

Scully, R., A. Panday, R. Elango, and N. A. Willis. 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* 20 (11):698-714.
Serebreni, L., L.-M. Pleyer, V. Haberle, O. Hendy, A. Vlasova, V. Loubiere, F. Nemčko, K. Bergauer, E. Roitinger, K. Mechtler, and A. Stark. 2022. Functionally distinct promoter classes initiate transcription via different mechanisms reflected in focused versus dispersed initiation patterns. *bioRxiv*:2022.2009.2030.509643.
Shandilya, J., and S. G. Roberts. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* 1819 (5):391-400.
Shen, Y., F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, and B. Ren. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488 (7409):116-120.
Shi, B., W. Li, Y. Song, Z. Wang, R. Ju, A. Ulman, J. Hu, F. Palomba, Y. Zhao, J. P. Le, W. Jarrard, D. Dimoff, M. A. Digman, E. Gratton, C. Zang, and H. Jiang. 2021. UTX condensation underlies its tumour-suppressive activity. *Nature* 597 (7878):726-731.
Shima, Y., K. Sugino, C. M. Hempel, M. Shima, P. Taneja, J. B. Bullis, S. Mehta, C. Lois, and S. B. Nelson. 2016. A Mammalian enhancer trap resource for discovering and manipulating neuronal cell types. *Elife* 5:e13503.
Shin, H., Y. Shi, C. Dai, H. Tjong, K. Gong, F. Alber, and X. J. Zhou. 2016a. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res* 44 (7):e70.
Shin, H. Y., M. Willi, K. HyunYoo, X. Zeng, C. Wang, G. Metser, and L. Hennighausen. 2016b. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet* 48 (8):904-911.
Shlyueva, D., G. Stampfel, and A. Stark. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15 (4):272-286.
Sima, J., A. Chakraborty, V. Dileep, M. Michalski, K. N. Klein, N. P. Holcomb, J. L. Turner, M. T. Paulsen, J. C. Rivera-Mulia, C. Trevilla-Garcia, D. A. Bartlett, P. A. Zhao, B. K. Washburn, E. P. Nora, K. Kraft, S. Mundlos, B. G. Bruneau, M. Ljungman, P. Fraser, F. Ay, and D. M. Gilbert. 2019. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell* 176 (4):816-+.
Singer, Z. S., J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell* 55 (2):319-331.
Singh, P. B., and A. G. Newman. 2020. On the relations of phase separation and Hi-C maps to epigenetics. *R Soc Open Sci* 7 (2):191976.
Solomon, M. J., P. L. Larsen, and A. Varshavsky. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53 (6):937-947.
Solovei, I., M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, J. Guck, and B. Joffe. 2009. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137 (2):356-368.
Solovei, I., A. S. Wang, K. Thanisch, C. S. Schmidt, S. Krebs, M. Zwerger, T. V. Cohen, D. Devys, R. Foisner, L. Peichl, H. Herrmann, H. Blum, D. Engelkamp, C. L. Stewart, H. Leonhardt, and B. Joffe. 2013. LBR and Lamin A/C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation. *Cell* 152 (3):584-598.
Soshnikova, N., and D. Duboule. 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science* 324 (5932):1320-1323.
Soutourina, J. 2018. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* 19 (4):262-274.
Spielmann, M., D. G. Lupianez, and S. Mundlos. 2018. Structural variation in the 3D genome. *Nat Rev Genet* 19 (7):453-467.
Spitz, F., F. Gonzalez, and D. Duboule. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 113 (3):405-417.
Su, J. H., P. Zheng, S. S. Kinrot, B. Bintu, and X. Zhuang. 2020. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182 (6):1641-1659.e1626.

Subramanian, K., M. Weigert, O. Borsch, H. Petzold, A. Garcia-Ulloa, E. W. Myers, M. Ader, I. Solovei, and M. Kreysing. 2019. Rod nuclear architecture determines contrast transmission of the retina and behavioral sensitivity in mice. *Elife* 8.

Sun, D., Y. Xi, B. Rodriguez, H. J. Park, P. Tong, M. Meong, M. A. Goodell, and W. Li. 2014. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol* 15 (2):R38.

Sun, M. A., G. Wolf, Y. Wang, A. D. Senft, S. Ralls, J. Jin, C. E. Dunn-Fletcher, L. J. Muglia, and T. S. Macfarlan. 2021. Endogenous Retroviruses Drive Lineage-Specific Regulatory Evolution across Primate and Rodent Placentae. *Mol Biol Evol* 38 (11):4992-5004.

Symmons, O., V. V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, and F. Spitz. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24 (3):390-400.

Szabo, Q., A. Donjon, I. Jerkovic, G. L. Papadopoulos, T. Cheutin, B. Bonev, E. P. Nora, B. G. Bruneau, F. Bantignies, and G. Cavalli. 2020. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* 52 (11):1151-1157.

Szabo, Q., D. Jost, J. M. Chang, D. I. Cattoni, G. L. Papadopoulos, B. Bonev, T. Sexton, J. Gurgo, C. Jacquier, M. Nollmann, F. Bantignies, and G. Cavalli. 2018. TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Sci Adv* 4 (2):eaar8082.

Tak, Y. G., and P. J. Farnham. 2015. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* 8:57.

Takebayashi, S., V. Dileep, T. Ryba, J. H. Dennis, and D. M. Gilbert. 2012. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proceedings of the National Academy of Sciences of the United States of America* 109 (31):12574-12579.

Tedeschi, A., G. Wutz, S. Huet, M. Jaritz, A. Wuensche, E. Schirghuber, I. F. Davidson, W. Tang, D. A. Cisneros, V. Bhaskara, T. Nishiyama, A. Vaziri, A. Wutz, J. Ellenberg, and J. M. Peters. 2013. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* 501 (7468):564-568.

Therizols, P., R. S. Illingworth, C. Courilleau, S. Boyle, A. J. Wood, and W. A. Bickmore. 2014. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346 (6214):1238-1242.

Thiecke, M. J., G. Wutz, M. Muhar, W. Tang, S. Bevan, V. Malysheva, R. Stocsits, T. Neumann, J. Zuber, P. Fraser, S. Schoenfelder, J. M. Peters, and M. Spivakov. 2020. Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep* 32 (3):107929.

Thomas, H. F., E. Kotova, S. Jayaram, A. Pilz, M. Romeike, A. Lackner, T. Penz, C. Bock, M. Leeb, F. Halbritter, J. Wysocka, and C. Buecker. 2021. Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol Cell* 81 (5):969-982.e913.

Torosin, N. S., A. Anand, T. R. Golla, W. Cao, and C. E. Ellison. 2020. 3D genome evolution and reorganization in the Drosophila melanogaster species group. *PLoS Genet* 16 (12):e1009229.

Trauernicht, M., M. Martinez-Ara, and B. van Steensel. 2020. Deciphering Gene Regulation Using Massively Parallel Reporter Assays. *Trends Biochem Sci* 45 (1):90-91.

Tucker, K. L., Y. Wang, J. Dausman, and R. Jaenisch. 1997. A transgenic mouse strain expressing four drug-selectable marker genes. *Nucleic Acids Res* 25 (18):3745-3746.

Vakoc, C. R., D. L. Letting, N. Gheldof, T. Sawado, M. A. Bender, M. Groudine, M. J. Weiss, J. Dekker, and G. A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 17 (3):453-462.

van Arensbergen, J., B. van Steensel, and H. J. Bussemaker. 2014. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol* 24 (11):695-702.

van Schaik, T., M. Vos, D. Peric-Hupkes, P. Hn Celie, and B. van Steensel. 2020. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep* 21 (11):e50636.

van Steensel, B., and A. S. Belmont. 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169 (5):780-791.
van Steensel, B., and S. Henikoff. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature biotechnology* 18 (4):424-428.
Vanhille, L., A. Griffon, M. A. Maqbool, J. Zacarias-Cabeza, L. T. Dao, N. Fernandez, B. Ballester, J. C. Andrau, and S. Spicuglia. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6:6905.
Vietri Rudan, M., C. Barrington, S. Henderson, C. Ernst, D. T. Odom, A. Tanay, and S. Hadjur. 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* 10 (8):1297-1309.
Villar, D., C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek, and D. T. Odom. 2015. Enhancer evolution across 20 mammalian species. *Cell* 160 (3):554-566.
Visel, A., S. Minovitsky, I. Dubchak, and L. A. Pennacchio. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35 (Database issue):D88-92.
Visel, A., E. M. Rubin, and L. A. Pennacchio. 2009. Genomic views of distant-acting enhancers. *Nature* 461 (7261):199-205.
Vogel, M. J., D. Peric-Hupkes, and B. van Steensel. 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nature protocols* 2 (6):1467-1478.
Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. *Proteins* 52 (4):573-584.
Wang, S., J. H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C. T. Wu, and X. Zhuang. 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353 (6299):598-602.
Wang, Z., K. Lin, and H. Xiao. 2022. A pan-cancer analysis of the FAT1 in human tumors. *Sci Rep* 12 (1):21598.
Weinberg, D. N., P. Rosenbaum, X. Chen, D. Barrows, C. Horth, M. R. Marunde, I. K. Popova, Z. B. Gillespie, M.-C. Keogh, C. Lu, J. Majewski, and C. D. Allis. 2021. Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. *Nature Genetics* 53 (6):794-800.
Weintraub, A. S., C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, Y. E. Guo, D. Hnisz, R. Jaenisch, J. E. Bradner, N. S. Gray, and R. A. Young. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171 (7):1573-1588 e1528.
Will, A. J., G. Cova, M. Osterwalder, W. L. Chan, L. Wittler, N. Brieske, V. Heinrich, J. P. de Villartay, M. Vingron, E. Klopocki, A. Visel, D. G. Lupianez, and S. Mundlos. 2017. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nature Genetics* 49 (10):1539-1545.
Williamson, I., L. A. Lettice, R. E. Hill, and W. A. Bickmore. 2016. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* 143 (16):2994-3001.
Wingett, S., P. Ewels, M. Furlan-Magaril, T. Nagano, S. Schoenfelder, P. Fraser, and S. Andrews. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* 4:1310.
Winick-Ng, W., A. Kukalev, I. Harabula, L. Z. Redondo, D. Szabo, M. Meijer, L. Serebreni, Y. Zhang, S. Bianco, A. M. Chiariello, I. Irastorza-Azcarate, L. Fiorillo, F. Musella, C. J. Thieme, E. Irani, E. T. Triglia, A. A. Kolodziejczyk, A. Abentung, G. Apostolova, E. J. Paul, V. Franke, R. Kempfer, A. Akalin, S. A. Teichmann, G. Dechant, M. A. Ungless, M. Nicodemi, L. Welch, G. Castelo-Branco, and A. Pombo. 2021a. Cell-type specialization in the brain is encoded by specific long-range chromatin topologies. *bioRxiv*:2020.2004.2002.020990.
Winick-Ng, W., A. Kukalev, I. Harabula, L. Zea-Redondo, D. Szabó, M. Meijer, L. Serebreni, Y. Zhang, S. Bianco, A. M. Chiariello, I. Irastorza-Azcarate, C. J. Thieme, T. M. Sparks,

S. Carvalho, L. Fiorillo, F. Musella, E. Irani, E. Torlai Triglia, A. A. Kolodziejczyk, A. Abentung, G. Apostolova, E. J. Paul, V. Franke, R. Kempfer, A. Akalin, S. A. Teichmann, G. Dechant, M. A. Ungless, M. Nicodemi, L. Welch, G. Castelo-Branco, and A. Pombo. 2021b. Cell-type specialization is encoded by specific chromatin topologies. *Nature* 599 (7886):684-691.

Wittkopp, P. J., and G. Kalay. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* 13 (1):59-69.

Woltering, J. M., D. Noordermeer, M. Leleu, and D. Duboule. 2014. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* 12 (1):e1001773.

Wu, H. J., A. Landshammer, E. K. Stamenova, A. Bolondi, H. Kretzmer, A. Meissner, and F. Michor. 2021. Topological isolation of developmental regulators in mammalian genomes. *Nat Commun* 12 (1):4897.

Wutz, G., C. Varnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S. Schoenfelder, G. Jessberger, M. Muhar, M. J. Hossain, N. Walther, B. Koch, M. Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, and J. M. Peters. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* 36 (24):3573-3599.

Xi, Y., and W. Li. 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10:232.

Xie, L., P. Dong, Y. Qi, T. S. Hsieh, B. P. English, S. Jung, X. Chen, M. De Marzio, R. Casellas, H. Y. Chang, B. Zhang, R. Tjian, and Z. Liu. 2022. BRD2 compartmentalizes the accessible genome. *Nat Genet* 54 (4):481-491.

Xu, Z., D. S. Lee, S. Chandran, V. T. Le, R. Bump, J. Yasis, S. Dallarda, S. Marcotte, B. Clock, N. Haghani, C. Y. Cho, K. C. Akdemir, S. Tyndale, P. A. Futreal, G. McVicker, G. M. Wahl, and J. R. Dixon. 2022. Structural variants drive context-dependent oncogene activation in cancer. *Nature* 612 (7940):564-572.

Yagi, M., M. Kabata, A. Tanaka, T. Ukai, S. Ohta, K. Nakabayashi, M. Shimizu, K. Hata, A. Meissner, T. Yamamoto, and Y. Yamada. 2020. Identification of distinct loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. *Nat Commun* 11 (1):3199.

Yang, H., Y. Luan, T. Liu, H. J. Lee, L. Fang, Y. Wang, X. Wang, B. Zhang, Q. Jin, K. C. Ang, X. Xing, J. Wang, J. Xu, F. Song, I. Sriranga, C. Khunsriraksakul, T. Salameh, D. Li, M. N. K. Choudhary, J. Topczewski, K. Wang, G. S. Gerhard, R. C. Hardison, T. Wang, K. C. Cheng, and F. Yue. 2020. A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* 588 (7837):337-343.

Zabidi, M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, O. Frank, and A. Stark. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518 (7540):556-559.

Zehnder, T., P. Benner, and M. Vingron. 2019. Predicting enhancers in mammalian genomes using supervised hidden Markov models. *BMC Bioinformatics* 20 (1):157.

Zhang, D., P. Huang, M. Sharma, C. A. Keller, B. Giardine, H. Zhang, T. G. Gilgenast, J. E. Phillips-Cremins, R. C. Hardison, and G. A. Blobel. 2020a. Alteration of genome folding via contact domain boundary insertion. *Nat Genet* 52 (10):1076-1087.

Zhang, X., M. Pavlicev, H. N. Jones, and L. J. Muglia. 2020b. Eutherian-Specific Gene TRIML2 Attenuates Inflammation in the Evolution of Placentation. *Mol Biol Evol* 37 (2):507-523.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle, and A. Jeltsch. 2010. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* 38 (13):4246-4253.

Zhang, Y., T. Li, S. Preissl, M. L. Amaral, J. D. Grinstein, E. N. Farah, E. Destici, Y. Qiu, R. Hu, A. Y. Lee, S. Chee, K. Ma, Z. Ye, Q. Zhu, H. Huang, R. Fang, L. Yu, J. C. Izpisua Belmonte, J. Wu, S. M. Evans, N. C. Chi, and B. Ren. 2019. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* 51 (9):1380-1388.

Zhang, Y., Y. X. See, V. Tergaonkar, and M. J. Fullwood. 2022. Long-Distance Repression by Human Silencers: Chromatin Interactions and Phase Separation in Silencers. *Cells* 11 (9).

Ziller, M. J., H. Gu, F. Müller, J. Donaghey, L. T. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, A. Gnirke, and A. Meissner. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500 (7463):477-481.

Zufferey, M., D. Tavernari, E. Oricchio, and G. Ciriello. 2018. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* 19 (1):217.

Zuin, J., G. Roth, Y. Zhan, J. Cramard, J. Redolfi, E. Piskadlo, P. Mach, M. Kryzhanovska, G. Tihanyi, H. Kohler, P. Meister, S. Smallwood, and L. Giorgetti. 2021. Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv*:2021.2004.2022.440891.

ACKNOWLEDGEMENTS

Mein Dank gilt in erster Linie Herrn Prof. Dr. Stefan Mundlos, der mich zu jedem Zeitpunkt unterstützt hat und mir die Möglichkeit gab, Teil einer so wunderbaren Forschungsgruppe zu sein. Ebenso bedanke ich mich bei Herrn Prof. Dr. Sigmar Stricker für die Übernahme des Zweitgutachtens und die Beantwortung meiner Fragen.

Ein besonderer Dank geht an Mike für die Betreuung meiner Arbeit, sein Vertrauen und seine außergewöhnliche Unterstützung. Ich habe viel durch ihn und mit ihm gelernt und unsere gemeinsame Zeit hat mich sehr geprägt. Des Weiteren bedanke ich mich bei Konrad, der mir stets geholfen hat, wenn ich ein weiteres Paar Hände brauchte, sowie bei Patricia, die eine wunderbare Studentin war. Ebenfalls gilt mein Dank Alex, die mir besonders zu Beginn meiner Arbeit unendlich viele Fragen beantwortet hat und immer mit Rat zur Seite stand.

Selbstverständlich gilt mein Dank allen ehemaligen und aktuellen Mitgliedern des Mundlos-Teams: Asita, Alex, Alicia, Andreas, Baiwei, Blanka, Björn, Bjort, Carola, Cesar, Chiara, Christina, Cinzia, Daniel, Fany, Felix, Fiona, Friederike, Gabi, Giulia, Guillaume, Henrike, Ivana, Jana, Janina, Jessy, Josh, Julia, Juliane, Konrad, Lena, Lila, Lion, Magdalena, Martin, Masha, Mike, Mikie, Milan, Mira, Nils, Nino, Norbert, Patricia, Philine, Philipp, Rene, Robert, Rocío, Sala, Tobi, Uirá, Ute, Vanessa, Vera und Verena. Ohne Euch wäre diese Gruppe nicht so dynamisch und bunt. Die Diskussionsbereitschaft und gegenseitige Unterstützung innerhalb der Gruppe ist immer wieder bemerkenswert. In diesem Sinne möchte ich mich speziell bei Alica, Mikie und Josh fürs Korrekturlesen dieser Arbeit bedanken. Ebenso bedanke ich mich bei Asita, Norbert und Ute für die experimentelle und organisatorische Unterstützung des gesamten Teams.

Darüber hinaus bedanke ich mich bei allen Kooperationspartnern für die Zusammenarbeit an gemeinsamen Projekten. Hier möchte ich insbesondere Quentin, Frédéric sowie Giacomo Cavalli und sein gesamtes Team erwähnen, die mir einen produktiven und wissenschaftlich stimulierenden Aufenthalt in ihrer Gruppe ermöglicht haben. Ebenso bedanke ich mich bei Andrea für zahlreiche Besprechungen, die unsere Arbeit vorangebracht haben.

Zuletzt, aber in meinem Herzen an erster Stelle, bedanke ich mich besonders bei meinen Eltern, meinen Brüdern, Lars, Fabian und Jonas, meinem Partner Paul sowie bei meinen Freunden, die mich auch in dieser Zeit immer unterstützt haben.

SCIENTIFIC PUBLICATIONS

Ringel AR, Szabo Q, Chiariello AM, Chudzik K, Schöpflin R, Rothe P, Mattei AL, Zehnder T, Harnett D, Laupert V, Bianco S, Hetzel S, Glaser J, Phan MHQ, Schindler M, Ibrahim DM, Paliou C, Esposito A, Prada-Medina CA, Haas SA, Giere P, Vingron M, Wittler L, Meissner A, Nicodemi M, Cavalli G, Bantignies F, Mundlos S, Robson MI.: ***Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes.*** Cell. 2022 Sep 29;185(20):3689-3704.e21. doi: 10.1016/j.cell.2022.09.006.

Allou L, Balzano S, Magg A, Quinodoz M, Royer-Bertrand B, Schöpflin R, Chan WL, Speck-Martins CE, Carvalho DR, Farage L, Lourenço CM, Albuquerque R, Rajagopal S, Nampoothiri S, Campos-Xavier B, Chiesa C, Niel-Bütschi F, Wittler L, Timmermann B, Spielmann M, Robson MI, **Ringel A**, Heinrich V, Cova G, Andrey G, Prada-Medina CA, Pescini-Gobert R, Unger S, Bonafé L, Grote P, Rivolta C, Mundlos S, Superti-Furga A.***: Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator.*** Nature. 2021 Apr;592(7852):93-98. doi: 10.1038/s41586-021-03208-9.

Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. Robson MI, **Ringel AR**, Mundlos S.**;** Mol Cell. 2019 Jun 20;74(6):1110-1122. doi: 10.1016/j.molcel.2019.05.032.

Ruoff K, Kilic T, Devant J, Koromyslova A, **Ringel A**, Hempelmann A, Geiss C, Graf J, Haas M, Roggenbach I, Hansman G.: ***Structural Basis of Nanobodies Targeting the Prototype Norovirus.*** J Virol. 2019 Mar 5;93(6):e02005-18. doi: 10.1128/JVI.02005-18.

Yien YY*, **Ringel AR***, Paw BH.: ***Mitochondrial transport of protoporphyrinogen IX in erythroid cells.*** Oncotarget. 2015 Aug 28;6(25):20742-3. doi: 10.18632/oncotarget.5124.
* equal contribution

Hill RJ, **Ringel A**, Knuepfer E, Moon RW, Blackman MJ, van Ooij C. ***Regulation and Essentiality of the StAR-related Lipid Transfer (START) Domain-containing Phospholipid Transfer Protein PFA0210c in Malaria Parasites*****.** J Biol Chem. 2016 Nov 11;291(46):24280-24292. doi: 10.1074/jbc.M116.740506.

van Ooij C, Withers-Martinez C, **Ringel A**, Cockcroft S, Haldar K, Blackman MJ. ***Identification of a Plasmodium falciparum phospholipid transfer protein*****.** J Biol Chem. 2013 Nov 1;288(44):31971-83. doi: 10.1074/jbc.M113.474189.