Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Massimo Petretich, M.Sc.
born in: Latisana, Italy
Oral-examination: 20/09/2016

Importance of Chromosomal Architecture to Organize Promoter-Enhancer Long-Range Interactions in *c-Myc locus*

Referees: Prof. Dr. Detlev Arendt
Prof. Dr. Ingrid Lohmann

To Herbert Steinbeisser (1958–2014)
Professor at Heidelberg University

Table of Contents

1 Summary

1.1 English Summary

In metazoans the regulation of transcription commonly relies on multiple types of regulatory elements. Often these elements are separated from their target promoters by great genomic distance and can even have other genes in between. Current models propose the chromosomes to be folded following a specific topology in order to allow these long-range interactions to occur. In addition to the *locus* topology, chromatin environment and DNA accessibility play an important role to determine the activity of regulatory elements.

Thanks to genome-wide studies these layers of regulation have been shown to correlate with gene expression to different extents, however it is still not clear how chromatin, DNA topology and genetic components integrate in order to carry out this process. With this project we aim to understand basic mechanisms organizing promoter–regulatory elements communication in a defined model *locus*.

A previous project from the laboratory used chromosomal deletions and *in vivo* reporter assays to identify and characterize regulatory elements in *c-Myc locus*. This study, performed by Veli Uslu showed that during mouse embryonic development, at 11.5 days of gestation, different enhancers coordinate *c-Myc* transcription in the facial tissues and embryonic liver. We analyzed the distribution inside *c-Myc locus* of histone modifications enriched at enhancer sites, restricting to few smaller putative regions the genetic position of facial and liver enhancers. We also assessed the occupancy of the architectural protein CTCF, enriched at sites involved in long-range interactions and on insulators, finding both tissue-invariant and tissue-specific binding sites at different parts of the *locus*.

To study if the topology of the *locus* influences enhancer promoter communication and how it does this, we created three lines of mice carrying different chromosomal inversions. Each of which reshuffles the relative positions of the different genetic elements: *c-Myc* promoter, enhancers and CTCF binding sites.

The observed effects of the genetic engineering include significant reductions of *c-Myc* expression in the facial tissues and redirection of regulatory activity to other

genes. In the liver we saw instead milder and non-significant effects of the inversions on *c-Myc* regulation.

We investigated the occupancy of CTCF and the Cohesin complex subunit RAD21 (another architectural protein) in the facial mesenchyme and liver, on the engineered lines, detecting in *c-Myc locus* just some minor changes compared to the Wt.

Our data suggest that the architecture of the *loci* may have a functional role in organizing promoters-regulatory elements interactions, and, if altered, can lead to genes misexpression. In case architectural proteins are involved in this process, our data suggests that their binding to the genome, not significantly altered by genetic reshuffling, may be mostly determined by the DNA sequence (genetics). On the other hand, it is possible that the long-range interactions engaged by these proteins may be influenced more by their working context.

1.2 Deutsche Zusammenfassung

Bei vielzelligen Tieren stützt sich die Regulierung der Transkription häufig auf mehreren regulativen Elementen. Oft besteht zwischen diesen Elementen und ihren Zielpromotoren ein großer genetischer Abstand und es können sich auch andere Gene zwischen ihnen befinden. Die aktuellen Modelle sehen vor, dass die Chromosomen nach einer bestimmten Topologie gefaltet sind, um das Entstehen dieser *long-range interactions* zu ermöglichen. Zusätzlich zu der *Locus* Topologie spielen die Chromatin-Umgebung und die DNA Zugänglichkeit eine wichtige Rolle, um die Aktivität der regulativen Elemente zu bestimmen.

Dank genomweiter Studien wurde bewiesen, dass diese Schichten der Regulierung mit der Genexpression unterschiedlichen Ausmaßes korreliert sind. Es ist jedoch noch nicht klar, wie Chromatin, DNA-Topologie und genetische Komponenten interagieren, um diesen Vorgang durchzuführen. Mit diesem Projekt zielen wir darauf ab, die grundlegenden Mechanismen der Organisation bei der Kommunikation von Promotor-Regulationselementen in einem definierten *Locus*-Modell zu verstehen.

In einem vorherigen Projekt des Labors wurden chromosomale Deletionen und *in-vivo* Reporter-Assays verwendet, um die regulatorischen Elemente in *c-Myc Locus* zu identifizieren und charakterisieren. Diese Studie, durchgeführt von Veli Uslu, zeigte, dass während der Embryonalentwicklung von Mäusen, bei 11,5 Tagen Trächtigkeit, verschiedene Transkriptionsverstärker die *c-Myc*-Transkription im Gesichts-Mesenchym mit der embryonalen Leber koordinieren.

Wir untersuchten die Verteilung innerhalb des *c-Myc-Locus* von an den Enhancer Sites angereicherten Histon-Modifikationen, um die genetische Position der Gesichts- und Leber Enhancer auf kleinere putative Regionen zu beschränken. Wir untersuchten auch die Belegung der Isolatoren als auch des architektonischen Proteins CTCF's, an den Standorten, die in die weitreichenden Wechselwirkungen involviert sind. Folglich fanden wir sowohl Gewebe-Invarianten als auch gewebespezifische Bindungsstellen in verschiedenen Teilen des *Locus*.

Wir betrachteten die Rolle der *Locus* Topologie in der Beeinflussung der Promoter-Enhancer-Kommunikation mit der Schaffung von drei ausgeglichenen chromosomalen Inversionen innerhalb des *Locus*. Jede dieser Inversionen soll die Positionen der *c-Myc*-Promotoren und die Langstrecken-Enhancer, sowie CTCF Bindungsstellen auf spezifische Weise neu vermischen. Wir beobachteten Umleitungen von Regulierungstätigkeiten (anhand der in-vivo-Reporter-Assays) und signifikante Reduktion von *c-Myc* Expression im Gesichts-Mesenchym. In der

Leber hingegen beobachteten wir eine leichtere und nicht erhebliche Auswirkung der Inversionen auf die *c-Myc*-Regulierung.

Wir untersuchten die Belegung von CTCF und der Untereinheit vom Cohesin komplexen RAD21 (ein weiteres architektonisches Protein) im Gesichts-Mesenchym und in der Leber, auf den technischen Linien, mit Erfassung im *c-Myc Locus* von nur einigen kleineren Veränderungen im Vergleich zu dem Wt.

Unsere Daten suggerieren, dass die Architektur der *Loci* eine funktionelle Rolle in der Organisation der regulativen Promotor-Elemente haben kann. Wenn diese verändert werden, kann dies zur Gen-Fehlexpression führen. Die gewebespezifischen Auswirkungen der Gentechnik, die wir am *c-Myc-Locus* beobachten konnten, könnte die Tatsache widerspiegeln, dass verschiedene Gewebe unterschiedlich lokalisierte Topologien implementieren, um spezifische, weitreichende Wechselwirkungen zu fördern. Wenn architektonische Proteine an diesem Prozess beteiligt sind, weisen unsere Daten darauf hin, dass deren Bindung zum Genom, das nicht signifikant durch genetische Umordnung verändert ist, vor allem durch die DNA-Sequenz (Genetik) bestimmt werden kann. Es ist jedoch möglich, dass die weitreichenden Wechselwirkungen, die sie durchleben, stärker durch den Kontext beeinflusst werden, in dem sie eingesetzt sind.

2 Introduction

2.1 Transcriptional Control of Gene Expression in Metazoans

In multicellular organisms, every cell needs to execute cell-type specific functions, maintain its lineage identity, respond to cell signaling and be able to tune its physiology according to fluctuations of chemical and physical parameters. The completion of these activities is ensured by the control of gene expression.

The flow of genetic information from DNA to protein is controlled at several levels: transcription, splicing, transcript export and localization, transcript degradation, translation, post-translational-modifications and protein degradation. Although not clearly demonstrated, one possible reason for having so many points of control is to confer robustness to the system, for example, in case one point of control is not working properly. Additionally, this redundancy can also provide flexibility to the system in terms of adaptation timing. In this last case, some changes require a fast and transient response, for example the production and release of specific hormones in dangerous situations. Other stimuli instead require a more stable and not necessarily fast response, for example the specification of cell fate in metazoans.

DNA transcription is controlled at the level of initiation, elongation and termination, (reviewed in Levine, 2011; Porrua & Libri, 2015). Transcriptional initiation is dependent on the assembly on the promoter of the pre-initiation complex (PIC). PIC is constituted by the RNA polymerase II (RNA Pol II) and the general transcription factors (GTF) (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH), which recognize the promoter and load it with the RNA Pol II (T. K. Kim, Ebright, & Reinberg, 2000). This complex allows the transcription of approximately the first 10 nucleotides. At this step the polymerase can either stall, or continue in the main gene body (Hirose & Ohkuma, 2007). Some studies revealed that 10–30% of the inactive promoters are loaded with RNA Pol II (Core, Waterfall, & Lis, 2008; Jonkers & Lis, 2015; Levine, 2011). However, in this case, RNA Pol II stalls and fails to proceed in the main gene body. Until a short time ago it was believed that the initiation step was the only limitation of transcription. Recent data, mostly due to

the great development in sequencing technologies (next generation sequencing, NGS), has started challenging this model. There are now protocols available to assess in a genome wide fashion RNA Pol II occupancy (RNA Pol II ChIP-seq), nascent transcripts (GRO-seq) (Core et al., 2008), unstable transcripts (GRO-cap) (Core et al., 2014a), as well as transcriptional elongation speed (4sUDRB-seq) (Fuchs et al., 2015). The quantity of data obtained by these methods support the idea that the control of pause/release of the RNA Pol II could be a further point of transcriptional regulation, yet it is still not completely clear what regulates the balance between transcriptional initiation and elongation (Zeitlinger et al., 2007).

Once the RNA Pol II overcomes the promoter-proximity and enters the gene-body, the transcriptional speed can differ as much as three fold between different genes (Danko et al., 2013; Fuchs et al., 2014; Jonkers, Kwak, & Lis, 2014; Veloso et al., 2014). There are important differences in transcriptional speed even inside the gene-body: the beginning (within ~ 15Kbp from the promoter) is the slowest phase, characterized by an elongation of 0.5Kbp/minute, while the rest of the gene is transcribed at 2–5Kbp/minute (figure 1) (Danko et al., 2013; Fuchs et al., 2014; Jonkers et al., 2014; Veloso et al., 2014). The increment of RNA Pol II speed likely occurs due to a progressive maturation of the transcriptional machinery while it is elongating, which includes the phosphorylation of the serine 2 of the RNA Pol II carboxy terminal domain (CTD), the removal of pausing factors (if present) and the recruitment of other subunits (such as splicing factors) (Cheng et al., 2012; Heidemann, Hintermair, Voß, & Eick, 2013; Jishage et al., 2012). In addition, it is possible that chromatin and histone modifications can affect the RNA Pol II advancement. Transcribed gene bodies are enriched for H2B ubiquitylation, H3H36me3, H3K56ac and H3K79me2, which can directly increase the DNA accessibility, or function as a scaffold for histone chaperones that can evict nucleosomes or alter their turnover (Bintu et al., 2012; Venkatesh & Workman, 2015; Zentner & Henikoff, 2013). Even inside the fast-transcribed zone, there can be RNA Pol II slowdowns, typically, they occur at exons, or at cleavage and polyadenylation sequences (Gromak, West, & Proudfoot, 2006; Kwak, Fuda, Core, & Lis, 2013). It is possible that splicing events—which can take 20–30 seconds *in vivo*—may be the cause for the transcriptional slowdown observed on exonic sequences (Alexander, Innocente, Barrass, & Beggs, 2010; Jonkers et al., 2014; Kwak et al., 2013; Veloso et al., 2014).

After the gene is fully transcribed, the polymerase has to dissociate from the DNA, in a process known as transcription termination. In yeast (and similarly in metazoans) the cleavage and polyadenylation factor-cleavage factor (CPF–CF)

complex recognizes termination signals in the 3' UTR of the nascent mRNA as well as the phosphorylated serine 2 of the RNA Pol II CTD. The CPF-CF complex recognizes and cleaves the mRNA on the poly(A) site and adds ATP nucleotides on the 3' end of the transcript thereby producing the poly(A)-tail, typical of the mRNA (Kuehner, Pearson, & Moore, 2011). It is still nonetheless debated how the RNA Pol II slips off the DNA, ending the transcription process (M. Kim et al., 2004; Richard & Manley, 2009).

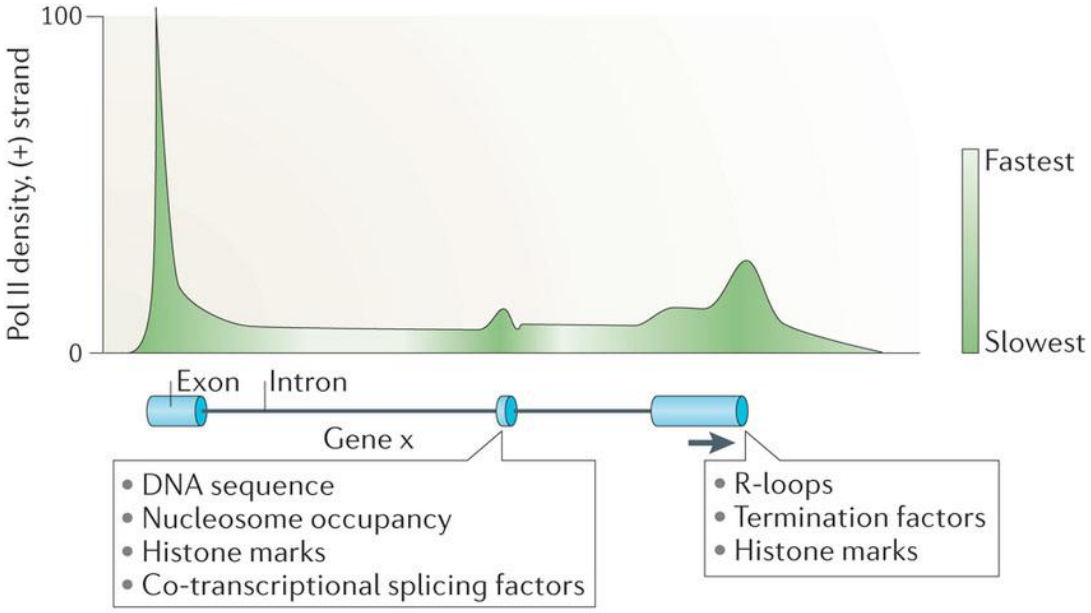


Figure 1 | Transcription elongation speed. Adapted from (Jonkers & Lis, 2015). The figure shows a hypothetical gene. The density of RNA Pol II and the speed of transcription across the gene are not constant. The factors that are thought to affect the transcriptional speed at specific parts of the gene are listed below the gene model.

2.1.1 Genetic Elements of Transcriptional Regulation

2.1.1.1 Promoters

The promoter is defined as the sequence encompassing the transcription start site (TSS) of a gene, necessary to initiate its transcription.

In mammals the definition of promoter includes the core promoter, extending for about 30bp upstream and downstream of the TSS (commonly referred to as

promoter), and the proximal promoter, within a few hundreds of base pairs (bp) upstream of the TSS, which comprises binding sites for TFs.

One distinction in promoter classes is done according to their sequence composition at the proximal promoter: high and low CG content or high and low CpG dinucleotide frequency (Carninci et al., 2006; Yamashita, Suzuki, Sugano, & Nakai, 2005). This distinction partially overlaps with the function of the genes. Promoters with high CG content tend to be found at broadly expressed genes and also, developmentally regulated ones (Carninci et al., 2006). In total, they represent about 60% of the human genes. They are characterized by having multiple TSSs close to each other and by having CpG-islands on the proximal promoter. Promoters with low CG content, are mostly found on tissue specific genes, which tend to have a single TSS (Maston, Evans, & Green, 2006; Ponjavic et al., 2006; Venter et al., 2001).

Promoters can also be classified based on the motifs found at the core promoter. These notably include the TFIIB recognition element (BRE), typically starting from 37 bp upstream of the TSS, the TATA box, 31 bp upstream of the TSS, the Motif Ten Element (MTE), 18 bp downstream of the TSS, and, the Downstream Promoter Element (DPE), just after the MTE (Maston et al., 2006). Of these, DPE and BRE are the most common, present in 25% of the human core promoters, and the TATA box, present in 13% of the human core promoters (Maston et al., 2006).

Another important criteria of classifying promoters, other than the ones based on DNA sequence, is represented by the epigenetic status of the promoter, namely, histone and DNA modifications (treated in more detail in the next sections) which reflect the activity of the promoter in different cell types. Tissue specific expressed genes usually carry the active H3K4me3 and H3K27ac histone modifications only downstream of the TSS whereas broadly expressed genes such as housekeeping genes, carry these marks throughout the promoter (Ernst & Kellis, 2010). In particular the H3K4me3 usually spans the entire CpG island (Deaton & Bird, 2011).

2.1.1.2 Regulatory Elements

It is well established in the literature that regulatory elements (RE(s)) are responsible for mediating transcriptional regulation, acting probably on every transcriptional phase. These regulatory elements are DNA sequences capable of recruiting TF(s). They are usually a few hundreds of bp long and carry specific DNA motifs

recognized by the TFs (reviewed in Spitz & Furlong, 2012). Other than sequence (motifs), indirect binding and DNA accessibility affect the recruitment of TFs to regulatory elements.

One example of indirect binding is given by the co-activator p300–CREB, which is in general, recruited to the DNA by other TFs, through protein–protein interactions (Merika, Williams, Chen, Collins, & Thanos, 1998).

TFs accessibility to motifs is affected by the presence of nucleosomes or because of unsuitable DNA shape. In these cases, remodeling complexes like SWI/SNF or architectural proteins like HMGA1 can respectively slide the nucleosomes away from the binding site or unbend the DNA, in both cases the result is the improvement of the binding of the TF to the RE (Voss & Hager, 2014; Yie, Merika, Munshi, Chen, & Thanos, 1999). ChIP-seq analysis showed that TFs bind only a fraction of the available motifs in the genome, supporting the idea that chromatin context, accessibility and DNA shape represent altogether a major determinant for TF recruitment (Arvey, Agius, Noble, & Leslie, 2012; John et al., 2011a; X.-Y. Li et al., 2011; Tsai, Shiu, & Tsai, 2015).

Regarding the function of REs instead, it is possible that a significant fraction of the TFs binding is non functional, and probably, just have to do with the fact that the DNA in those locations is accessible enough (John et al., 2011b).

Conversely, the binding of TFs to the DNA can have functional implications in transcription, promoting or inhibiting it. In these cases the REs are respectively called enhancers and silencers. TFs can affect transcription by recruiting other co-activators/co-repressors or nucleosome-remodeling complexes. TFs can also act more directly, by promoting the recruitment of the GTF (in case of enhancers), or by targeting their assembly (in case of silencers) (Koch et al., 2011). Lastly, TFs can also recruit DNA and histone modifying enzymes, like for the case of the protein complexes Trithorax and Polycomb (reviewed in Schuettengruber, Martinez, Iovino, & Cavalli, 2011; Steffen & Ringrose, 2014). The post-translational modification of the histones (most often on the histone tails, protruding outside the nucleosome) can in turn have many consequences (discussed in more detail in the chromatin section). They can directly affect the progression of the RNA Pol II on the gene body. Alternatively, they can recruit other TFs, which specifically recognize the histone modification. Finally, they can recruit other proteins that alter the chromatin compaction.

2.1.1.2.1Enhancers

2.1.1.2.1.1 Enhancers Characterization

Enhancers are defined as DNA sequences that can activate transcription through the binding of TFs (activators).

Enhancers were originally discovered in the simian virus 40 (SV40) genome, already in the early eighties (Conrad & Botchan, 1982; de Villiers, Olson, Tyndall, & Schaffner, 1982). Soon after their description in the SV40 system, enhancers were described also in a few other *loci*, including the *Immunoglobulin heavy chain* (*IgH*) genes (Banerji, Olson, & Schaffner, 1983; Gillies, Morrison, Oi, & Tonegawa, 1983).

Later on, the development of the enhancer-trap allowed the discovery of several enhancers and became a standard technique (used for long time) for their identification in many systems, ranging from mammals to *Drosophila* and *Arabidopsis* (Dornan, Gailey, & Goodwin, 2005; Engineer, Fitzsimmons, Schmuke, Dotson, & Kranz, 2005; Korn et al., 1992; Weber, de Villiers, & Schaffner, 1984).

Nowadays, two main different approaches are being used for the identification of enhancers on a genome-wide scale.

The first approach is based on the use of NGS technologies. One method exploits the fact that enhancers tend to lie on accessible DNA regions; techniques such as DNaseI-seq and FAIRE-seq aim to map on the genome these regions (Giresi et al., 2007; Gross & Garrard, 1988; M. Liu et al., 2015; Stalder et al., 1980). Alternatively, histone modifications such as H3K4me1 and H3K27ac, or the coactivator p300, all enriched at enhancer sites, can be mapped on the genome by ChIP-seq (Creyghton et al., 2010a; Heintzman et al., 2007; Rada-Iglesias et al., 2011; Visel, 2009; Zentner, Tesar, & Scacheri, 2011).

The second approach is based on high throughput reporter assays, which allow testing the enhancer potential of thousands of sequences in parallel (W. Akhtar et al., 2013; Arnold et al., 2013; Melnikov et al., 2012; Mogno, Kwasnieski, & Cohen, 2013). One example of this approach is given by the STARR-seq method, firstly applied in *Drosophila* cells, which allowed the identification of thousands of cell-type-specific enhancers (Arnold et al., 2013).

2.1.1.2.1.2 Enhancers Mechanisms of Action

Enhancers were shown to affect every step of transcriptional regulation.

Regarding the transcription initiation, enhancers can promote the assembly of the PIC at genes promoters. It is believed that in most of the cases the way this is achieved occurs through the recruitment of the Mediator protein complex, rather than by a direct targeting. According to this view, enhancers-bound-activator(s) interact with the Mediator and form an active hub where the signals of one or more regulatory elements converge (reviewed in Malik & Roeder, 2010). The Mediator complex, in turn, promotes the formation of the PIC (Esnault et al., 2008; Pavri et al., 2005).

Enhancers can also promote the RNA Pol II release from stalled promoters, improving the elongation step (Hargreaves, Horng, & Medzhitov, 2009; K. Lee, Hsiung, Huang, Raj, & Blobel, 2015; Lokody, 2014; Zippo et al., 2009).

An important feature of enhancers concerns their compatibility with the different promoters. Some reports showed that not any enhancer could act on any promoter.

In *Drosophila*, one study identified that the motifs at the core promoter confer the specificity to the interaction. In this study the capability of a set of known enhancers to activate (*in vivo*) the expression of a reporter gene having either the DPE or TATA motifs at the core promoter was assessed. It was shown that certain enhancers activate only one of the two types of promoters. Regarding the function of the promoter–enhancer specificity, it was speculated that this mechanism could in principle be important in cases where enhancers do not lie in the proximal promoter, but rather are separated from their target gene(s) by a large distance (as described in the next section). Alternatively, this mechanism may be important in the case that the enhancers are closer to a non-target gene then the target one. In these situations, the compatibility between the two could help in sorting promoter–enhancer interactions. On the other hand, in the same study it was also shown that likely this is not a universal mechanism, as it was limited to 4 out of the 18 enhancers tested (Butler & Kadonaga, 2001).

More recent reports performed in a genome-wide fashion suggested a slightly different scenario. In one study the responsiveness of a housekeeping and a developmental gene promoter to all the 11364 candidate enhancers in *Drosophila* S2 cells was analyzed. It was reported that in 32% of the cases, enhancers did not have a preference for either of the promoters (Zabidi et al., 2014).

In the next sections other mechanisms for organizing promoter–enhancer communication will be mainly considered.

2.1.1.2.2 Silencers

Silencers are defined as DNA sequences bound by TFs (repressors) capable of repressing transcription.

Compared to enhancers, the identification of silencers is complicated by the fact that there are no hallmarks available for their high-throughput identification. In addition, the fact that as from definition, silencers repress transcription, makes the reporter assays a less suited method (compared to enhancers) for their discovery and validation. Repressors, namely the TFs that bind silencers and exert the repressing effects, were mostly described. For these reasons, repressors are treated in more detail in this section.

Similarly to activators, repressors were shown to impair transcription in different ways.

The dorsal switch protein (DSP1) in *Drosophila* was shown to repress the *Zerknullt* (*Zen*) gene by binding TBP and removing TFIIA from the DNA. As a consequence, DSP1 targets the assembly of the pre-initiation complex and prevent the loading of RNA Pol II to the promoter (Doyle, Kraut, & Levine, 1989; Kirov, Lieberman, & Rushlow, 1996). Similarly, in mouse the repressor Mohawk binds the pre-initiation complex proteins TBP, TFIIA1 and TFIIB, as well as co-repressors histone deacetylases enzymes (Anderson, Beres, Wilson-Rawls, & Rawls, 2009). Other studies showed that repressors can target transcription initiation by binding to the Mediator (reviewed in Malik & Roeder, 2010).

Another mechanism of transcriptional silencing involves the pausing of the RNA Pol II. A few proteins—including the negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF)—in both mammals and *Drosophila* were shown to increase the RNA Pol II stalling on the promoter-proximal region (20–60bp away from the TSS) (Missra & Gilmour, 2010; Natarajan et al., 2013; Yamaguchi, Shibata, & Handa, 2013). In one study, the transcription factor SP3 was shown to inhibit the transcriptional elongation at the $p21^{CIP}$ gene. The depletion of SP3 led to increase $p21^{CIP}$ mRNA levels, and, was accompanied by an increment of H3K36me3 (histone modification associated with transcriptional elongation) and by a reduced NELF occupancy at the $p21^{CIP}$ promoter (Valin, Ouyang, & Gill, 2013). Similarly, the SNAIL repressor in *Drosophila* was shown to inhibit the release of the RNA Pol II from the genes promoters (Bothma, Magliocco, & Levine, 2011).

Often repressors recruit on silencer elements other co-repressors, commonly histone deacetylases. One example of this mechanism is given by the repressor RUNX2, which suppresses transcription at the rDNA in a HDAC1 (histone deacetylase) dependent manner (Ali et al., 2012).

Finally, some repressors such as Polycomb group proteins affect transcription by chemically modifying histone tails. Their activities influence the properties of the chromatin, and hence, are described in more detail in the chromatin section.

2.1.1.2.3 Short-Range and Long-Range Regulatory Elements in Metazoans

The regulation of gene expression at transcriptional level exists in all living organisms from bacteria to eukaryotes. However there are a few fundamental differences between these two *taxa*.

In bacteria the process is fairly simple, it is based on regulatory sequences located very close to, or even at the very promoter. These sequences are bound by TFs that can either help recruit the RNA polymerase, or, occupy the promoter, in this way preventing the RNA polymerase binding it. A famous example of this mechanism is given by the Lactose operon discovered by Francois Jacob, Andre´ Lwoff and Jacques Lucien Monod (awarded the Nobel prize in physiology or medicine, in 1965).

In metazoans, the spatial relationship between regulatory elements and promoters is more complicated; regulatory elements lie both upstream and downstream of the target gene, as well as inside gene introns (Ott et al., 2009). The distance between them also varies, ranging from hundreds of bp to hundreds of Kbp (respectively, short and long-range regulatory elements). For example in mammals, in erythroid cells, the expression of the *β-globin* genes depends on regulatory sequences located 20–30Kbp upstream of the gene cluster (figure 2) (Curtin, Liu, Liu, Chang, & Kan, 1989; Hardison et al., 1997; Moon & Ley, 1990).

The most accepted model—explaining how long-range regulatory sequences influence their target promoters—proposes that the DNA molecule is flexible and can allow the formation of looped structures (Bulger & Groudine, 1999). The validity of the "looping" model is thought to extend to all the eukaryotes, from yeast to mammals (Petrascheck et al., 2005).

It is still not clear how many regulatory sequences act on a given gene at any time, and the physiological meaning of having both short and long-range regulatory elements. Recent studies performed with adaptations of the 3C and Hi-C methods (Capture-C and HiCap) were particularly useful in order to map interactions between distal regulatory elements and promoters (Hughes et al., 2014; Sahlén et al., 2015).

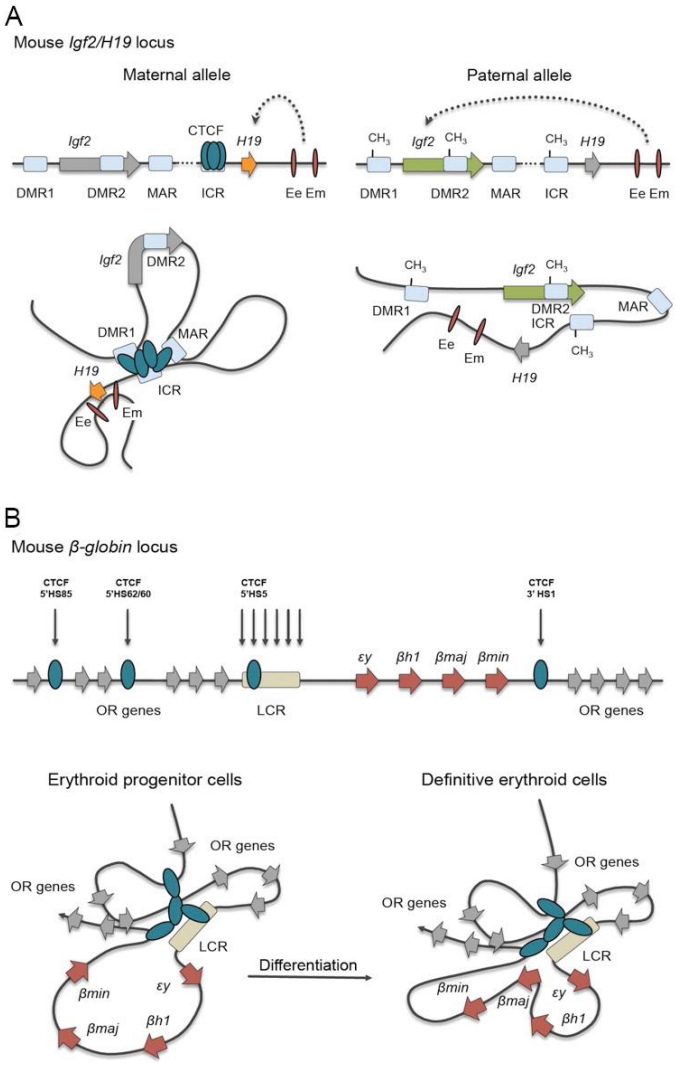


Figure 2 | *Igf2-H19 and Hbb locus* topology. Adapted from (Ribeiro de Almeida, Stadhouders, Thongjuea, Soler, & Hendriks, 2012). A) *The Igf2-H19 locus* is subjected to differential methylation (DMR) on paternal and maternal alleles. In the maternal allele CTCF binds the imprinting control region (ICR) and mediates DNA looping, important for preventing the endodermal enhancer (Ee) and the mesodermal enhancer (Em) to act on *Igf2* (leading to *Igf2* silencing and to the expression of *H19*). In the paternal allele, CTCF is unable to bind the methylated ICR, and as a result, the enhancers reach *Igf2* and activate its expression (Bell & Felsenfeld, 2000; Hark et al., 2000). B) The *β-globin genes (Hbb locus)* lie within the olfactory receptor (OR) gene-cluster. CTCF binds four DNaseI hypersensitive sites (HS) scattered throughout the region, one of them in the *locus* control region (LCR). CTCF-mediated looping is important to condensate the OR gene-cluster. This compaction brings several regulatory sequences into proximity, and as a result, it forms at the LCR an "active chromatin hub". During cells maturation only certain *β-globin* genes interact with the LCR while others stick out of the loop and remain silent. In eryhroid progenitor cells, the *εy* and *βh1* globin genes are expressed; on differentiated erythroid cells, *εy* and *βh1* become silenced, and the adult genes *β-major* (*βmaj*) and *β-minor* (*βmin*) gain expression.

The HiCap method performed in mESC, showed that distal interactions often occur within 100Kbp, and that distal regions interact most often with just one promoter whereas promoters on average interact with approximately 6 distal regions (Sahlén et al., 2015). Interestingly, this study also showed that 65% of the total interactions occur between a given element and the closest promoter. This evidence suggests that in a significant proportion of the cases promoter–distal element interactions can skip gene(s) and argues against some old models present in the literature, proposing that promoter–enhancer interactions are uniquely based on proximity.

2.1.1.3 Insulators

In eukaryotes genetic insulators have been described in many *taxa*, ranging from yeast, to *Drosophila* and mammals (Kirkland, Raab, & Kamakaka; Vogelmann, Valeri, Guillou, Cuvier, & Nollmann). The physiological importance of insulators is particularly evident in metazoan genomes, where the long-distance activity of REs may (in principle) not be confined only to their target genes, but also extend to others, leading to ectopic expression of the latter, with potential detrimental consequences. Insulators were therefore proposed to limit the range of action of REs.

The definition of insulator encompasses two different genetic mechanisms. The first category, "enhancer-blocking" represents DNA sequences capable of disrupting the communication between a regulatory element and a promoter when placed in between (Conte, Dastugue, & Vaury, 2002; Geyer & Corces, 1992). The second category is represented by DNA sequences capable of containing repressive effects, notably the spreading of heterochromatin. This function is carried out by insulators with a "barrier" activity (Gaszner & Felsenfeld, 2006; Valenzuela & Kamakaka, 2006).

Despite the fact that some insulators behave either as enhancer blockers or as barriers, some others can act as both. Examples of the last category are the short interspersed element (SINE) in mouse and the gypsy insulator in *Drosophila* (Gdula, Gerasimova, & Corces, 1996; Tatiana I Gerasimova, Lei, Bushey, & Corces, 2007; Lunyak et al., 2007; Román, González-Rico, & Fernández-Salguero, 2011).

In both mouse and yeast it was observed that the tRNA gene promoters have a barrier function, blocking the spread of repressive heterochromatin into actively transcribed domains. The barrier activity is not dependent on RNA Pol III transcription and is carried out by the general transcription factors (for the RNA Pol

III) TFIIIB and TFIIIC (Noma, Cam, Maraia, & Grewal, 2006; Valenzuela, Dhillon, & Kamakaka, 2009). Another mechanism of barriers is observed at the *β-globin locus*. In chicken cells this *locus* is insulated from an adjacent heterochromatic domain by the HS4 insulator. In this case the barrier activity requires the proteins CTCF, USF1 and VEZF1. The subsequent recruitment of histone modifying enzymes bring active modifications that counteract the heterochromatin spreading (Dickson et al., 2010).

Regarding the enhancer-blocking insulators, reporter assays in *Drosophila* allowed the identification of several sequences that possess this property (Parnell & Geyer, 2000). In mammals, an example of enhancer-blocking insulator is the imprinting control region (ICR) located between the *Igf2* promoter and its enhancer (figure 2). In the maternal allele, the binding of CTCF to the ICR is essential to maintain the insulating properties of this sequence and does not allow the enhancer to act on the *Igf2* promoter, keeping the gene silenced (Kurukuti et al., 2006), (reviewed in P. Singh, Lee, & Szabó, 2012).

Two independent models can describe the mechanisms of the enhancer-blocking insulators. One proposes that the enhancer does not reach the promoter because it is physically sequestered by the enhancer-blocking insulator (Nolis et al., 2009). Another model proposes that enhancer-blockers and perhaps, insulators in general, are bound by architectural proteins and interact with each other *via* DNA loops between them, affecting the 3D architecture of the *loci*. As a consequence, this would constrain one enhancer in one loop, preventing it reaching other promoters (T I Gerasimova & Corces, 2001). According to this view, several recent evidence support the idea that enhancer-blocking insulators do not work autonomously. Rather, their activity seems to be largely influenced by the surrounding sequence, showing significant differences according to the context in which the insulators are placed (reviewed in Chetverina, Aoki, Erokhin, Georgiev, & Schedl, 2014). For example, in *Drosophila*, if the enhancer-blocker su(Hw*)* is placed between a promoter and its enhancer, as defined, it can block the activity of the enhancer. However, this activity is lost if two copies of su(Hw) are used instead of one (Muravyova, 2001). Supporting the context-dependency model, other observations showed that enhancer-blockers can indeed impair the communication between enhancer and promoter if arranged as described above, but can also improve enhancer-promoter interactions in other configurations. For example, both insulators yellow-1A2 and white-WARI in *Drosophila* lead to reporter activation if placed downstream of a reporter and Gal4 motif construct (reviewed in Chetverina et al., 2014).

In *Drosophila*, several proteins are enriched at enhancer-blocking sequences, amongst them are CTCF, the boundary element associated factor (BEAF) and the GAGA factor (Van Bortle et al., 2014). In human, CTCF is enriched between adjacent genes with low correlation in expression level, compared to sites where the flanking genes have similar expression levels (Xie et al., 2007). In *Drosophila* similar findings were obtained for the protein BEAF and in a few cases, CTCF was also shown to act as enhancer blocker (Bell, West, & Felsenfeld, 1999; Wood et al., 2011; Yang, Ramos, & Corces, 2012; Yusufzai, Tagami, Nakatani, & Felsenfeld, 2004). Despite these few reported cases, the role of these TFs in mediating insulating effects, and their precise mechanism of action remain to be clarified.

2.1.1.4 Tethering Systems

The evidence that the physical interaction between long-range regulatory elements and promoters occur due to DNA-loops between them, leads to the consideration that some regulatory elements may have evolved just to carry out the function of mediating the formation of DNA-loops.

One example of such a system is given by the *Antennapedia* gene complex in *Drosophila* (Calhoun, Stathopoulos, & Levine, 2002; Ohtsuki, Levine, & Cai, 1998). The *Antennapedia* is one of the most important *Hox* gene-clusters in *Drosophila*. Misexpression of these *Hox* genes during development impairs the correct embryonic segmentation, causing homeotic transformation, and in some cases, death. The expression of the *Src* gene, in different tissues of the parasegment-2, is dependent on the T1 enhancer. T1 lies 25Kbp telomeric to the gene promoter and in between the two there is another gene, *Ftz*. Interestingly, the T1 enhancer does not activate *Ftz* (which in addition to being the closest, has a stronger promoter than *Src*). Somehow T1 skips *Ftz* and activates *Src*. This phenomenon seems to be mediated by the tethering element Src450 located just upstream of the *Src* promoter. Remarkably, if the Src450 sequence is placed upstream of the *Ftz* promoter, *Ftz* acquires expression in the same tissues as *Src* as a result of the recruitment of the T1 enhancer.

The molecular mechanisms and the TFs responsible for this behavior are unknown. In human erythroid cells, in the *globin* gene-cluster, it is speculated that GATA1–AP1/NF-E2 interactions are serving a similar purpose (Gong & Dean, 1993; Gourdon et al., 1992).

2.1.2 Topological Component of Transcriptional Regulation: From Chromosome Territories to DNA Loops

2.1.2.1 Nuclear Compartments

As recently as the late 1970s, with the development of DNA–FISH protocols for "chromosome painting", it was observed that chromosomes tend to have ordered arrangements inside the nucleus—termed chromosome territories (CT(s)). In particular, larger chromosomes tend to stay at the periphery, while smaller ones are mostly at the interior (Sun, Shen, & Yokota, 2000). Correlations between gene activity and nuclear positioning were also described. Active genes and gene-rich chromosomal regions are preferentially located towards the interior of the nucleus, while inactive genes and gene-poor chromosomal regions tend to stay at periphery (Bolzer et al., 2005; Croft et al., 1999).

Regarding the dynamicity of the DNA movements inside the nucleus, *in vitro* studies performed in mammalian cells, showed that during interphase, DNA sequences are constrained in a volume of about 1% of the total volume of the nucleus, corresponding to a radius of 1μm (Edelmann, Bornfleth, Zink, Cremer, & Cremer, 2001). The early G1 phase is characterized by movements greater than 2μm and represents the only phase of the cell cycle where important DNA repositioning is observed. After that, chromosomal segments seem to be more constrained within their CTs (Walter, Schermelleh, Cremer, Tashiro, & Cremer, 2003).

2.1.2.2 Lamina Associated Regions

One mechanism presumably important for organizing the CTs (and the topology of the chromosomes) may be the tethering of DNA segments to the nuclear periphery. The nuclear-lamina has been hypothesized to mediate this interaction.

The nuclear-lamina is composed of a network of V-intermediate-filaments (lamins) attached to the nuclear membrane by lamin-associated membrane proteins (LAP) (reviewed in Georgatos, Meier, & Simos, 1994; Gruenbaum et al., 2003). Lamins are split into two groups: type A and type B, both able to bind a class of DNA sequences called matrix attachment regions (MARs) (Dijkwel & Hamlin, 1988; Forrester, van Genderen, Jenuwein, & Grosschedl, 1994; Loc & Strätling, 1988; Ludérus et al., 1992). These are typically 100–300bp long, with a high content of A/T nucleotides. It was also shown that lamins can bind the core histones in a

MARs independent manner (Goldberg et al., 1999; Taniura, Glass, & Gerace, 1995).

From a physiological perspective, lamins deficiencies cause several diseases termed "laminophaties" that range from muscular dystrophies to neurophaties (Camozzi et al., 2014). It was described that human fibroblasts carrying dominant-negative mutations on the lamin-A gene had genome-wide perturbations in DNA–lamin interactions, which as consequences, led to the loss of definition of CTs and loss of peripheral heterochromatin (McCord et al., 2013). Similarly, other studies performed in mouse fibroblasts, suggested that lamins deficiencies have consequences at the level of CTs, for example, with the relocation of chr18 towards the nucleus interior (Malhas, Lee, Sanders, Saunders, & Vaux, 2007). On the other hand, other studies on mESC showed that the knockout of lamin-A and lamin-B did not cause any significant change of interaction between the DNA and nuclear-lamina. As suggested by the authors, other nuclear-lamina components could be involved in the process (Amendola & van Steensel, 2015). For example, it was observed that the lamin-B receptor is important for anchoring heterochromatic regions to the nuclear envelope (Solovei et al., 2013). Other screenings in human fibroblasts showed that several nuclear envelope trans-membrane proteins promote DNA localization to the nuclear periphery. In particular, individual proteins were shown to affect the nuclear localization of only certain chromosomes or to have an effect only on certain cell types (Zuleger et al., 2013).

Nuclear lamina was shown as an important determinant to organize the positioning of heterochromatic regions (McCord et al., 2013). One histone mark found in heterochromatin is the H3K9me3, and interestingly, in mammalian cells H3K9me2 is enriched at the nuclear periphery. G9a, the methyl-transferase that catalyzes this modification, promotes DNA–nuclear-lamina interactions (Kind et al., 2013). In addition to the G9a enzyme, the modifications H3K9me2/3 themselves were also shown in few contexts to be involved in the process (Bian, Khanna, Alvikas, & Belmont, 2013). Finally, other repressive chromatin marks, namely Polycomb and the histone H1 were proposed to be involved in nuclear-lamina DNA localization; however, their roles in the process remain to be clarified (Guelen et al., 2008; Ikegami, Egelhofer, Strome, & Lieb, 2010; Izzo et al., 2013).

Because of its role in the maintenance of nuclear architecture, nuclear-lamina seems to have some effects on gene expression. Consistent with the idea that heterochromatic parts of the genome are located at the nuclear periphery, most genes in contact with nuclear-lamina are very lowly expressed (Pickersgill et al., 2006). However, correlation between spatial position and expression level does not prove

causality. In one study, performed in human cells, with the aim of investigating this question, specific chromosomes were tethered to the nuclear periphery by binding to a nuclear envelope protein. It was observed that after this perturbation, there was a reversible silencing of some, but not all, of the genes involved. This silencing was shown to be dependent on the activity of histone deacetylases enzymes. Importantly, the observation that several genes did not show significant reductions in gene expression, suggests that transcription at the nuclear periphery is possible (Finlan et al., 2008). Similarly, more recent studies showed that the selective anchoring of active genes to the nuclear-lamina in some but not all the cases can reduce their expression level (Towbin, Gonzalez-Sandoval, & Gasser, 2013). The role of nuclear-lamina in transcriptional regulation is also supported by the observation that during mESC differentiation, several genes move towards the nuclear-lamina and become silenced, others instead move away from it and gain expression (Peric-Hupkes et al., 2010). Therefore, it is possible that the DNA interaction with the nuclear-lamina could be an important mechanism not only for the maintenance of the nuclear/chromosomal architecture, but also for gene silencing.

It is still debated in the literature what makes the nuclear periphery such a repressive environment. One possibility could be that the enrichment of histone modifying enzymes like histone deacetylases and the methyl-transferase G9a are responsible for the gene silencing (Finlan et al., 2008; Milon et al., 2012). Another possibility could be that the higher DNA compaction observed at the periphery compared to the nuclear interior presents many obstacles for the TFs to bind their motifs (W. Akhtar et al., 2013; Nagano et al., 2013).

2.1.2.3 Chromatin Conformation

Next generation sequencing technologies have recently allowed the exploration of the genome architecture at a smaller scale than previously possible with the DNA–FISH (used to map the chromosome territories). The DNA–FISH can also be used to determine spatial distances between probed DNA sequences, however the throughput is low. With chromosome-capture derived methods it is now possible to assess DNA interaction in a genome-wide manner (Barutcu et al., 2015).

Within nuclear territories, chromosomes are organized into large compartments, typically of several Mbp. They are referred to as A and B compartments, representing active and inactive chromatin respectively (Lieberman-Aiden et al., 2009; Simonis et al., 2006). Inside the nucleus, both A and B compartments tend to

cluster with other compartments of the same kind (reviewed in Dekker & Heard, 2015).

At a smaller scale, Hi-C data revealed that the genomes of eukaryotes are arranged in topologically-associated-domains (TAD) (J. R. Dixon et al., 2012). TADs are defined as chromosomal regions where DNA sequences have enriched contact frequency within the domain, compared to sequences located outside. In the mouse genome there are approximately 2200 TADs, with average size of 900Kbp. Their position in the chromosomes is relatively constant across different cell types and even across different, but related species (J. R. Dixon et al., 2012). In *Drosophila* TADs share most of the characteristics found in mammals except that they are smaller (~ 60Kbp), probably due to the higher gene density in this model (reviewed in Dekker & Heard, 2015).

In mammals, TAD-boundaries—namely the regions separating adjacent TADs—are enriched for housekeeping genes, tRNA genes, SINE retrotransposons and CTCF motifs. They are instead depleted of tissue specific genes, H3K4me1 and H3K9me3 histone modifications (J. R. Dixon et al., 2012).

It was shown that despite the fairly conserved organization of TADs across different cell-types, the specific long-range interactions within them can change considerably (Nora et al., 2012). In fact, due to their large size, in mouse, TADs contain multiple genes, on average 32 (the median is 19). One early study identified 9888 dynamic interacting regions in the mouse genome, which in more than 96% of cases, occurred inside the same TAD. These regions were enriched for genes differentially expressed across different cell-types and contained 20% of all genes with expression changes of at least four fold. The authors proposed that the contacts inside TADs could be dynamic and engage tissue specific REs (J. R. Dixon et al., 2012). Therefore it is possible that either TADs do not completely overlap with the concept of regulatory domains (RD(s)) (genomic regions within which a given set of regulatory elements exerts its activity), or that they contain several regulatory domains (Symmons et al., 2014).

In mammals CTCF and Cohesin complex are thought to be important for chromosome architecture. Some early evidence of this comes from the *H19-Igf2 locus*. Many of the long-range interactions in this *locus* depend on the presence of CTCF motifs at the *H19* imprinting control region (ICR) (figure 2A) (Hark et al., 2000; Ribeiro de Almeida et al., 2012; Zhao et al., 2006). Similar observations regarding the need for CTCF binding sites for maintaining the *locus* architecture were also made for the *Hbb locus* (figure 2B) (Splinter et al., 2006). Besides these *locus*-based cases, some studies investigated the role of CTCF and Cohesin complex

at genomic level using the Hi-C method (Seitan et al., 2013; Sofueva et al., 2013; Solovei et al., 2013). Despite these studies leading to slightly different conclusions (probably due to the different cellular models used), they agree that in both cases when Cohesin or CTCF are depleted, there is a general decrease in intra-TAD interactions. In the case of the Cohesin knockout, this relaxation had no effect at the level of TAD organization. Instead, for CTCF knockout, the decrement in intra-TAD interactions was balanced by a slight increase of inter-TADs interactions (Zuin et al., 2013). This evidence indicates that although TADs are still present in protein knockout experiments, both CTCF and Cohesin complex could play a role in TADs definition and in the genome organization. CTCF and Cohesin complex probably contribute in different ways to these processes and may not be the only factors involved.

Other data suggest that TFIIIC may also be important to further stabilize long-range interactions and specify TAD-boundaries (Raab et al., 2012), (reviewed in Kirkland, Raab, & Kamakaka, 2013; Van Bortle & Corces, 2012).

Finally, the Mediator complex was proposed to be important for the formation of DNA-loops connecting enhancers to promoters. Firstly, genome-wide data showed a highly significant correlation of the binding to the genome of the Mediator and NIPBL (Cohesin loading factor) (Kagey et al., 2010). Secondly, knockout experiments showed that the MED1 subunit of the Mediator complex is required to bring into close proximity the T3 response element to the *Crabp1* gene promoter and induce its expression (Park et al., 2005).

2.1.2.4 Chromosome Architecture Factors

The two factors involved in genome organization discussed here are CTCF and the Cohesin complex.

CTCF, also known as CCCTC-binding-factor is a zinc-finger protein, conserved in bilaterian animals (Heger, Marin, Bartkuhn, Schierenberg, & Wiehe, 2012). It recognizes and binds DNA on the consensus sequence CCGCGNGGNGGCAG.

Cohesin is a protein complex formed by the 4 subunits SMC1, SMC3, SCC3/SA and SCC1/RAD21. It is evolutionarily conserved in all eukaryotes. The primary function of Cohesin is to keep sister chromatids together during DNA replication, G2 and early M phases. More recently this complex was described as having a function in interphase, in particular in gene expression. Because of its

capability of linking different DNA molecules, observed in G2 and early M phases, it was hypothesized that Cohesin plays a role in gene expression by mediating DNA looping, important for promoter–regulatory element interactions (Rollins, Morcillo, & Dorsett, 1999; West, Gaszner, & Felsenfeld, 2002).

Chromatin-immunoprecipitation, coupled with high throughput sequencing (ChIP-seq) methods, allowed the mapping of the genomic location of both CTCF and Cohesin binding sites. In mammalian cells, CTCF binds around 55000 sites, most of them conserved across cell types. About half of binding sites occur at intergenic regions, while about 15% occur nearby the promoters and about 40% in intragenic regions (H. Chen, Tian, Shu, Bo, & Wang, 2012).

Cohesin binding in the genome is somewhat more variable, ChIP-seq for the SMC1 subunit revealed 46000 binding sites in mouse cortex tissue and 32000 in mouse pancreatic tissue (Cuadrado, Remeseiro, Graña, Pisano, & Losada, 2015). Similar to CTCF, about 12% of the SMC1 binding sites occur nearby annotated promoters, while about 45% are found in intragenic and intergenic regions.

There is an extensive overlap between CTCF and Cohesin occupancy, about 50–80% of the CTCF binding sites are also bound by Cohesin (Hou, Dale, & Dean, 2010; Parelho et al., 2008a).

Regarding the cell-type conservation of the CTCF binding sites, it is shown that 20–50% of the them have, at least to some extent, cell-type variability (Heintzman et al., 2009; T. H. Kim et al., 2007). It is not completely clear what determines CTCF (and perhaps Cohesin) binding at these sites. A first clue is that in the CTCF motif there is a CpG dinucleotide, suggesting that differential methylation determines which sites can be bound in different cell types. Indeed, one study showed that 41% of the tissue specific CTCF binding sites do correlate with differential methylation (Wang et al., 2012). The case of the methylation-sensitive binding of CTCF at the *H19-Igf2 locus,* which controls the epigenetic imprinting as well as the function of the ICR insulator, is one clear example of this mechanism (Guibert et al., 2012). On the other hand, the presence of many cell-type specific sites that are not differentially methylated indicates that other mechanisms—in addition to DNA methylation—must be involved in the process (Wang et al., 2012). One possibility is that the binding of CTCF to the DNA may be stabilized by other proteins/cofactors. Few proteins were proven at specific *loci* to be necessary for CTCF function. Amongst these there are NUCLEOPHOSMIN, PARP1, YY1 and Cohesin (Zlatanova & Caiafa, 2009). However, only Cohesin so far has been shown to be important for CTCF activity on a genomic scale (Parelho et al., 2008b; Rubio et al., 2008; Wendt et al., 2008; Xiao, Wallace, & Felsenfeld, 2011).

In mammals as well as *Drosophila*, architectural factors binding sites are slightly enriched at TAD-boundaries (J. R. Dixon et al., 2012; Van Bortle et al., 2014). Genome-wide studies suggest their involvement in maintaining the topological organization of the chromosomes and in delineating TADs (Seitan et al., 2013; Sofueva et al., 2013; Solovei et al., 2013). In mouse, TADs cover about 90% of the genome (excluding centromeres and telomeres) and contain 85% of CTCF binding sites. TAD-boundaries instead cover about 10% of the genome (excluding centromeres and telomeres) and contain 15% of the CTCF binding sites. Importantly, despite a slight increase of CTCF occupancy at TAD-boundaries, the majority of CTCF sites still occur intra-TAD (J. R. Dixon et al., 2012). This simple evidence, valid also for Cohesin, suggests that architectural factors may be involved in other functions (not necessarily architectural) other than TAD specification.

Among these functions, it is proposed that Cohesin and CTCF could be important to mediate promoter–REs long-range interactions on a sub-TAD scale. To go deeper into this question, the Hi-C technique, performed with standard sequencing depth, does not have the resolution to map intra-TAD topologies. The 5C method instead, is usually implemented over genomic regions of around 1Mbp and can overcome this limitation. In mESC the use of this technique, revealed that intra-TADs interactions occur between sites bound by CTCF and/or Cohesin. Knockdown of SMC1 and MED12 (subunit of the Mediator complex) had consequences at the level of chromosome architecture, and also resulted in the downregulation of genes engaging long-range interactions mediated by the Cohesin (Phillips-Cremins et al., 2013).

2.1.3 Chromatin Component of Transcriptional Regulation

Chromatin has a wide influence on transcription. Firstly, it affects DNA accessibility, an important feature for TFs binding. Secondly, through many different chemical modifications (mostly on histones and DNA), it labels specific sites for recruiting enzymes and proteins (reviewed in V. W. Zhou, Goren, & Bernstein, 2011) (figure 3). Lastly, it contributes to maintaining cell identity and the transcriptional program across cell divisions. This form of phenotypic inheritance is known as "epigenetic" and occurs by passing modified histones to the daughter cells in mitosis (Campos, Stafford, & Reinberg, 2014). In this project, only the role of chromatin in gene expression and not in epigenetics is considered.

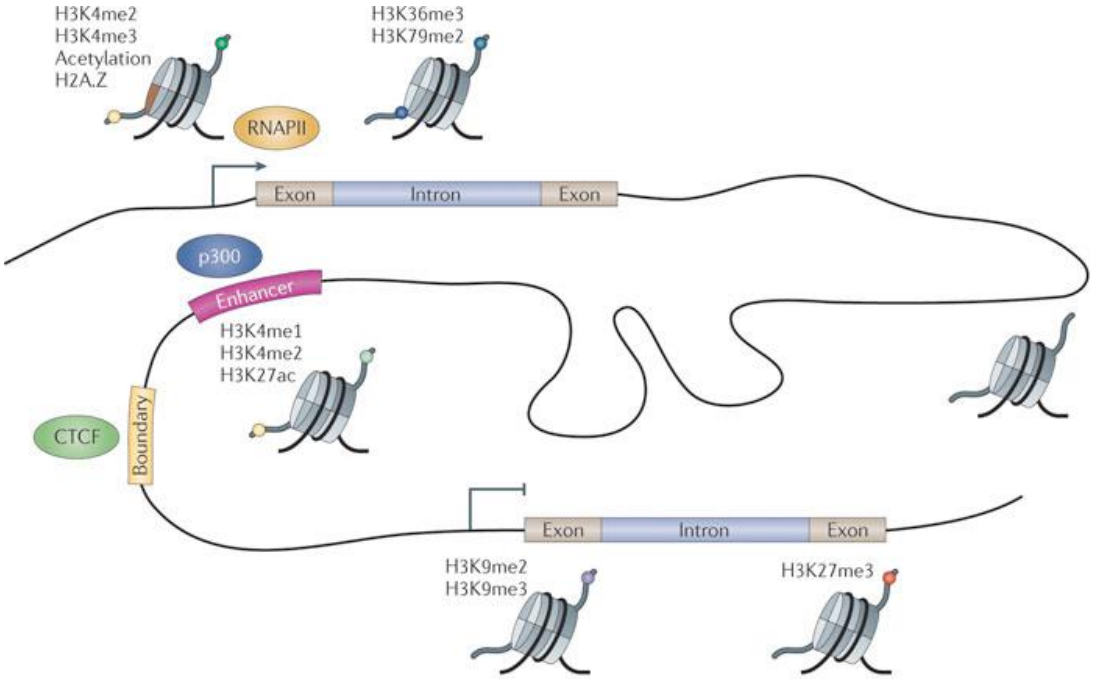


Figure 3 | Histone modifications and their context. Adapted from (V. W. Zhou et al., 2011).

The binding of TFs to their motifs is typically affected by their accessibility on the DNA, which depends on the presence of the nucleosome at these sites as well as on higher order of compaction. Except for "pioneer transcription factors", a class of TFs that can bind nucleosomal DNA, most often histones and TFs, compete with each other to bind the DNA (reviewed in (Iwafuchi-Doi & Zaret, 2014). DNA accessibility is controlled by different mechanisms. These include the nucleosome repositioning, done by the remodeling complexes; the replacement of histone subunit with histone variants, giving the nucleosome different affinity for the DNA and different binding turnover; and the post-translational modifications of the histone tails.

Histones interact with the DNA to a large extent *via* hydrophobic interactions and ionic bonds. Modification of the histones by lysine acetylation removes the positive charge on that residue, thereby decreasing the interaction with the negative charge of the DNA. This, results in a looser interaction of the histone with the DNA. As a consequence, the DNA is more accessible and the remodeling complexes can move the whole nucleosome more freely. Acetylated histones are commonly found on active gene promoters and gene bodies as well as on active enhancers. Proteins that have a bromodomain recognize the acetylated lysines. In higher eukaryotes there are 5 bromodomain-extraterminal (BET) proteins: Brd2, Brd3,

Brd4, and Brdt, some of them, subunits of remodeling complexes such as SWI/SNF-like and the general transcription factor TFIID (Krogan et al., 2003; Matangkasombut, Buratowski, Swilling, & Buratowski, 2000).

One way of reducing DNA accessibility is *via* condensation. One important factor mediating chromatin compaction is the protein HP1, which binds the H3K9me3 histone modification (enriched in heterochromatin), and was shown *in vivo* to mediate large-scale chromatin condensation (Cowell et al., 2002; Mosch, Franz, Soeroes, Singh, & Fischle, 2011; Verschure et al., 2005). The histone H1 was also shown to mediate chromatin compaction. Unlike the other histones, it binds the linker DNA between the nucleosomes and promotes condensation. Rather than a general transcriptional repressor, the histone H1 seems to be mainly involved in repression of transposable elements and specific sets of genes (Vujatovic et al., 2012). In *Drosophila* H1 seems to be involved in repression of ~ 50% of transposable elements and ~ 10% of the protein coding genes (Lu et al., 2013). In mammals, the study of the histone H1 is complicated by that presence of five different H1 isoforms that can complement each other's function when single knockout experiments are performed (Drabent, Saftig, Bode, & Doenecke, 2000; Sirotkin et al., 1995).

Despite the large number of chromatin proteins and histone modifications described in the literature, an important study in the chromatin field revealed that from a functional perspective, there are just few major types of chromatin (Filion et al., 2010).

In this study, the group of Bas Van Steensel mapped the location of more than fifty chromatin proteins in *Drosophila* cells. Principal component analysis of the coverage profiles revealed that altogether; the genome can be divided into five major chromatin categories. Using multivariate segmentation algorithms they then segmented the genome into five states, characterized by similar protein composition. Almost concomitantly other studies reported similar observations in mammals. Although the authors did not describe the chromatin types in particular, they found similar correlations of DNA binding proteins and histone modifications with the different transcriptional activities of the chromatin (Ernst et al., 2011).

The five main chromatin types found in *Drosophila* by Filion and colleagues are briefly summarized in the next sections.

2.1.3.1 Transcriptionally Active Chromatin

There are two distinct kinds of active chromatin: one is found on genes with tissue specific expression, the other, on genes with broad expression across different tissues. Both types are characterized by the active histone marks H3K4me3 and H3K27ac at the promoter, however, while in the case of broadly active genes these modifications span the whole promoter, the tissue-specific genes present them only downstream of the TSS (Deaton & Bird, 2011; Ernst & Kellis, 2010).

The H3K4me3 histone modification is catalyzed by the SET domain present in the Trithorax group proteins (TrxG) in *Drosophila*, and in the myeloid/lymphoid or mixed-lineage leukemia (MLL) (ortholog of Trithorax) in mammals (Katsani, Arredondo, Kal, & Verrijzer, 2001; Yu, Hanson, Hess, Horning, & Korsmeyer, 1998). H3K27ac is catalyzed by p300 (Tie et al., 2009).

The comparison of the transcript levels in broadly active and tissue-specific genes shows that there is almost no difference between the two categories. However, the chromatin associated with broadly expressed genes is the only chromatin type in *Drosophila* to have the H3K36me3 modification, previously considered a universal marker of transcriptional elongation (Rando & Chang, 2009; Sims & Reinberg, 2009). Consistently, in mammals, H3K36me3 is also enriched at CpG rich promoters (M. M. Akhtar, Scala, Cocozza, Miele, & Monticelli, 2013).

2.1.3.2 Enhancer Associated Chromatin

The chromatin found at active enhancer sites is an active type of chromatin. Filion and colleagues included it in the two types of active chromatin and did not attribute it to an independent class (Filion et al., 2010). It shares many common features with the chromatin found at active genes, but it also has a few differences.

Similarly to all the active chromatin types, enhancer-associated chromatin is characterized by high DNA accessibility and low nucleosome density, which is an important feature for the binding of TFs (Boyle et al., 2008), (reviewed in (Zaret & Carroll, 2011).

On the other hand, while active genes have H3K4me3 and H3K27ac histone modifications at the promoter, enhancers do not have the H3K4me3. Instead, they often have the H3K4me1 histone mark, and, acquire the H3K27ac only if they are active.

Active enhancers are enriched for the remodeling complex SWI/SNF, for the Mediator complex and for the coactivators Brd4 and p300 (Euskirchen et al., 2011; Lovén et al., 2013).

Inactive enhancers instead can carry the H3K27me3 modification, indicating Polycomb repression (Simon & Kingston, 2009).

One feature of the enhancers-associated chromatin, recently brought to light thanks to the improvements of the methods aimed to find unstable transcripts, is that there is transcription ongoing at enhancer sites. It is not known with precision how common this phenomenon is, nonetheless, in the human genome there are thousands of transcribed enhancers (Andersson et al., 2014). The RNA molecules thereby produced (eRNA) are short, unspliced RNAs, with small half-lives, degraded shortly after they are generated (probably by the exosome complex) and therefore invisible using the standard transcriptomics methods (Core et al., 2014a). Despite the lack of knowledge regarding its biological function, there is evidence suggesting that enhancer transcription may not be just a result of leaking RNA Pol II transcription on accessible DNA. Rather, enhancer transcription shares many common points with gene transcription, including the presence of core promoters motifs and highly positioned nucleosomes flanking the TSS, indicating that enhancer transcription may be a defined and regulated process (Core et al., 2014a).

2.1.3.3 Repressive Chromatin

In the *Drosophila* genome, three different kinds of repressive chromatin have been described, altogether accounting for almost 3/4 of the entire genome and containing almost half of the genes. Genes embedded in repressive chromatin are either not expressed or expressed at very low levels.

One kind of chromatin found in all eukaryotes is the classical heterochromatin, positive for the histone modification H3K9me3 and condensed by the protein HP1. H3K9me3 modification is catalyzed in multiple steps: first the lysine 9 of the histone H3 is mono and di-methylated by the histone-lysine N-methyltransferase G9a, then SETDB1 and SUV39H1/H2 catalyze the last methylation (J.-K. Lee & Kim, 2013; Schuhmacher, Kudithipudi, Kusevic, Weirich, & Jeltsch, 2015; Shinkai & Tachibana, 2011; Vaquero et al., 2007).

The second kind of repressive chromatin is the Polycomb associated chromatin, also present in all eukaryotes. Polycomb group proteins are essential for cell differentiation and development (Martinez & Cavalli, 2006; Pasini, Bracken, Hansen, Capillo, & Helin, 2007). They are found in two independent complexes that have slightly different mechanisms of repression, and catalyze two different histone modifications: H3K27me3, and the ubiquitylation of the histone H2A,

H2AK119ub. H3K27me3 modification completely covers promoters and transcribed regions of repressed developmental genes (Kalb et al., 2014). This modification cannot coexist, in the same residue, with the active histone mark H3K27ac (found at the promoter of every transcribed gene). Therefore it is thought that H3K27me3 has a repressive effect on transcription by counteracting the deposition of the acetylation mark (reviewed in (Schwartz & Pirrotta, 2013). H2AK119ub instead, mediates transcriptional repression by interfering with the transcriptional elongation (W. Zhou et al., 2008). The chromatin condensation occurs independently of both histone modifications and it is likely to be mediated by Polycomb proteins themselves (Eskeland et al., 2010).

The last repressive chromatin type is different from those previously described, as it does not contain the heterochromatin markers HP1 and H3K9me3, nor Polycomb group proteins and associated modifications. Instead it is enriched for the histone H1. Because of the technical challenge to perform histone H1 ChIP-seq in mammalian tissues (the few H1-ChIP-seq datasets available were obtained using tagged versions of the histone H1, (J.-Y. Li, Patterson, Mikkola, Lowry, & Kurdistani, 2012)), this chromatin type has been described mainly in *Drosophila*. In this organism it is the most abundant chromatin type and covers 48% of the genome. It seems, like in the case of the other repressive chromatin types, that it actively represses transcription and, contains silent genes that are active in only few tissues.

2.1.4 Genomic Imprinting

2.1.4.1 Imprinting Functions

Imprinting is defined as the epigenetic mechanism that causes the mono-allelic expression of a gene, dependent on the parental origin of the allele.

It was first discovered thirty years ago by the observation that in pronuclear transplantation experiments, both male and female genomes were necessary for correct embryonic development (McGrath & Solter, 1984).

Imprinting is found only in flowering plants and mammals and has independently evolved in these two *taxa* (Köhler, Wolff, & Spillane, 2012). In mammals it probably evolved together with the placenta (125 million years ago).

In mouse, 150 genes are imprinted and about half this number in human. However, the imprinted genes are not necessarily conserved between these two organisms (Prickett & Oakey, 2012).

More than 80% of the imprinted genes are found in gene clusters, in mouse there are 13 of them spread on 8 chromosomes (a complete list can be found at www.mousebook.org). Imprinted genes show clear enrichment with specific gene ontology categories including metabolism, thermoregulation and embryonic growth (Cannon & Nedergaard, 2004; Charalambous, da Rocha, & Ferguson-Smith, 2007; M. Chen et al., 2005, 2012; Constância et al., 2002; Guillemot et al., 1995; Kelly et al., 2009).

Its effects are seen predominantly on embryo-nourishing tissues such as placenta, adipose tissue and brain (M. Chen et al., 2005; Constância et al., 2002). In the latter case, it affects behaviors associated with feeding, for example, the neonatal milk suckling (Ball et al., 2013; Plagge et al., 2004). In 72% of the cases, imprinted genes are imprinted in multiple tissues, whereas in the remaining 28%, they are imprinted in a single tissue (Schulz et al.). In more than half of the cases, tissue-specific imprinted genes show mono-allelic expression in extra-embryonic structures, such as placenta (48%) and yolk sac (9%). In most of the remaining cases, tissue-specific imprinted genes are imprinted in the whole brain or in specific parts of the brain (Schulz et al., 2008).

The reasons why imprinting evolved are not known. Two hypothetical models coexist in the literature, the kinship model and the coadaptation model.

The kinship model proposes that maternal and paternal genes have different "interests" in the use of nutrients while the fetus is dependent on the mother. For the paternal genes in order to spread, it is convenient to maximize the intake of nutrient by the embryo from the uterus, even though this could be detrimental for the mother. Maternal genes instead, benefit from the opposite situation, where the little uptake from the fetus guarantees more pregnancies for the mother. According to this theory, often, imprinted genes that promote or inhibit growth are expressed from the paternal or maternal allele respectively (Isles, Davies, & Wilkinson, 2006).

The coadaptation model instead, is based on the fact that the fetus develops inside the mother and the two of them have a continuous communication/exchange. The model proposes that some genes (most of the imprinted genes are involved in growth) in the fetus are expressed from the maternal allele because they just fit better with the maternal environment (Wilkins, 2014). For example the pregnancy and delivery of a larger fetus may be difficult for a smaller mother.

2.1.4.2 Mechanisms of Imprinting

Imprinting is established on the male and female germline via CpG DNA methylation on the imprinting control region (ICR) of a gene cluster (Lees-Murdock & Walsh). These regions are referred to as germline differentially methylated regions (gDMRs). Some other allele specific methylation differences are acquired after oocyte fertilization, during early embryonic development. These regions are termed acquired differentially methylated regions (aDMRs) (S.-P. Lin et al., 2003; Peters et al., 1999; Redrup et al., 2009; Thorvaldsen, Duran, & Bartolomei, 1998).

During embryonic development, harsh genomic reprograming events are necessary to acquire totipotent traits, among these, the complete DNA demethylation of the paternal pronucleus (Iqbal, Jin, Pfeifer, & Szabó, 2011). It is a big question in the field what epigenetic signals allow the methylation of the ICRs to remain stable throughout development and adulthood despite these reprograming events.

The parent specific DNA methylation at the ICRs is responsible for the local establishment of a specific chromatin environment. This is characterized by the deposition of repressive histone modifications, among which are H3K9me3 and H4K20me3 (Henckel et al., 2009). Consistently, other studies showed that the methyl transferase activity of the G9a enzyme (responsible for the mono and di-methylation of the H3K9) is necessary for the imprinting maintenance on the trophoblast (Wagschal et al., 2008).

Maternal and paternal imprinted gene clusters show important differences.

Maternal imprinting is seen on ICRs that contain the promoter of a long non-coding RNA (lncRNA) spanning a protein-coding gene of the cluster, but transcribed in antisense direction. As their transcription is mutually exclusive, in the paternal chromosome, the lncRNA transcription blocks the transcription of the protein-coding genes of the cluster. On the other hand, in the maternal chromosome the lncRNA expression is silenced by DNA methylation, allowing the transcription of the protein-coding genes of the cluster. This mechanism is supported by the study of a few cases, including the *Kcnq1* and *Ube3a* clusters (figure 4A) (Mancini-Dinardo, Steele, Levorse, Ingram, & Tilghman, 2006; Sleutels, Zwart, & Barlow, 2002).

Paternal imprinting also involves lncRNA transcription, however, it tends to occur at intergenic regions, probably on insulator elements, without spanning protein-coding genes. The *Igf2-H19 locus* is one example of parental imprinting (figure 4B). In this *locus*, in the maternal chromosome, the binding of CTCF to the

ICR prevents an enhancer acting on the promoter of *Igf2*, which leads to the silencing of *Igf2*. From the ICR there is the transcription of the lncRNA *H19*, which is dispensable, for maintaining the insulating effect of the ICR. In the paternal chromosome the methylation of the ICR does not allow CTCF to bind, and as a result, the enhancer is able to activate *Igf2* transcription (Bell & Felsenfeld, 2000)(Hark et al., 2000).

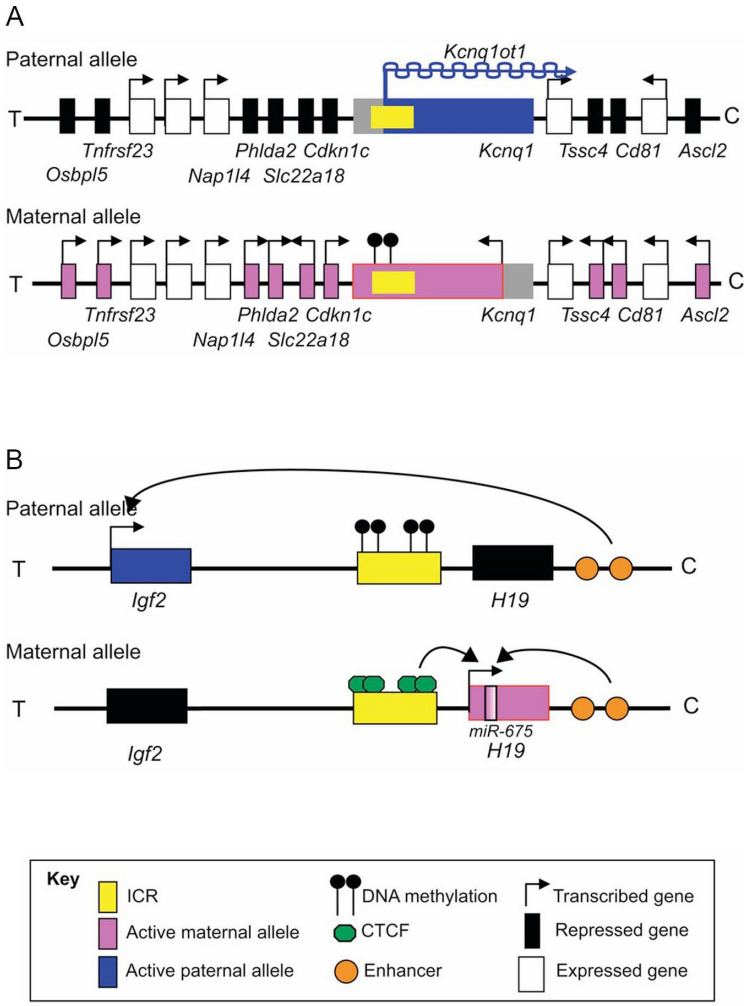


Figure 4 | Mechanisms of paternal and maternal imprinting. Adapted from (Plasschaert & Bartolomei, 2014). A) Maternal imprinting. At the *Kcnq1 locus* in the paternal allele the antisense transcription of the lncRNA *Kcnq1ot1* blocks the transcription of *Kcnq1.* In the maternal allele, the methylation of *Kcnq1ot1* promoter keeps *Kcnq1ot1* silenced. *Kcnq1* and other genes of the cluster acquire expression. B) Paternal imprinting. At the *Igf2/H19 locus,* in the maternal allele an enhancer (located downstream *H19* transcribed region) activate the expression of the non-coding gene *H19.* The binding of CTCF to the ICR (lying between *Igf2* and *H19*) confers insulating properties to the ICR. As a consequence, the enhancer is not able to activate the expression of *Igf2*. In the paternal allele, the methylation of the ICR prevents the binding of CTCF, and causes the ICR to lose its function as insulator. Because of this, the enhancer can activate *Igf2.*

2.2 *c-Myc locus*

2.2.1 MYC Roles as Transcription Factor

c-Myc gene was discovered more than thirty years ago, from chromosomal translocations in Burkitt's lymphoma cells. The analysis of the translocation breakpoint revealed a coding sequence similar to the myelocytomatosis viral oncogene *v-Myc*, and for this reason it was named *c-Myc* (Bishop, 1982; Sheiness, Fanshier, & Bishop, 1978).

MYC is a basic helix-loop-helix transcription factor, evolutionarily conserved in all vertebrates with a considerable amount of sequence similarity (Atchley & Fitch, 1995). It binds to thousands of promoters in mammalian cells as MYC-MAX heterodimer (Blackwood & Eisenman, 1991; C. Y. Lin et al., 2012). In particular it binds the motif CACGTG of the enhancer box (E-box) in the core promoter of active genes. Depending on the target gene, MYC can act as transcriptional activator or repressor, and, can affect transcription at both initiation and elongation steps (Rahl et al., 2010).

MYC mediates the transcriptional response of growth-factors stimulation. Importantly, MYC does not only regulate the expression of mRNA(s), it also regulates ribosomal and tRNA genes, transcribed by the RNA Pol I and RNA Pol III respectively (Campbell & White, 2014; Dai, Sun, & Lu, 2010; Mitchell et al., 2015). Amongst the major gene ontology categories of protein-coding genes under the control of MYC there are: ribosome biogenesis, apoptosis, cell adhesion, cell size, angiogenesis and metabolic pathways (Nieminen, Partanen, & Klefstrom, 2007; Peterson & Ayer, 2011; A. M. Singh & Dalton, 2009; Uslu et al., 2014; van Riggelen, Yetil, & Felsher, 2010).

Because of its influence on all these processes, MYC was shown early on to have a role in both development and cancer.

2.2.1.1 MYC Roles in Embryonic Development

c-Myc is essential for proper mouse development. Homozygotes null mutations cause lethality between 9.5 and 10.5 days of gestation, most likely due to early hematopoietic defects (Davis, Wims, Spotts, Hann, & Bradley, 1993). *c-Myc* null embryos have developmental delays compared to the littermates and are considerably

smaller in size. In addition, they also have a few phenotypes at the moment of death, including heart and neural tube abnormalities (Davis et al., 1993).

Heterozygote *c-Myc* animals are viable, but are smaller and have smaller organs than their littermates after birth. In particular, muscle, connective tissue, skin, bones, white fat and lymphoid organs are the most affected by the lower *c-Myc* levels. Fibroblast cultures showed that *c-Myc* heterozygotes cells have increased doubling time compared to Wt cells under the same conditions, but have the same size and cell cycle distribution, indicating that the reduction in body and organ weight in *c-Myc* heterozygote animals is due to reduced cell proliferation (Trumpp et al., 2001).

In *Drosophila* decreased *c-Myc* expression also lead to decreased body and organ weight. However, in contrast to mouse, this phenotype is due to reduction in cell size, more than cell number (Johnston, Prober, Edgar, Eisenman, & Gallant, 1999). It is possible that despite the fact that in both organisms *c-Myc* is involved in tissue and organism growth, there may be substantial differences in the extent *Drosophila* and mouse couple cell growth to cell proliferation (Trumpp et al., 2001).

2.2.1.2 MYC Roles in Cancer

The majority of the data in the literature regarding MYC, refers to the important role this protein has in cancer rather than in physiological conditions like development.

In contrast to other proto-oncogenes like RAS and BCR, MYC does not need to be mutated to promote tumor formation (although some cancers, like Burkitt's lymphoma do have mutated MYC forms) (Land, Parada, & Weinberg; Langdon, Harris, Cory, & Adams, 1986). Its only overexpression is a common feature in many MYC-driven cancers. About half of the human tumors show MYC overexpression, commonly, reaching more than two orders of magnitude compared to the physiological level (C. Y. Lin et al., 2012b; Rauen, 2013). Non-transformed cells typically enter apoptosis if MYC expression is forced to such levels (Askew, Ashmun, Simmons, & Cleveland, 1991).

The ways in which cancers benefit from such high levels of MYC are numerous, the most important probably being the enhanced cell growth and protein synthesis as well as the influence on the tumor environment by increasing the production and secretion of angiogenic factors (Baudino et al., 2002; Rosenwald, 1996).

2.2.2 *c-Myc* Transcriptional Regulation

Cell growth and proliferation are proportional to the amount of *c-Myc* transcripts. *In vitro*, quiescent cells have 10–40 fold less *c-Myc* mRNA than cycling ones (Facchini & Penn, 1998; Souleimani & Asselin, 1993). Removal of growth factors *in vitro,* brings *c-Myc* transcript to a basal level and has consequences at the level of growth and proliferation (Grandori, Cowley, James, & Eisenman, 2000; Henriksson & Lüscher, 1996; Lemaitre, Buckle, & Méchali, 1996).

During embryonic development, higher levels of MYC compared to terminally differentiated tissues, sustain the embryo growth and counteract cellular differentiation. Several signaling pathways at later stages are involved in the downregulation of *c-Myc*, the most important being TGF-β, WNT and BMP, (reviewed in J Liu & Levens, 2006). Amongst the main classes of TFs acting downstream of the signals there are LEF-TCF and SMAD(s) (Wierstra & Alves, 2008).

In adult tissues, mitogens, growth factors and antiproliferative signals like retinoic acid, vitamin-D, estrogen, PDGF, EGF, interferons and interleukins keep *c-Myc* expression under tight control. These pathways relay the signal to several TFs, including the estrogen receptor (ER), Stat(s), NFκB, Sp1–3, E2F(s), Rfx(s), AP1/FOS-JUN (reviewed in J Liu & Levens, 2006; Wierstra & Alves, 2008).

2.2.2.1 *c-Myc* Promoters

c-Myc is a housekeeping gene, and as commonly seen for this category of genes (and developmentally regulated ones), it has a high CpG content promoter, in this specific case forming a large CpG island (~ 3Kbp, containing 221 CpG di-nucleotides).

c-Myc transcription can start from two independent promoters 160bp apart from each other, P1 and P2 (with P2 being downstream of P1). Most transcripts initiate at P2 (Albert et al., 2001; Moberg, Logan, Tyndall, & Hall, 1992).

In addition to the signals conveyed by the discussed pathways, *c-Myc* transcription is also controlled by a large proximal-promoter. Ongoing transcription leads to accumulation of negative supercoiling, necessary for DNA unwinding. The 2Kbp upstream P1 contain elements capable of sensing this supercoiling and detecting the overall transcription rate. A series of non B-DNA sequences, including Z-DNA and G-triplex/quadruplex affect the local properties of the DNA polymer

and tunes *c-Myc* transcription by reacting to DNA coiling changes (Brooks & Hurley, 2009).

There are three sites in a Z-DNA shape nearby *c-Myc* TSSs, two of them are positioned upstream of P1 and one occurs on the junction between the first intron and the second exon. Z-DNA is a left-handed double helix formed by alternated purine–pyrimidine nucleotides, stiffer than B-DNA. The conversion of one turn of B-DNA into one turn of Z-DNA, possible at the Z-B-DNA junction, absorbs two negative supercoils and therefore could be a way to counteract the spreading of coiling stresses due to transcription (Wittig, Wölfl, Dorbic, Vahrson, & Rich, 1992).

Another highly evolutionarily conserved element, also located nearby the P1 core promoter, is the NHE element. It is placed 142–115 bp upstream of P1 and contains four imperfect repeats of the motif CCCTCCCCA. This sequence was shown *in vitro* to form G-triplex and quadruplex on the purine (G) rich strand and I-motif on the pyrimidine (C) rich strand (Mathad, Hatzakis, Dai, & Yang, 2011). The I-motif similarly to the G-quadruplex, is a four-stranded intra-molecular DNA secondary structure. DNaseI hypersensitivity assays showed that the NHE is a particularly accessible DNA, suggesting that it is usually in a nucleosome-free state (González & Hurley, 2010). Regarding its functions, there are a few hypotheses proposed so far. One, related to the function of the Z-DNA, is that G-quadruplex may counteract the propagation of supercoiling to the upstream promoter elements (like for the FUSE element) (Brutzer, Luzzietti, Klaue, & Seidel, 2010; Wada & Netz, 2009). Another possibility, probably the most important, is that NHE may affect *c-Myc* transcription through the several proteins and complexes it binds. A few examples are given by the hnRNP A1, A2, B and K and NUCLEOLIN (Brooks & Hurley, 2009; González, Guo, Hurley, & Sun, 2009; González & Hurley, 2010).

The Far Upstream Sequence Element (FUSE) element, located 1.7 Kbp upstream of the P1 promoter, was also shown to be involved in *c-Myc* regulation. Because of its AT rich sequence composition (characterized by a low melting temperature), it can open in the presence of negative supercoils accumulated during transcription. In the presence of basal transcription, it partially opens and binds the FUSE Binding Protein (FBP). FBP recruits the general transcription factor TFIIH, involved in transcription initiation and promoter escape, thereby increasing *c-Myc* transcription (Juhong Liu et al., 2001, 2006). This leads to further accumulation of negative supercoils and further opening of the FUSE element. At this stage, the FUSE Interacting Repressor (FIR) is recruited by FBP, and binds the ssDNA of the FUSE through a RNA Recognition Motif. FIR also interacts with the TFIIH and somehow reduces *c-Myc* transcription, with a consequent reduction of negative

supercoiling. As a result, the FUSE ssDNA re-annealing ejects both FIR and FBP, restoring *c-Myc* basal transcription (Cukier et al., 2010).

In conclusion, *c-Myc* proximal promoter, which is not sufficient to drive *c-Myc* expression in transgenic assays, is equipped with DNA sequences that can undergo structural changes (Lavenu, Pournin, Babinet, & Morello, 1994). These sequences seem to be important for *c-Myc* auto-regulation, ensuring that its expression levels do not exceed the physiological thresholds. This system may be particularly important in case growth-factor responding REs (located outside *c-Myc* proximal promoter) over-stimulate *c-Myc* transcription (Lavenu et al., 1994).

2.2.2.2 Long-Range Regulation of *c-Myc*

2.2.2.2.1 Genome Wide Association Studies

c-Myc gene lies in a gene-poor (or gene-desert) region of ~ 3Mbp, in the chromosome 8 in humans and in the chromosome 15 in mouse. In mouse *c-Myc locus* extends from the gene *Trib1* at the centromeric side, to the end of the *Gsdmc* gene-cluster at the telomeric side. In the whole gene-desert there are few annotated transcripts, *c-Myc*, *Fam84b*, *A1bg* and *Pvt1*. *Fam84b* and *A1bg* may be protein-coding genes, while *Pvt1* is a non-coding gene (Hsu et al., 2015). In addition to the annotated coding and non-coding genes, *c-Myc locus* has several evolutionarily conserved regions, scattered throughout the gene-desert. Their synteny is preserved in mammals as well as in chicken and zebrafish, in a few instances (figure 5). It is remarkable that despite the fact that some organisms are hundreds of millions of years apart on the phylogenetic tree, they have preserved the relative disposition of transcribed regions and conserved elements.

Genome wide association studies (GWAS) showed that throughout *c-Myc locus* in human, there are several linkage disequilibrium blocks of single nucleotide polymorphisms (SNP), with high susceptibility for genetic disorders, and different types of cancer (Amundadottir et al., 2006; Easton et al., 2007; Ghoussaini et al., 2008; Gudmundsson et al., 2007; Haiman et al., 2007; Kiemeney et al., 2008; Sebastiani, Timofeev, Dworkis, Perls, & Steinberg, 2009; Tomlinson et al., 2008; Zanke et al., 2007).

From the same GWAS, it was reported that in about 40% of cases there was no overlap between transcribed regions and linkage disequilibrium blocks, indicating that non-coding polymorphisms may account for a considerable portion of genetic

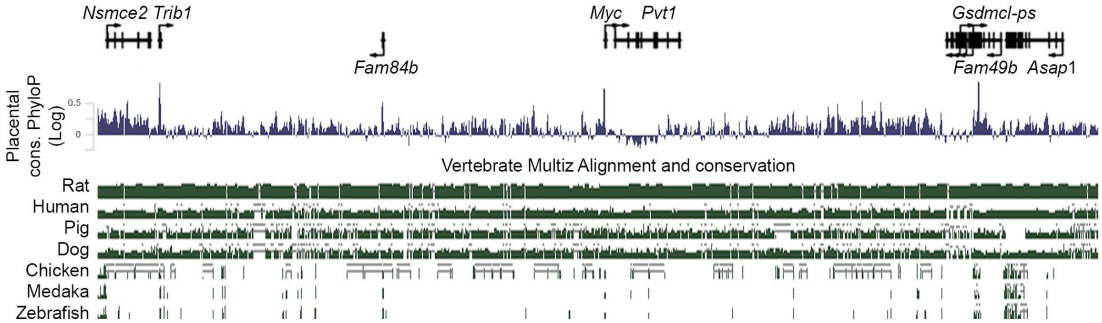


Figure 5 | *c-Myc locus* evolutionary conservation. Data available at: http://genome-euro.ucsc.edu/cgi-bin/hgTracks. The middle panel shows the degree of conservation in mammals. The lower panel shows the alignment of homolog sequences in other *species.* The figure indicates the presence of several homolog coding and non-coding sequences, conserved from mouse to zebrafish, and scattered throughout the *locus.*

diseases. One possible explanation could be that the observed polymorphisms occurred in regulatory elements, thereby affecting functional long-range regulation (reviewed in Visel, Rubin, & Pennacchio, 2009). According to this view, some of the linkage disequilibrium blocks in *c-Myc locus*, found hundreds of Kbp away from *c-Myc* promoter, could contain regulatory elements whose activity may differ between individuals due to polymorphisms (Wasserman, Aneas, & Nobrega, 2010).

GWAS studies in human showed that in the telomeric side of *c-Myc locus,* the 8q24.21 region delineates a linkage disequilibrium block of SNPs associated with the genetic disease cleft lip/cleft palate (CLP) (figure 6) (Amundadottir et al., 2006; Easton et al., 2007; Ghoussaini et al., 2008; Gudmundsson et al., 2007; Haiman et al., 2007; Kiemeney et al., 2008; Sebastiani et al., 2009; Tomlinson et al., 2008; Zanke et al., 2007). CLP is one of the most common genetic diseases in humans, observed in one out of seven hundred people. It is a multifactorial disorder, with many *loci* involved, including *Irf6, Vax1* and many others, in addition to 8q24.21. 8q24.21 however, was shown in one study to be the major susceptibility *locus* for CLP, with attributable risk of 0.41. It corresponds to a region of 640Kbp (chr8:129,700,000-130,340,000, NCBI36/hg18 assembly) containing multiple SNPs statistically associated with the disease, among which is "rs987525" (the one with the lowest *p*-value) (Birnbaum et al., 2009). This 640Kbp window has a corresponding ortholog in mouse, located on the telomeric side of *c-Myc locus* (chr15:62610410-63188813, NCBI37/mm9 assembly).

In another study, putative enhancers important during face morphogenesis were identified by p300 ChIP-seq on E11.5 mouse facial mesenchyme. Four enriched regions were included in the mouse ortholog of the 8q24.21 CLP susceptibility region, about 1.1Mbp telomeric to *c-Myc*. Two of these sequences (hs1877 and

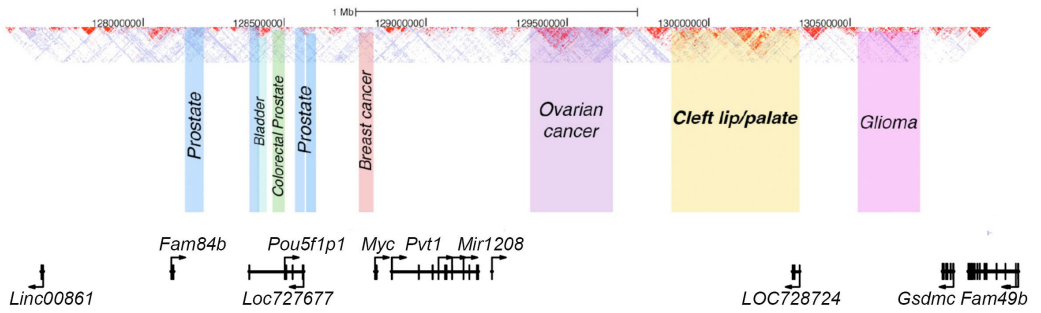


Figure 6 | Human genetic variation in *c-Myc locus*. Adapted from (Wasserman et al., 2010b). The picture shows a portion of the *c-Myc locus* (genomics coordinates: chr8:126800000-131000000, hg19 assembly). Each point in the heatmap represents the degree of linkage disequilibrium between two given SNPs. The colored rectangles emphasize the association of linkage-disequilibrium blocks with different pathologies.

mm458) were tested for enhancer activity by reporter assays and were shown to drive reporter expression in the facial mesenchyme and maxillary process respectively (Attanasio et al., 2014).

GWAS in human also showed that a region 440Kbp upstream of *c-Myc* contains five blocks of linkage disequilibrium associated with prostate cancer. In one study the whole region was analyzed for carrying prostate-specific regulatory elements (Wasserman, Aneas, & Nobrega, 2010). The authors tagged three overlapping human BAC clones with a LacZ reporter under the control of the core promoter of the *β-globin* gene. They injected the construct in fertilized mouse oocytes and evaluated the LacZ staining at different stages after birth. Two clones showed reporter expression in the prostate. They further restricted the region driving reporter expression to a 5Kbp sequence carrying the key polymorphism rs6983267-G. The staining pattern observed in the prostate was similar to the one of the endogenous *c-Myc*, as seen by RNA *in situ* hybridization. The sequence containing this susceptibility allele for prostate cancer, occurring about 400Kbp upstream of *c-Myc* promoter was proposed to be a prostate enhancer, perhaps acting on *c-Myc* in a long-range manner (Wasserman et al., 2010b).

2.2.2.2.2 Cancer Studies

A few data, obtained from cancer cell-lines suggest that not only long-range interactions are important for the endogenous promoter–enhancers communication, but also that if altered, for example by viral integration into the genome or by

chromosomal rearrangements (often seen in cancer), they can lead to ectopic gene expression.

One early example showing long-range regulatory behavior in *c-Myc locus* comes from the study of several T-cells lymphomas. A screening of rat thymic lymphoma obtained by infection of T-cells with the retrovirus Moloney murine leukemia virus (MoMuLV), produced several cell lines with the characteristics of T-cell lymphomas (Lazo, Klein-Szanto, & Tsichlis, 1990). The mapping of the integration site of MoMuLV in several leukemia lines thereby obtained, revealed that the majority of the integrations occurred in *c-Myc locus* at various distances from *c-Myc* promoter, the furthest located 270Kbp away (Clark, 1997; Lazo, Lee, & Tsichlis, 1990; Palumbo et al., 1990). These insertions were shown to increase *c-Myc* expression by long-range effect, thereby causing T-cell leukemias.

Human papillomaviruses (HPV16 and HPV18) are the major cause of cervical cancers seen worldwide (World Health Organization data). Even though the involvement of HPV in interfering with the endogenous gene expression has never been described, in the HeLa cell lines (cervical cancer) the HPV18 integrated in the genome in multiple places, two of which in *c-Myc locus* (8q23 and 8q24) (Ambros & Karlic, 1987; Joos, Haluska, et al., 1992; Popescu, DiPaolo, & Amsbaugh, 1987).

Burkitt's lymphomas provide an example of *c-Myc* misexpression due to interfering long-range activity. Often these cancers arise because of chromosomal translocations that fuse *c-Myc locus* with the *Immunoglobulin heavy chain* (*IgH*) *locus*. The mapping of the translocation breakpoints revealed a quite heterogeneous distribution in *c-Myc locus*, ranging from 14 to 340Kbp upstream of *c-Myc* promoter. In the *IgH locus*, the breakpoints occur within a small window, located close to the enhancer of the *IgH* gene. In this case it was proposed that the *IgH* enhancer interfered with *c-Myc* regulation, specifically in B-cells, causing *c-Myc* misexpression and consequently, leukemia (Joos, Falk, et al., 1992; Joos, Haluska, et al., 1992).

3 Aim of the Project

Eukaryotic transcription is a process regulated at different levels that can be summarized in three major categories: genetic, chromatin and topology. In the last decade, especially thanks to the development of next-generation-sequencing (NGS) technology, there has been a remarkable increase in the knowledge of each of these areas. From a genetic perspective, the improvement in annotation of transcripts and non-coding elements such as repetitive sequences, CpG islands, endogenous retro-transposons, regulatory elements, evolutionarily conserved regions and many others, have fueled genomics studies. In the chromatin field, binding sites of several transcription factors, as well as chromatin modifications, have been mapped into the genomes in many cell lines and tissues, showing that many histone modifications have different degrees of correlations with transcription. More recently it has been shown that chromosomes and *loci* are folded in a controlled fashion, and that probably, their topological organization helps sorting the regulatory elements that act on certain promoters.

Despite the vast amount of data suggesting that DNA sequence, chromatin and topology play an important role in transcriptional regulation, it is still not clear how all these components functionally integrate in order to carry out this process. The aim of this project is to understand basic mechanisms organizing promoter–regulatory elements communication. We plan to achieve this by dissecting the contribution of chromatin and DNA topology individually, on a model *locus*.

c-Myc locus represents an ideal system for this project because of the amount of information available regarding the genetic component. Veli Uslu showed that long-range enhancer(s) boost *c-Myc* expression selectively on facial tissues during mouse embryonic development, and are important to form the upper lip and the primary palate. Moreover, the analysis of the regulatory landscape of *c-Myc locus* in mouse embryos, also performed by Veli Uslu, suggests the presence of other enhancer(s) active on the liver. These enhancer(s) seem to exert their action from *c-Myc* itself to the end of *Pvt1* transcribed region. Importantly, their location does not overlap with the facial enhancer(s). The presence of two (sets of) remote enhancers in *c-Myc locus,* active during mouse development on two tissues, give us the possibility to tackle our question from two independent systems. In addition, the presence of several tissues

where *c-Myc* expression is kept to the basal level, provides us with important internal controls.

Regarding the experimental setup, our approach is based on combining chromatin analysis with targeted genetic engineering and functional assays. The advantage of this approach compared to genome-wide studies, which provide clues only from a correlative point of view, is that the genetic engineering allows us to experimentally test the current models regarding topological and regulatory domains, providing functional evidence rather than descriptive correlations.

We performed the chromatin analysis using the ChIP-seq method. We focused on two main goals with this approach. Firstly, map and identify important regulatory elements, using antibodies recognizing chromatin signatures for gene promoters, enhancers and architectural proteins. Secondly, map broad chromatin domains, along the *locus*, to describe its overall organization in different tissues and compare it with the different features of the regulatory landscapes that have been described by Veli Uslu. For the second purpose, we took advantage of what was reported by Filion and colleagues, namely that despite the vast number of chromatin proteins, histone and DNA modifications described in the literature, there are just few major types of chromatin. They also reported that there are a few key histone modifications that can be used as markers of each chromatin type (Filion et al., 2010).

The contribution of the topological organization of the *locus* to *c-Myc* regulation was assessed by three balanced chromosomal inversions. The advantage of inversions over duplications and deletions is that they provide a "clean" way to reshuffle the genetic elements like promoters, regulatory elements and boundaries without changing them in number. The inversions breakpoints were chosen in order to perturb the TAD arrangements in specific ways. Finally, functional assays such as LacZ reporter assay and transcript detection by qPCR were used as readouts to assess the effects of the topological perturbations introduced.

Because of the role of CTCF and Cohesin to mediate DNA-DNA interactions and in the maintenance of the chromosome topology, we decided to test to what extent their binding to the genome is influenced by the *locus* architecture. For this question, the mouse lines carrying chromosomal inversions were particularly useful because of the reshuffling of the genetic elements that they introduce. We firstly investigated the occupancy of CTCF in *c-Myc locus* in our tissues of interest, namely embryonic face and liver. We then assessed, on the same tissues, the effect of the genetic reshuffling due to the inversion on CTCF binding.

Finally, the telomeric end of the *locus* was reported by Veli Uslu to contain an imprinting control region (ICR), responsible for the silencing of some regions of the

locus in the maternal allele. We used the inversion mouse lines to test the effect of the genetic reshuffling on the imprinted regions. For the sake of clarity, the data concerning this phenomenon are presented at the end of the results chapter, and are discussed separately from the rest of the work. The material presented and discussed in the sections not treating the imprinting, for example the material relative to the reporter assays, refer to the paternal allele, which is not imprinted.

4 Materials and Methods

4.1 Materials

4.1.1 Equipment

4.1.1.1 Thermomixers

Product	Manufacturer	Description
Eppendorf thermomixer® compact	Eppendorf	Heating
Eppendorf thermomixer® comfort	Eppendorf	Heating, programmable
Eppendorf thermostat™plus	Eppendorf	Heating and cooling

4.1.1.2 Shakers

Product	Manufacturer	Description
Maxq 416 / 430 hp tabletop digital orbital benchtop shakers	Thermo fischer	Benchtop orbital shaker
Polymax 1040	Heidolph	Swinging benchtop shacker
SB3	Stuart	Speed and angle adjustable rotator

4.1.1.3 Incubators

Product	Manufacturer	Description
HyBaid	Thermo fischer	Oven with rotating system integrated.

shacke'n'stack		Speed and temperature adjustable.
Hibridization oven	Binder	Temperature adjustable
TW8	Julabo	Water bath

4.1.1.4 Pipettes

Gilson Pipetman classic™, p2, p10, p20, p200, p1000.

4.1.1.5 Vortexes

Product	Manufacturer	Description
Vortex Genie 2	Scientific Industries Inc	Speed adjustable vortex with plate/tube adapters
MS 3 basic Lab Shaker	Laboratory mixers	Speed adjustable vortex

4.1.1.6 Sonicators

Product	Manufacturer	Description
Bioruptor® Plus sonication device	Diagenode	Fully adjustable in power, cycles and duty cycles. Including 1.5ml tubes adapter.
S220	Covaris	Focused-ultrasonicator (500KHz).

4.1.1.7 Bioanalyzer

Product	Manufacturer	Description
2100	Agilent technologies	Including chip-priming station, chip vortexer and 2100 Expert software

4.1.1.8 Qubit

Product	Manufacturer	Description
Qubit™ 3.0 Fluorometer	Life technologies	Target specific fluorescence nucleic acids quantitation.

4.1.1.9 Centrifuges

Product	Manufacturer	Description
Microcentrifuge5424	Eppendorf	Benchtop centrifuge
Microcentrifuge 5415 R	Eppendorf	Benchtop cooling centrifuge
MultiFuge 2SRT	Thermo fischer	Cooling centrifuge for falcons and plates
Megafuge2.0R	Heraeus	Centrifuge for falcons

4.1.1.10 Microscopes

Product	Manufacturer	Description
Leica M80	Leica	Dissection stereo microscope
Leica M16	Leica	Dissection stereo microscope
Leica M16F	Leica	Dissection stereo microscope with camera

4.1.1.11 Balances

Product	Manufacturer	Description
TE3135-DS	Sartorius	Accuracy: 1mg
TE601	Sartorius	Accuracy: 10mg

4.1.1.12 Thermocyclers

Product	Manufacturer	Description
C1000 Thermal Cycler	Bio-Rad	Two independent blocks of 48 wells
C1000 Touch Thermocycler	Bio-Rad	One 96 wells block
S1000 Thermocycler	Bio-Rad	Four machine connected as “tetrablock”, maximal capability 2 x 96 wells blocks and 4 x 48 wells blocks
ABI7500 Light cycler	Applied biosystems	RT-qPCR
BioMark	Fluidigm	Microfluidic RT-qPCR

4.1.1.13 Nanodrops

Product	Manufacturer	Description
Nanodrop8000	Thermo fischer	Eight channels spectrophotometer
ND-1000	NanoDrop	One channel spectrophotometer

4.1.1.14 Magnets

Product	Manufacturer	Description
DynaMag™2	Invitrogen	Magnets for ChIP
DynaMag™5	Invitrogen	Magnets for ChIP

4.1.1.15 DNA electrophoresis

Product	Manufacturer	Description
PerfectBlue Maxi Gel System L	Peqlab	Small size electrophoresis chamber
PerfectBlue Maxi Gel System M	Peqlab	Medium size electrophoresis chamber
PerfectBlue Maxi Gel System S Plus	Peqlab	Large size electrophoresis chamber

4.1.1.16 Power supplies

Product	Manufacturer	Description
EPS301	GE healthcare	Power supply

4.1.1.17 Computing

Minimal resources: 6 cores, 30Gbytes RAM.

4.1.2 Reagents

4.1.2.1 Primers

ID	Name	Sequence	Application
426	SB-L3	AAGTAGATGTCCTAACTGACTTGC	Standard genotyping
429	SB-R3	TCCTAACTGACCTAAGACAGG	Standard genotyping
3180	196231 L	CCTGGAATCTCCTTTTGTTCAGG	Standard and inversion genotyping
3589	196231 R	ATCCAACACCCCCTTTCTGC	Standard and inversion genotyping
3033	194578 L	CAGGAGTTTGCCAATCAACAGTG	Standard and inversion genotyping
3034	194578 R	GAAAGCAAGTGGGGAAGTCAGAG	Standard and inversion genotyping
790	SB-179039 R	GTTCCTCCCAAGGTTCATGCTC	Standard and inversion genotyping
2432	*c-Myc* nonRepeat(L) 179039	GGTTCTCTCTCATGGAGTGTATCAGG	Standard and inversion genotyping
3	SB_Rir1	GTTTTGGCAAGTCAGTTAGGACATC	Inversion genotyping
6	SB_Lir1	TTTCATCACATTCCCAGTGGGTC	Inversion genotyping
5534	*Sdha*_MRNA_F	TTGCAGCACAGGGAGGTATC	Fluidigm RT-qPCR
5535	*Sdha*_MRNA_R	ATGCCATCTCCAGTTGTCCTC	Fluidigm RT-qPCR
5536	*Trib1*_MRNA_F	CGAGGAGAGAACCCAGCTTAGA	Fluidigm RT-qPCR

5537	*Trib1*_MRNA_R	ATCTGACAGCGCATCATCTTC	Fluidigm RT-qPCR
5538	*Asap1*_MRNA_F	ATGTGACGGCTGAGACATGAG	Fluidigm RT-qPCR
5539	*Asap1*_MRNA_R	TCTGGTCCGGCATCCGATT	Fluidigm RT-qPCR
5543	*Fam49b*_MRNA_F	GCCGCTCGCGAACCT	Fluidigm RT-qPCR
5544	*Fam49b*_MRNA_R	TGTCCTTCTATCTGGCGGC	Fluidigm RT-qPCR
5545	*Tbp*_MRNA_F	TGGCGGTTTGGCTAGGTTTC	Fluidigm RT-qPCR
5546	*Tbp*_MRNA_R	GAAATAGTGATGCTGGGCACT	Fluidigm RT-qPCR
5547	*Fam84b*_MRNA_F	TCCGCGCAGGTGGCTTTAG	Fluidigm RT-qPCR
5548	*Fam84b*_MRNA_R	TTACGCTGTGGACTCTCGTTC	Fluidigm RT-qPCR
1317	Fwd_SYBR_MmGusb-qPCR	CTCTGGTGGCCTTACCTGAT	Fluidigm RT-qPCR
1318	Rev_SYBR_MmGusb-qPCR	CAGTTGTTGTCACCTTCACCTC	Fluidigm RT-qPCR
3006	c*c-Myc*(1) qPCR - Fwd	CCCTAGTGCTGCATGAGGAGACAC	Fluidigm RT-qPCR
3007	c*c-Myc*(1) qPCR - Rev	CCACAGACACCACATCAATTTCTTCC	Fluidigm RT-qPCR
3012	*Pvt1* qPCR -Fwd	CTGAGGTGGAGGAAGTTGCCCTTG	Fluidigm RT-qPCR
3013	*Pvt1* qPCR -Rev	GGCCACCTCAATCAGGCAGTGTC	Fluidigm RT-qPCR
5571	*Hprt*_mRNA_F	CTTCCTCCTCAGACCGCTTTT	Fluidigm RT-qPCR
5572	*Hprt*_mRNA_R	CATCATCGCTAATCACGACGC	Fluidigm RT-qPCR
5559	*Nsmce2*_MRNA_F	GGGAGCTGAAGAAGCAATATGGTA	Fluidigm RT-qPCR
5560	*Nsmce2*_MRNA_R	ATTGGTTTGGCTCTGGGT	Fluidigm RT-

		CA	qPCR
5561	*E430025E21Rik*_MRNA_F	GTGTTCCTGCTGAAAGACCG	Fluidigm RT-qPCR
5562	*E430025E21Rik*_MRNA_R	CCAGAACTCTGGACCCTTAAAGT	Fluidigm RT-qPCR
5590	chr15-60988828-60989797_F	TCCGATTCCGATGCAGTTCC	Fluidigm ChIP
5591	chr15-60988828-60989797_R	TTAGAAGGCCACCATGCTTGT	Fluidigm ChIP
5592	chr15-63602886-63603631_F	CTTGCTGCTGGACAATGCTG	Fluidigm ChIP
5593	chr15-63602886-63603631_R	TGGCAGGATTGTGTGGTGAA	Fluidigm ChIP
5594	chr15-60591260-60592101_F	TATGCCACATGAGGGTCAGG	Fluidigm ChIP
5595	chr15-60591260-60592101_R	AGCTCTGCTCTCTGCTCTCT	Fluidigm ChIP
5596	chr15-63483813-63484536_F	GTCCTCCACACCCAAATGGA	Fluidigm ChIP
5597	chr15-63483813-63484536_R	CGGCTAAGAGGAGGAAACAACT	Fluidigm ChIP
5598	chr15-63572139-63572734_F	ACGGACTACAAGACCAGGCA	Fluidigm ChIP
5599	chr15-63572139-63572734_R	GGGCCCATAGACTTTCACCTC	Fluidigm ChIP
5600	chr15-61814519-61815473_F	GGATGACCGGAAGCTTGTCT	Fluidigm ChIP
5601	chr15-61814519-61815473_R	CTGCAGGGTTTCCAGGTTCT	Fluidigm ChIP
5602	chr15-63553022-63553963_F	AGACCTCCAGGTGGGAAGAT	Fluidigm ChIP
5603	chr15-63553022-63553963_R	TTCAGAGTTGGTGGTTGGCA	Fluidigm ChIP
5604	chr15-60933612-60934521_F	GGCTCTAACTGTGGGCATTCA	Fluidigm ChIP
5605	chr15-60933612-60934521_R	CACCTCCCCCACCCATCTA	Fluidigm ChIP

5606	chr15-63171736-63172525_F	GAGGCACATTGTGGAGGGAG	Fluidigm ChIP
5607	chr15-63171736-63172525_R	CACTGTTCTCTGGGAGAGCC	Fluidigm ChIP
5608	chr15-61816690-61817501_F	AAGAGCCGTGTGTGCAGAG	Fluidigm ChIP
5609	chr15-61816690-61817501_R	CTCACTCCAGAGCTGCCTTC	Fluidigm ChIP
5610	chr15-63432906-63433773_F	GTGACAGGGTATTGCCCAGC	Fluidigm ChIP
5611	chr15-63432906-63433773_R	ATGGCAAGCTCAGTCCAACT	Fluidigm ChIP
5612	chr15-63660558-63660911_F	ACATTCTGAACGCTAGGGGAC	Fluidigm ChIP
5613	chr15-63660558-63660911_R	TCCCTGTTGGTTCCCTTCCA	Fluidigm ChIP
5614	chr15-59928547-59929378_F	CGAGAACCCGCATGGTGAT	Fluidigm ChIP
5615	chr15-59928547-59929378_R	CATCATCACCCTGGCAGACA	Fluidigm ChIP
5616	chr15-62159709-62160443_F	TGGAATGCCACCGAAAAGGA	Fluidigm ChIP
5617	chr15-62159709-62160443_R	AAAGCCCCGAGGATTGTGAG	Fluidigm ChIP
5618	chr8-47020195_Pos1_F	CTTCCCGACCTCAAATCCCC	Fluidigm ChIP
5619	chr8-47020195_Pos1_R	AAATAGGGGTGAAGGCGGC	Fluidigm ChIP
5620	chr1-183819016_Pos2_F	CCGTGTCCACCCATCAACAT	Fluidigm ChIP
5621	chr1-183819016_Pos2_R	GCCCTCTGCTGGCTAATGTC	Fluidigm ChIP
5622	chr12-82889780_Pos3_F	CTGGCACAGCAAGACCATGA	Fluidigm ChIP
5623	chr12-82889780_Pos3_R	AACCCCAGAGAGGTGCCATA	Fluidigm ChIP
5624	chr17-	CCTTGGCTACATGGTGAG	Fluidigm ChIP

	29508031_Pos4_F	GT	
5625	chr17-29508031_Pos4_R	AGCCAAGAGCAGGCATACAG	Fluidigm ChIP
5626	chr1-84730497_Mid1_F	ACCTTGTTATTTTGCTCTCCACTC	Fluidigm ChIP
5627	chr1-84730497_Mid1_R	GACAACATCTGTAACTGCGTGTC	Fluidigm ChIP
5628	chr3-36036726_Mid2_F	CTGACACAGGACACAAGCATAAAG	Fluidigm ChIP
5629	chr3-36036726_Mid2_R	TCCAAACTGATACTGTCTCACACC	Fluidigm ChIP
5630	chr5-65394338_Mid3_F	AAAGTCCCAGTGGCTCTGTTAT	Fluidigm ChIP
5631	chr5-65394338_Mid3_R	GCCACTTCCAAAGCTTGACTTAG	Fluidigm ChIP
5632	chr10-100659186_Neg1_F	TAGCCAGCATTTCTGTTTCTAGG	Fluidigm ChIP
5633	chr10-100659186_Neg1_R	CAAACACCAGACATACTGCTTCAT	Fluidigm ChIP
5634	chr5-59680664_Neg2_F	TGCCGGGCATACAATGCTTA	Fluidigm ChIP
5635	chr5-59680664_Neg2_R	GGTGAGTGAGACCAGCGAAA	Fluidigm ChIP
5636	chr14-81658234_Neg3_F	TGTCAATTTTTGACCACTCCCAC	Fluidigm ChIP
5637	chr14-81658234_Neg3_R	TCATACTTGGAAACACAGGAGGT	Fluidigm ChIP
5638	chr16-53987613_Neg4_F	TGCCACCCTAGATGAGAGACTAA	Fluidigm ChIP
5639	chr16-53987613_Neg4_R	ACAACTTATCATGAGCCACACTATG	Fluidigm ChIP

4.1.2.2 Chemicals

All the chemical reagents were purchased from Merck or Sigma-Aldrich except for the listed ones.

Product	Supplier	Description
Phenol non stabilized: Chloroform: Isoamyl Alcohol 25 : 24 : 1	Applichem	Phenol-chloroform used for ChIP DNA purification
cOmplete Protease Inhibitor Cocktail Tablets	Roche	Protease inhibitor small molecules, used during ChIP

4.1.2.3 Enzymes

Product	Supplier	Description
Taq DNA polymerase	Home made	Taq polymerase used for the standard genotyping

4.1.2.4 Commercial Kits

Product	Supplier	Description
Qubit® dsDNA HS Assay Kit	Invitrogen	Quantifies DNA in the range 0.2-100ng/µl
Power SYBR® Green Master Mix	Applied biosystems	Master mix including enzyme for RT-qPCR
E-Gel® 48 Agarose Gels, 2%	Invitrogen	Gel electrophoresis for DNA size selection for ChIP-seq library preparation
High Sensitivity DNA Analysis Kits	Agilent technologies	Reagents for bioanalyzer. Compatible with DNA in the range of concentration 100pg/µl-50ng/µl.
NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina®	New England biolabs	Library preparation kit for ChIP-seq
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	New England biolabs	Multiplexing oligonucleotides, compatible with NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina®
ProtoScript® II First Strand cDNA Synthesis Kit	New England biolabs	cDNA synthesis kit, compatible with Applied biosystems and Fluidigm
Expand Long Template PCR System	Roche	Long-range PCR, used for inversion genotyping

4.1.2.5 ChIP Antibodies

Product	Vendor	Description
ab8895	Abcam	Anti-Histone H3 (mono methyl K4) antibody - ChIP Grade
ab4729	Abcam	Anti-Histone H3 (acetyl K27) antibody - ChIP Grade
ab6002	Abcam	Anti-Histone H3 (tri methyl K27) antibody [mAbcam 6002] - ChIP Grade
ab4174	Abcam	Anti-Histone H2A.Z antibody - ChIP Grade
ab8898	Abcam	Anti-Histone H3 (tri methyl K9) antibody - ChIP Grade
07-729	Millipore	Anti-CTCF Antibody
ab992	Abcam	Anti-RAD21 antibody - ChIP Grade
sc-899	Santa Cruz	Pol II Antibody (N-20): sc-899

4.1.2.6 Magnetic Beads

Product	Supplier	Description
Dynabeads® Protein A for Immunoprecipitation	Invitrogen	Magnetic beads for ChIP
SPRIselect reagent kit	Beckman Coulter	Magnetic beads for ChIP-seq library adaptor cleanup

4.1.3 Consumables

4.1.3.1 Purification kits

Product	Supplier	Description
Quiaquick PCR purification kit	Quiagen	Column DNA purification
Minelute PCR purification kit	Quiagen	Column DNA purification. Used for ChIP-seq library preparation
RNeasy Plus Mini Kit	Quiagen	Whole RNA extraction kit (column based).

4.1.3.2 Cell Culture

Product	Supplier	Description
Nunc™ Cell Culture/Petri Dishes	Nunc	56.7 and 150 cm^2
DMEM	Gibco® Cell Culture Media	Low glucose medium
BW12-492F	Lonza bio whittacker	Fetal Serum Bovine (FBS)
Trypsin-EDTA (0.25%), phenol red	Gibco	Trypsin

4.1.3.3 Embryo Dissection

Product	Supplier	Description
Greiner petri dish	Greiner	60 x 15 mm, 94 x 16mm petri dishes
#5	Dunmont SA	Dissection forceps

4.1.3.4 Tubes

Product	Supplier	Description
Flex Tubes®	Eppendorf	Standard centrifuge tube. 0,5, 1,5, 2 ml
Eppendorf Safe-Lock Tubes	Eppendorf	Centrifuge tube suitable for nitrogen freezing. 1,5 ml
DNA LoBind Tubes	Eppendorf	Centrifuge tube for ChIP and low concentration samples. 1,5 and 2 ml
Greiner centrifuge tubes	Greiner	Falcon tubes. 15 and 50ml

4.1.3.5 Fluidigm Chips

Product	Supplier	Description
Flex Six IFC	Fluidigm	6 compartments of 12 samples x 12 primers.
96.96 IFC	Fluidigm	96 samples x 96 primers

4.1.3.6 Bioanalyzer Chips

Product	Supplier	Description
Agilent DNA 1000 Kit	Agilent technologies	

4.1.3.7 PCR Plates

Product	Supplier	Description
MicroAmp™ Fast Optical 48-Well Reaction Plate	Applied Biosystems	100µl 96 wells plate. Used for ChIP quality control.
Natural 96 well non-skirted PCR plate	Greiner	200µl 96 wells plate. Used for standard genotyping.

4.2 Methods

4.2.1 Genome Regulatory Organization Mapping With Integrated Transposons (GROMIT) System

GROMIT is a reporter system based on a LacZ gene under the control of the *β-globin* core promoter. It is equipped with a *LoxP* site (upstream of the β-globin:::LacZ cassette) that can be used for genetic engineering purposes. The whole construct is inserted into a "sleeping beauty" (SB) transposon (Tc-1 like fish transposon). Because of this feature, GROMIT can be remobilized to new genomic positions using the specific transposase HSB16. This transposase was obtained by genetic engineering and has higher remobilization efficiency than the original SB transposase. It was shown in mammals to effectively mediate the transposition in a cut and paste manner (Luo, Ivics, Izsvák, & Bradley, 1998). The expression of the HSB16 transposase is controlled by the *Protamine1* (*Prm1*) promoter (active only during spermatogenesis), and for this reason, the remobilization occurs only on the male germline. Male mice carrying both GROMIT SB-transposon and the Prm1:::HSB16 transgene produce spermatocytes carrying the SB-transposon in new

locations. These new insertions can be obtained as lines, and fixed very simply, by crossing these males with Wt females. The HSB16 has the tendency to remobilize the SB transposon to locations close to the parental insertion ("local hopping" (Keng et al., 2005)). This feature is useful because it allows the fine mapping of a given *locus* once a starting insertion is obtained.

4.2.2 Balanced Chromosomal Inversions

The chromosomal inversions are performed using the Cre-LoxP system, according to the STRING method (described by Spitz, Herkenne, Morris, & Duboule, 2005). The two LoxP required must be in the same chromosome and in opposite orientation. In this configuration the Cre enzyme recombines the DNA at the LoxP inverting the sequence in between. The LoxP are obtained from GROMIT insertions and must span the region of interest to be inverted. Their arrangements are listed in the table below.

Inversion name	Centromeric insertion breakpoint	Telomeric insertion breakpoint	mm9 coordinates (chr15)	Veli Uslu's defined coordinates
Inversion 1	196231 (+)	194578 (-)	60133316-62168343	c16–3a
Inversion 2	196231 (+)	179039 (-)	60133316-63550550	c16–17a
Inversion 3	194578 (-)	192857 (+)	62168343-63550550	3a–17a

The whole process takes four generations. In order to have both GROMIT insertions in the same chromosome, it is necessary to first generate an animal having the two insertions in the sister chromosomes. In case a meiotic-crossingover happens in the region in between, the two insertions are brought into the same chromosome. Once this animal is obtained, it is bred with an animal carrying a constitutively expressed Cre recombinase (Hprt:::Cre), in order to obtain pups carrying both alleles (it is more convenient, for the following breeding, to screen for a male double positive, rather than a female). At this stage the CRE enzyme recombines the DNA, and according to the efficiency of the process, a different percentage of the gametes will be recombined. The double positive male is bred against Wt C57BL/6J females, from this breeding, mice positive for the inversion and negative for the Cre recombinase are sought. Because of the fact that the sperm is considerably smaller than the oocyte, it likely carries little amount of CRE enzyme, and therefore will not interfere later on, at somatic level. The inversion leaves the primary sequences of GROMIT insertions (inversion's breakpoints) unaltered.

4.2.3 GROMIT Removal From Inversion Lines

The inversion lines carry both sleeping beauty transposons at the breakpoint. The removal of either of the two is done using the transposase HSB16. HSB16 is placed under the control of the *Protamine1* promoter (*Prm1*) that guarantees the transposase expression only during spermatogenesis. In the first generation, a male double positive for the inversion allele and the HSB16 transposase must be created. According to the remobilization efficiency (different for each GROMIT insertion), this male has a certain percentage of spermatocytes where one of the two transposons is remobilized to a new location or is just removed without being reintegrated. The double positive male is bred with Wt C57BL/6J females and the pups are screened for the lack either of the two transposons. At this stage, whether the missing transposon has reintegrated or not in the genome can be determined by either qPCR, determining the number of total transposons, or by further breeding with Wt animals. In the last case the presence in the litter of pups positive for GROMIT, but negative for the remaining parental insertion, indicates that the transposon missing from the inversion breakpoint, reintegrated somewhere else. However in the case no unmapped transposon is found in the litter, the lack of reintegration, or the reintegration in the same chromosome cannot be distinguished. Because of the fact that the HSB16 has the tendency of remobilizing the sleeping beauty transposon to locations close to the original one, the eventuality of a remobilization within the chromosome must not be neglected. For these reasons the qPCR method is chosen whenever possible.

4.2.4 DNA Extraction From Mouse Tails

The sliced fragments of tails are lysed over night (ON), at 55°C, 1000 rpm shacking, in 200µl of Tail-lysis-buffer for southern-blot (100mM Tris pH 8.5, 5mM EDTA, 20% SDS 0.2%, 200mM NaCl, in H_2O), plus 100µg/ml Proteinase K (PK). The tubes are spun for 1 minute and the supernatant is transferred to a new tube. An equal volume of isopropanol and 30µl of NaCl 5M are added in each tube. The tubes are shacked, incubated at 4°C for 10 minutes and spun for 10 minutes > 13000rpm on a benchtop centrifuge. The supernatant is discarded and washed with 70% ethanol by spinning for 10 minutes > 13000 rpm on a benchtop centrifuge. The pellet is resuspended in 300µl H_2O, warmed for 30 minutes at 37°C with mild shacking and stored at -20°C.

4.2.5 DNA Extraction From Extraembryonic Membranes

E11.5 yolk sacs are lysed in 200µl of Quick-tail-lysis-buffer (50mM KCl, 10mM Tris pH 8.00, 2mM $MgCl_2$, 0.01% Gelatin, 0.0045% Igepal, 0.0045% Tween-20, in H_2O) plus 100µg/ml PK ON, at 55°C, 1000rpm shacking. The PK is inactivated by 10 minutes incubation in 95°C, 1000rpm shacking. The solution is diluted 5 times in H_2O, spun 2 minutes >13000rpm on a benchtop centrifuge to clear it from the debris, and stored at -20°C.

4.2.6 Standard Genotyping

Each reaction is carried out in a total volume of 20µl with 0.4µl of home made Taq-DNA-polymerase, 0.2mM dNTPs (PeqGOLD dNTP set from Peqlab), PCR buffer (50mM Tris, pH 9.5, 15mM $(NH_4)_2SO_4$, 1,17mM $MgCl_2$, in H_2O), 250µM primers and 2µl DNA solution, typically around 100µg/µl concentrated. The reaction is performed with the following conditions.

Temperature (°C)	Duration (seconds)	Description
95	180	Step1, initial denaturation
95	20	Step2, denaturation
60	20	Step3, annealing
72	72	Step4, extension
		Step5, go to step 2, 34 additional times
72	300	Step6, dA adition (not necessary)
16	∞	Step7, hold

The primers combinations for the genotyping of each allele are listed in the table below. Every sleeping beauty allele is genotyped from both sides, left and right (L and R). In both cases one primer anneals in the sleeping beauty (always #426 for the left and #429 for the right), while the other one on the genomic DNA.

Allele	Forward (database ID)	Reverse (database ID)
SB-196231 (L)	426	3180
SB-194578 (L)	2033	426
SB-192857 (L)	426	790
SB-179039 (L)	2432	426
SB-196231 (R)	3589	429

SB-194578 (R)	429	3034
SB-192857 (R)	2432	429
SB-179039 (R)	429	790

4.2.7 Genotyping for Inversions

The genotyping for inversion is performed using the commercially available kit Expand long range template from Roche, according to the manufacturer's specifications.

Temperature (°C)	Duration (seconds)	Description
92	120	Step 1, initial denaturation
92	10	Step 2, denaturation
60	15	Step 3, annealing
68	300	Step 4, extension
		Step 5, go to step 2, 9 additional times
92	10	Step 6, denaturation
60	15	Step 7, annealing
68	320	Step 8, extension
		Step 9, go to step 6, 24 additional times
68	420	Step10, dA adition (not necessary)
16	∞	Step 11, hold

The primers combinations are listed in the table below.

Allele	Breakpoint	Forward (database ID)	Reverse (database ID)
Inversion 1	Centromeric	3180	3034
Inversion 2	Centromeric	3180	2432
Inversion 3	Centromeric	3033	2432
Inversion 1	Telomeric	3033	3589
Inversion 2	Telomeric	3589	790
Inversion 3	Telomeric	3034	790
SB-196231 Wt		3180	3589
SB-194578 Wt		3033	3034
SB-192857 and SB-		790	2432

179039 Wt			
Transposon (positive control)		3	6

The genotyping is performed on GROMIT positive, Hprt:::Cre negative pups. PCR for both breakpoints, from the genomic sequences encompassing the GROMIT transposon are done. The centromeric (genomic) primer of the centromeric breakpoint is used in combination with the centromeric (genomic) primer of the telomeric breakpoint. A second PCR is performed for the remaining breakpoint, using the telomeric (genomic) primer of the centromeric breakpoint in combination with the telomeric (genomic) primer of the telomeric breakpoint. Positive animals for long-range PCR on both breakpoints are considered as positive for the inversion. In case the LoxP, Hprt:::Cre double positive parent is female, additional PCRs spanning the Wt GROMIT insertions, on both breakpoints, must be done in order to rule out the possibility of chimerism. This is done because despite the Hprt:::Cre negative pups, the mother being positive for this allele can produce oocytes containing CRE protein, and therefore, the activity of the enzyme during early developmental stages may give rise to chimeras. Animals positive for long-range PCR product, for both the Wt and the inverted configuration are considered chimeras.

4.2.8 MEF Culture

CD1 Wt E13.5 embryos are decapitated and cleared from all the internal organs. The remaining carcasses are minced, deepened in 5ml trypsin and incubated 5 minutes at 37˚C in a water bath. The activity of the trypsin is blocked by addition of 5ml of culture medium (DMEM, 10% FBS). The dissociated cells are spun at 430g for 3 minutes at room temperature (RT), resuspended in 12ml of culture medium and plated in a 56.7cm^2 petry-dish. The cells are split every 2 days with 1:5 dilutions. The cells are harvested and processed according to the ChIP procedure until sonication. MEF sonicated chromatin can be stored at -80˚C indefinitely and is used for antibody testing.

4.2.9 Embryo Dissection

Pregnant females are sacrificed by cervical dislocation at 11.5 days of gestation (E11.5), the date of the vaginal plug is considered as day–0. The uterus is collected and plunged in Phosphate buffered saline (PBS) (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, in H_2O). The embryos, together with the extra-embryonic membranes, are separated from the uterus. The membranes are collected and stored in -20°C, and can be further processed for DNA extraction. The embryos are subjected to downstream processing according to the procedure.

4.2.10 LacZ Staining

The embryos are collected as described above, with the exception that the $CaCl_2$ and $MgCl_2$ are added to the PBS, to a final concentration of 1 and 0.5mM respectively. The embryos are fixed for 30 minutes in 4% formaldehyde (w/v) in PBS at 4°C with gentle shaking. They are washed in PBS (complemented with $CaCl_2$ and $MgCl_2$) for 10 minutes one time at 4°C, and one time at RT. Finally, the staining is performed by LacZ-staining-solution (0.01% Na Deoxycholate, 0.02% Igepal, 2mM NaCl, 10mM $K_4Fe_2(CN)_6$, 10mM $K_3Fe_3(CN)_6$, 1% spermidine and 2mM X-Gal in DMSO, in H_2O), in a dark humidified chamber at 37°C, ON. The embryos are washed twice in standard PBS at RT, and are stored indefinitely in 2% formaldehyde in PBS at 4°C.

4.2.11 Real Time Quantitative PCR (qPCR)

qPCRs for all the primer testing and for ChIP quality control, done before library preparation, are performed using the StepOne™ kit from Applied Biosystems (thermocycler and software, StepOne™ PCR software 2.0). The reaction is done in 15µl total, 7,5µl Power SYBR® Green Master Mix from Applied Biosystems (2X), 2µl DNA sample, 250nM primers, with the following cycling phases.

Temperature (°C)	Duration (seconds)	Description
95	600	Step 1, initial denaturation
95	15	Step 2, denaturation
60	60	Step 3, annealing + extension + fluorescence acquisition

		Step 4, go to step 2, 39 additional times
		Optional (amplicon melting curve calculation)
95	15	Step 5, denaturation
60	60	Step 6, annealing + extension
		Step 7, ramp 0.3°C/s + fluorescence acquisition
95	15	Step 8, denaturation + fluorescence acquisition

4.2.12 *c-Myc* and Neighboring Transcripts Quantification

4.2.12.1 Experimental Setup

The experiment is performed on the lines $INV1^{-/-}$, $INV1^{-/+}$, $INV1^{+/+}$, $INV2^{-/-}$, $INV2^{-/+}$, $INV2^{+/+}$, $INV3^{-/-}$, $INV3^{-/+}$ and $INV3^{+/+}$. The analyzed tissues are E11.5 facial mesenchyme (FM), proximal fore limb mesoderm (PFLM), heart and liver. For each sample type, three independent biological replicates are analyzed, and are referred together as a biological group.

For INV1 and INV2 (all established on C57BL/6J genetic background) heterozygotes males and females are crossed in order to obtain all the genotypes (Wt, heterozygotes and homozygotes). Despite the fact that the C57BL/6J line is inbred, and has little amount of polymorphisms, we decided to minimize even further the difference in genetic background between the comparison groups, this, in order to reduce any litter effects as much as possible. This effect may be seen, for example, in case all the homozygotes samples come from one litter and all the Wt samples come from another one, and as a result, this can confound the analysis. For this reason, we decided to balance the litter provenience of the samples between the three genotypes. Moreover, the tissue specimens are taken from embryos whose somite count differs at most by two (corresponding to about four hours of embryonic development).

4.2.12.2 RNA Extraction From Embryonic Tissues

The RNA is purified from embryonic tissues using the RNeasy mini kit from Quiagen, according to the manufacturer's specification and choosing the following options. Proximal fore limb mesoderm (PFLM), facial mesenchyme, liver and heart tissues are collected from E11.5 mouse embryos, placed in safe lock 1.5ml tubes, snap frozen in liquid N2 and stored at -80°C. As an exception from the original protocol, regardless of the tissue type and size, the tissues are always resuspended in

350µl of RLT buffer. The tissues are homogenized for 45 seconds using an electric mortar and pestle. The optional DNase treatment is performed on column, as indicated in the manual. The RNA is eluted from the columns with 30µl of RNase free water in two subsequent times. The expected total RNA yield varies from tissue to tissue, for liver and PFLM is about 4.5µg, for facial mesenchyme 4µg, and for heart 1.5µg. DNA contaminations are not detected by qPCR using primers specific for genomic DNA.

4.2.12.3 Reverse Transcription

The first strand cDNA synthesis is done with the NEB ProtoScript II kit according to the manual. The RNA is quantified for all the samples by nanodrop spectrophotometer. 200ng are used for the cDNA synthesis. The standard procedure suggests including in the experiment a negative control reverse transcription, done by replacing the ProtoScript II enzyme with H_2O, for each sample. The procedure is split into two batches, processing half of the total samples at a time, as an exception, the negative control is done only for two randomly chosen samples, for each batch. The reaction uses random hexamers as primers, is set in 200µl 8 wells RNase free PCR strips and proceeds with the listed cycling phases (the option standard protocol is chosen instead of the easy protocol).

Temperature (°C)	Duration (minutes)	Description
65	5	Step 1, initial RNA denaturation (with only water, RNA and random hexamers)
		Step 2, spin briefly and put on ice. Add the reaction buffer and the ProtoScript II enzyme
42	5	Step 3, pre-synthesis incubation
42	60	Step 4, denaturation and cDNA synthesis
80	5	Step 5, reverse transcriptase inactivation

4.2.12.4 Revers Transcription qPCR (RT-qPCR)

The DNA concentration of the cDNAs does not need to be quantified. The samples transferred into 96 wells plates and are handed to the GeneCore facility at EMBL. The samples are subjected, without dilution, to the standard Fluidigm sample preparation and downstream qPCR. The reactions are set in 2 chips Flex Six IFC, using ten independent compartments. Every compartment is loaded with eleven samples and one technical negative control (water). The chips are run in a Fluidigm Biomark machine using EvaGreen fluorescent dye and the following cycling conditions.

Temperature (°C)	Duration (seconds)	Description
95	60	Step 1, initial denaturation
96	5	Step 2, denaturation
60	20	Step 3, annealing + extension + fluorescence acquisition
		Step 4, go to step 2, 29 additional times
60	3	Step 4, ramp start point
To 95		Step 5, ramp: 1°C / 3s + fluorescence acquisition

4.2.12.5 RT-qPCR Data Analysis

The principle of the qPCR analysis is to compare the threshold cycle (Ct) of target genes, to the reference one(s). The expression level of the targets is expressed as multiple of the expression of the reference(s). The reference genes are selected on the basis of being expressed in the analyzed tissues, to have low variability in expression levels from sample to sample, and to be unaffected by the experimental treatments (in this case genetic engineering) applied to the animals (assumption). In this experiment, the target transcripts, object of the analysis are: *E430025E21Rik*, *Nsmce2*, *Trib1*, *Fam84b*, *c-Myc*, *Pvt1*, *Fam49* and *Asap1*. The reference transcripts are: *Gusb*, *Hprt*, *Sdha*, *Tbp*. The method applied for the comparison is similar to the ΔCt method as described in (Schmittgen & Livak, 2008) (equation 2), with a small adaptation.

Equation 1: $\Delta Ct = CtX - CtR$

Where *X* represent the target gene and *R* the reference one.

Equation 2: $X = 2^{\wedge(-\Delta Ct)}$

The method assumes that the primers efficiency is exactly 1 (100% efficient, every cycle the number of product doubles) for both targets and references. To avoid this inaccuracy, the method used for the quantification in this project is described by the equation 3.

Equation 3: $$X = \frac{(1+EffR)^{\wedge CtR}}{(1+EffX)^{\wedge CtX}}$$

Where *X* represent the target gene, *R* the reference gene and *Eff* the efficiency of the primers (from 0 to 1).

However, all the selected primers have efficiency close to 1, and the introduced error is negligible. Moreover, as the comparison is done for a given target in different genotypes, the error would not lead to any bias, as it is the same in all the compared genotypes. Therefore the standard ΔCt method can also be used.

There are two options regarding how to proceed in case several references are used. The first strategy is to average the Ct of all the reference genes and use this mean as reference Ct. The second strategy, the one used in this study, is to choose as reference, the gene whose expression levels are most stable across samples, and, do not differ significantly across different biological groups.

To quantify expression level of each reference, the equation 3 is used, the reference Ct is given in this case by the mean of the other reference genes. The coefficient of variation (Cv) is estimated for each reference gene within each biological group as described by the equation 5. The reason why equation 5 is chosen instead of equation 4, is that for small sample sizes (3–4 in this study) it better approximates the Cv. Cv estimates of all biological groups are summed for each reference, obtaining a total Cv, equation 6.

Equation 4: $$Cv = \frac{s}{\bar{x}}$$

Where $\bar{x}$ represents the estimate of the sample mean and *s* the estimate of the sample standard deviation.

Equation 5: $$Cv^* = \left(1+\frac{1}{4n}\right)Cv$$

Where n is the sample size.

Equation 6: $$totalCv = \sum_i Cv_i$$

Where Cv_i represents the Cv estimate for the i^{th} biological group.

To test whether the expression levels of reference genes differ between biological groups, and therefore introduce bias in the estimate of the targets, an ANOVA test is applied (using functions from the stats package in R). The test works by comparing the variance between groups with the variance within groups. The ANOVA is applied to each reference gene, setting the significance level threshold to 0.05. Higher p-values, calculated from the F-distribution, indicate insufficient evidence that the expression values in the biological groups differ.

Once the best-keeper gene is chosen, the enrichment compared to the reference gene can be calculated. The hypothesis that Wt and homozygotes inversion samples come from the same normal distribution is tested by the standard two-tailed *t-test*, for each combination of line and tissue (stats package in R). The *p*-values for each comparison group (for example INV1-face) are corrected using the Benjamini-Hochberg method (stats package in R).

4.2.13 ChIP

Facial mesenchyme and liver tissues are collected from the embryos at E11.5 and fixed in a 1.5ml tube (up to 10 slices of tissues) with 1% formaldehyde in PBS at RT for 10 minutes. The formaldehyde is quenched with the addition of glycine to a final concentration of 125mM, 5 minutes on ice. The tissues are washed in PBS once and can be stored in -80°C. The tissues can be pooled at this stage according to the needs (not more than 30 slices per tube, see table below for more precise estimate) and are lysed in 1ml of lysis-buffer-A (10mM Hepes, 10mM EDTA, 0.5mM EGTA, 0.25% Triton X 100, in H_2O), for 10 minutes at 4°C. Afterwards, the lysed tissues are washed in lysis-buffer-B (10mM Hepes pH8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.01% Triton X-100, in H_2O) for 10 minutes at 4°C. The pellet is resuspended in 300µl of sonication buffer (10mM Tris-HCl pH 8, 1mM EDTA ph 8, 0.1% SDS, in H_2O) complemented with protease inhibitor cocktail from Roche, and sonicated in one round on a Bioruptor plus from Diagenode, setting the machine on high power, for 24 cycles 30 seconds on and 30 seconds off, constantly

at 4°C. If larger volume of sonication buffer is required the sonication is performed in multiple rounds each with at most 300µl of sonication buffer. The chromatin is pooled (if sonicated in multiple rounds) and spun for 5 minutes at >13000rpm on a benchtop centrifuge at 4°C, to get rid of the cell debris. The supernatant is collected in a new tube and if needed, can be stored indefinitely at -80°C. The chromatin is thawed on ice (if frozen), and the volume is estimated with a pipette. The buffer composition is equilibrated to RIPA buffer by addition of one ninth of the volume of sonication-to-RIPA-buffer (10X) (10mM Tris-HCl pH8, 1mM EDTA, 10% Triton X100, 0.1% SDS, 1% NaDeoxycholate, 1.4M NaCl, in H_2O). 2% of the total material is taken as input chromatin and is stored at 4°C until indicated. The remaining chromatin is diluted from 2 to 10 times in RIPA buffer complemented with protease inhibitors from Roche. The antibody is added to the chromatin and the tubes are incubated ON on a rotator at 4°C. The immunocomplexes are captured by the addition of 20µl of magnetic beads conjugated with Protein A, resuspended in RIPA buffer. The slurry is left in rotation at 4°C for other 5 hours, in order for the beads to capture most of the immunocomplexes. Before the first wash, the beads are immobilized on the side of the tube with a magnetic rack and 100µl of flow-through chromatin is collected. The beads are washed three times in RIPA buffer, one time in RIPA buffer 500mM NaCl, and one time in TE buffer (10mM Tris-HCl pH8, 1mM EDTA pH 8, in H_2O) leaving the tube in rotation at 4°C for a few minutes every wash. After the last wash, the tube is changed with a clean one and the immunocomplexes are eluted with 95µl of elution buffer (10mM Tris-HCl pH8, 1mM EDTA pH 8, 150mM NaCl, 0.1% SDS, in H_2O). The input chromatin samples are also equilibrated in 95µl of elution buffer and from this step on, together with the flow-through fractions, they are subjected to the same treatments as the ChIP samples. The crosslinks between DNA and protein are reversed by incubating the samples ON at 65°C in a thermomixer with 800rpm shacking. 20µg of PK are added to the tubes and incubated for 2 hours at 55°C in a thermomixer, shacking at 800rpm. Only applied to the ChIP samples: the beads are separated from the solution using a magnetic rack and the supernatants are collected into new tubes. The prepared DNA is purified by phenol chloroform. One volume of phenol-choloroform-isoamyl-alcohol (25:24:1) is added to each tube, mixed with a vortex at max speed for 30 seconds, and spun for 5 minutes at RT at 15000g. The aqueous phases are collected into new tubes. 2.5 volumes of cold 100% ethanol, 20µg of glycogen and 0.1 volumes of NaAcetate 3M pH 5 are added to the tubes. They are then vortexed for 30 seconds at max speed, incubated for 1h at -20°C and spun for 10 minutes at >13000rpm on a benchtop centrifuge at 4°C. The DNA pellets are

washed with cold 70% ethanol by spinning for 10 minutes at >13000rpm on a benchtop centrifuge, at 4°C. The DNA is resuspended in 40µl H_2O and can be stored at -20°C. 1µg of DNA from the flow through is run on a 1.5% agarose gel. It is expected that the majority of the DNA is found in the range 100–400bp. If the fragmentation considerably exceeds this range, the DNA cannot be used for downstream analysis. The quality of the ChIP is assessed by qPCR. The enrichment of the ChIP DNA relative to the input is compared for different genomic locations representing known enriched/non-enriched regions. If the signal to noise ratio is higher than 30, the samples can be subjected to library preparation and sequencing.

Epitope	Antibody	Tissues per ChIP-seq		Sonication volume (µl)	Antibody (µg)
		Slices	10^6 cells		
H3K27ac	ab4729	10	3	300	1
H3K4me1	ab8895	10	3	300	0.5
H2A.Z	ab4174	20	6	300	1
RNA Pol II	sc-899	80	24	900	12
H3K27me3	ab6002	10	3	300	0.5
H3K9me3	ab8898	20	6	300	0.7
CTCF	07-729	80	24	900	10
RAD21	Ab992	80	24	900	3

4.2.13.1 Library Preparation and Sequencing

The libraries for ChIP-seq are prepared using the NEB next kits (for libraries and multiplexing, listed in the materials section) according to the manufacturer's instructions, applying the following adaptations. DNA purification between every reaction is carried out using the Quiaquick PCR purification kit, instead of the (recommended) AMPure beads. The size selection is performed before the PCR amplification, like described in the protocol, but using 2% E-gels from Invitrogen instead of AMPure beads. The quality of the library is assessed by DNA electrophoresis on chip, Agilent Bioanalyzer, DNA 1000 kit. It is expected that all the samples have a unique fragments distribution centered on 320bp. In case some samples carry adaptor dimer contaminants, characterized by a typical peak around 80bp in the Bioanalyzer chromatogram, the samples are processed using SPRI select beads. The samples are all brought to exactly 50µl. 55µl of beads are used for the purification. This ratio of aqueous phase and beads efficiently removes DNA with a size equal or lower than 80bp. The affected samples are reanalyzed on the

Bioanalyzer after the purification. Once all the samples show a single distribution of fragments, they are diluted to 10nM and can be pooled (if required). The sequencing is done on Illumina HiSeq 2000 machines, choosing the option 50bp single-end reads with a dedicated short read for the index. The procedure requires that both input and ChIP samples are sequenced to at least 1X coverage. For mammalian genomes like mouse this corresponds to approximately 10^7 uniquely mapped reads with average genomic fragment size of 250bp. Usually up to 12 samples can be pooled in one HiSeq 2000 lane.

4.2.13.2 ChIP-seq Data Processing

The reads obtained from the sequencing are aligned to the NCBI37/mm9 reference mouse genome using the bowtie2 software (version 2.1.0), using the default settings (one mismatch allowed). Alignments are converted from sam to bam-format, They are then sorted and indexed using samtools software (version 0.1.19). Peaks are called using macs14, also in this case using the default parameters. Coverage tracks are done using the GenomicRanges, GenomicAlignments and rtracklayer R packages (Bioconductor version: release 3.1). Peaks lift over to all the mouse reference genomes coordinates is done using rtracklayer R package. Sequencing quality is assessed using ShortRead R package (same Bioconductor release as before).

4.2.13.3 ChIP-qPCR on *c-Myc* Inversion Lines

The occupancy of CTCF and RAD21 on *c-Myc locus* is analyzed on Wt, INV1, INV2 and INV3 lines on 13 selected genomic locations, plus 10 other control regions (as listed in the table below).

Primer name	Genomic location	Type	Figure nomenclature*
chr15-59928547-59929378	chr15:59928547	Assay	F1
chr15-60591260-60592101	chr15:60591260	Assay	F2
chr15-60933612-60934521	chr15:60933612	Assay	FM1
chr15-60988828-60989797	chr15:60988828	Assay	FM2
chr15-61814519-61815473	chr15:61814519	Assay	M1
chr15-62159709-62160443	chr15:62159709	Assay	M2
chr15-63171736-63172525	chr15:63171736	Assay	M3
chr15-63432906-63433773	chr15:63432906	Assay	TEL1

chr15-63483813-63484536	chr15:63483813	Assay	TEL2
chr15-63553022-63553963	chr15:63553022	Assay	TEL4
chr15-63572139-63572734	chr15:63572139	Assay	TEL5
chr15-63602886-63603631	chr15:63602886	Assay	TEL6
chr15-63660558-63660911	chr15:63660558	Assay	TEL8
chr8-47020195_Pos1	chr8:47020195	Control	Positive 1
chr1-183819016_Pos2	chr1:183819016	Control	Positive 2
chr12-82889780_Pos3	chr12:82889780	Control	Positive 3
chr1-84730497_Mid1	chr1:84730497	Control	Middle 1
chr3-36036726_Mid2	chr3:36036726	Control	Middle 2
chr5-65394338_Mid3	chr5:65394338	Control	Middle 3
chr10-100659186_Neg1	chr10:100659186	Control	Negative 1
chr5-59680664_Neg2	chr5:59680664	Control	Negative 2
chr14-81658234_Neg3	chr14:81658234	Control	Negative 3
chr16-53987613_Neg4	chr16:53987613	Control	Negative 4

(*) F = Fam84b-TAD; FM = Fam84b-c-Myc-TAD-boundary; M = c-Myc-TAD; TEL = Telomeric End of *c-Myc locus* (the numbers represent the CTCF peaks numeration on figure 23A).

The experiment is done on E11.5 face and liver. For each line at least three independent biological replicates are processed. The samples for INV1 are obtained from heterozygote-heterozygote breeding, from which only homozygote tissues are used. The samples for INV2 and INV3 are generated from homozygote-homozygote breeding, while the Wt from C57BL/6J-C57BL/6J breeding. In order to minimize litter-based biases, several litters for each line are collected and, tissue slices from each litter are randomly subdivided between the replicates (see table below for more details). Each sample is processed as described earlier, and the chromatin thereby obtained is split into two aliquots, used for CTCF and RAD21 ChIP. The ChIP and input DNA (84 samples in total) is subjected to qPCR on a Fluidigm system (96 primers, 96 samples chip). This setup guarantees at least 4 independent qPCR reactions for each sample, using 23 primers.

Line	Tissue	Antibodies	Biological replicates	Total litters	Slices per ChIP
Wt	Liver	CTCF, RAD21	3	7	8
Inversion 1	Liver	CTCF, RAD21	3	6	2.5
Inversion 2	Liver	CTCF, RAD21	3	9	10.5

Inversion 3	Liver	CTCF, RAD21	3	13	7.5
Wt	Face	CTCF, RAD21	4	7	6
Inversion 1	Face	CTCF, RAD21	3	6	2.5
Inversion 2	Face	CTCF, RAD21	4	9	7.5
Inversion 3	Face	CTCF, RAD21	4	13	5.5

4.2.13.4 ChIP-qPCR on *c-Myc* Inversion Lines Data Processing

Enrichment compared to input is calculated according to the equation 2, where *Ref* represents the input and *X* the ChIP.

To test for difference in enrichment on inversion lines and Wt (control genotype) the standard *t-test* has been used. However, the fact that the variance estimated for the technical and biological replicates does not represent the total one, may give some inaccuracies. In ChIP-qPCR the variance is modeled according to the equation 7.

Equation 7: $$\sigma^2 = \sigma^2_{biological} + \sigma^2_{detection} + \sigma^2_{residual}$$

The $\sigma^2_{detection}$ represents the variability of the measurement itself (qPCR variability on both input and ChIP). The $\sigma^2_{biological}$ can be further subdivided into two components, modeled by equation 8.

Equation 8: $$\sigma^2_{biological} = \sigma^2_{replicate} + \sigma^2_{genotype}$$

The biological replicates do not allow estimating the whole biological variance. There is another source of biological variability, represented by having different lines. One way to overcome this problem could be to perform heterozygotes-heterozygotes crosses for each line, and compare Wt and homozygotes tissues obtained from the same breeding.

If this breeding setup is not available, like in this case, an analytical correction to the σ^2 can be applied. The aim is to find the $\sigma^2_{genotype}$ term, whose addition to the equation 8 increases the variance, resulting in a slight decrease in statistical power.

The estimate of $\sigma^2_{genotype}$ must be done on the control regions, which are assumed to be unaffected by the genetic engineering. The first step in the procedure is to apply a glog transformation to the data, this in order to rule out mean-variance relationships (equation 9). Afterwards, the variance components are obtained by

fitting linear mixed-effect models (lme4 R package), specifying genotype and replicate random-effects. The total variance σ^2 is calculated by summing all the variance components in the model.

Equation 9: $$g\log(x) = \left[ArcSin(x) - \log(2)\right] / \log(2)$$

Finally, *p*-values are calculated and corrected by Benjamini-Hockberg method.

5 Results

5.1 Genetic Characterization of *c-Myc locus*, Done By Veli Uslu

5.1.1 Regulatory Landscape of *c-Myc locus* in Mouse Embryos

The regulatory potential of *c-Myc locus* was assessed by a former student in the laboratory, Veli Uslu, using the Genome Regulatory Organization Mapping with Integrated Transposons (GROMIT) system (Ruf et al., 2011; Uslu et al., 2014). GROMIT is based on a LacZ reporter gene under the control of the minimal promoter of the *β-globin* gene. This promoter does not allow the expression of the LacZ reporter unless regulatory inputs are conveyed. Therefore, GROMIT can be seen as a regulatory sensor that reveals the transcriptional capability of the genome at the place where it is inserted, providing a readout of the integrated effects of regulatory elements, DNA topology and chromatin environment altogether.

Veli Uslu started this screening by remobilizing the GROMIT insertion 179039 (located inside *c-Myc locus*, established by Sandra Ruf) and obtained more than 50 new insertions in *c-Myc locus*, covering mostly the telomeric side.

From the analysis of the LacZ expression, performed at 11.5 days of gestation (E11.5) it was possible to notice that many insertions on the telomeric side of the *locus* showed expression of the LacZ reporter—at least in some tissues—overlapping with the endogenous *c-Myc* expression seen by RNA *in-situ* hybridization (figure 7). Remarkably, some of the insertions that gave LacZ staining patterns reproducing most closely the endogenous *c-Myc* expression, occurred 1.7Mbp away from *c-Myc* promoter, at the telomeric end of the *locus* (179039 (TRACER database entry ID) or 17a (Veli Uslu's nomenclature, expressing the distance from *c-Myc* promoter in hundreds of Kbp), figure 7), the tissues nomenclature is summarized in the supplementary figure 1). Tissues such as facial mesenchyme (FM), somites, branchial arches (BA), rhombic lip (RL) and proximal limb mesoderm (PLM) appeared stained in multiple insertions. Liver staining was observed only for two insertions (one of which is 194578 (3a), figure 7), while tissues like tail bud, forebrain, midbrain, neural tube and terminal limb mesoderm were never stained.

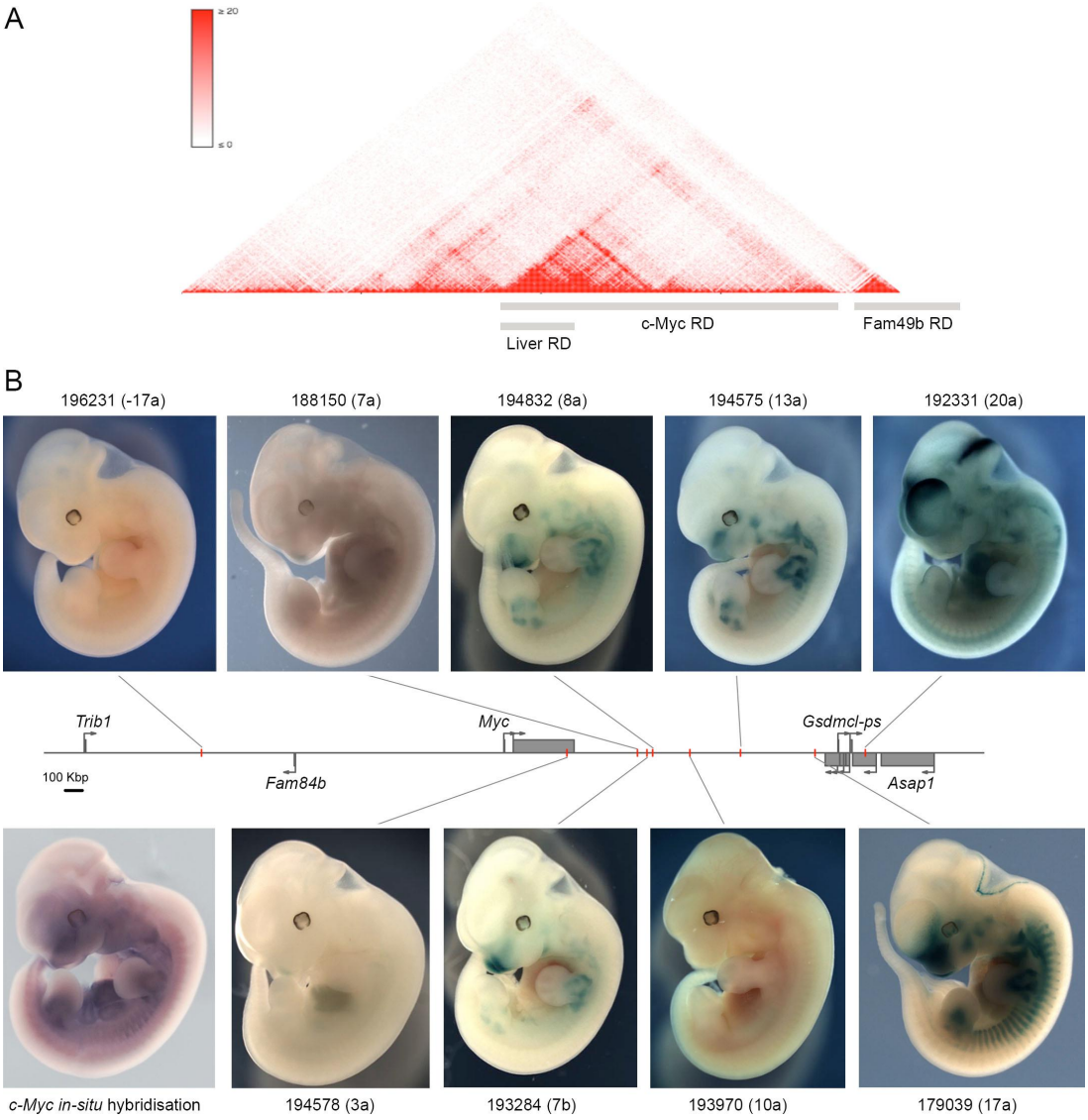


Figure 7 | *c-Myc locus* regulatory landscape. A) Hi-C interaction heatmap of mouse CH12 cells (mm9 genomic coordinates chr15:60000000-64000000) available at: http://promoter.bx.psu.edu/hi-c/view.php, with annotation of the *c-Myc locus* regulatory domains (RDs) (Rao SS et al., 2014). The resolution is 25Kbp, while the intensity reflects the number of sequencing reads for each interaction: ≤1, white; ≥80, red. B) Material presented thanks to kind concession of Veli Uslu. This panel shows the LacZ reporter staining of E11.5 mouse embryos from selected GROMIT insertions (located on the paternal allele) representative of *c-Myc* and *Fam49b* RDs. The name of each insertion refers to the entry ID in the public database TRACER (http://www.ebi.ac.uk/panda-srv/tracer/). The name in brackets refers to the approximate distance from *c-Myc* promoter, expressed as multiples of $10^{\wedge 5}$bp.

Three insertions occurred further away, after the telomeric end of the gene-desert, in the *Fam49b locus*. These insertions showed a completely different scenario than in *c-Myc locus*, they had a LacZ staining scattered throughout the embryo, being particularly marked on the tail bud, forebrain, midbrain-hindbrain boundary and neural tube (192331 (20a), figure 7). Finally, three insertions were obtained at the centromeric side of *c-Myc locus* and were in all the three cases negative.

This screening revealed the presence of three distinct regulatory domains (RD) on the telomeric side of the *locus* and on the neighboring *loci* (Symmons et al., 2014). The first one extends from *c-Myc* to the end of *Pvt1* transcribed region, showing mainly liver specific transcriptional inputs. The second RD includes the first one and extends from *c-Myc* to the *Gsdmc* gene cluster. It contains different insertions that show LacZ expression in all the tissues with higher *c-Myc* expression (often in just a combination of them), and therefore termed c-Myc-RD. Finally, the third RD lies telomeric to the *Gsdmc* gene cluster, on the *Fam49b locus* (referred to as Fam49b-RD), showing completely distinct staining patterns than the ones on c-Myc-RD.

Importantly, in this region the disposition of RDs extensively overlap with the disposition of the topological domains. Hi-C interaction maps of *c-Myc locus* available for mESC, mouse cortex and CH12 cell line show that *c-Myc* is embedded in one TAD extending from just after the gene *Fam84b* to the *Gsdmc* cluster, termed c-Myc-TAD (figure 8). c-Myc-TAD can be further subdivided into two smaller sub-TADs, having *c-Myc* itself at the border. The c-Myc-RD lies inside the c-Myc-TAD, overlapping with the telomeric c-Myc-subTAD. At the *Gsdmc* gene-cluster a TAD-boundary separates the c-Myc-TAD from the adjacent and more telomeric one. This latter TAD contains the *Fam49b* gene (Fam49b-TAD) and overlaps with the Fam49b-RD. The correlation between TADs and the reporter staining of GROMIT insertions suggests that the activity of regulatory elements may be indeed confined within the TAD and supports the model proposed by Symmons and colleagues describing an extensive overlap between TADs and regulatory domain (Symmons et al., 2014).

A more detailed look reveals that inside c-Myc-TAD not all the insertions led to reporter expression in the same tissues. In many cases the staining was present only in a subset of the tissues where the endogenous *c-Myc* expression is higher. In addition, some insertions inside c-Myc-TAD gave no reporter expression at all (referred to as "cold spots" or "blank spots", 193970 (10a), figure 7). Focusing on the nasal epithelia (NE) the LacZ staining was seen only in insertions located towards the telomeric edge of the *locus.* Others, for example the FM, somites and PLM showed LacZ staining of different intensities/strength when comparing GROMIT

insertions integrated in different parts of the c-Myc-TAD. Finally in the case of the liver, the reporter staining was captured only within ~ 300Kbp to *c-Myc,* on the telomeric side of the *locus*.

In conclusion, the fact that the regulatory landscape varies from tissue to tissue indicates that the *locus* could have tissue-specific properties. In addition, the fact that not all the insertions occurring intra-TAD gave the same reporter staining, shows that not all the positions inside the TAD are equally likely to receive the same regulatory inputs. One possible explanation for this may be that in *c-Myc locus* different intra-TAD sub-topologies or chromatin features vary from tissue to tissue.

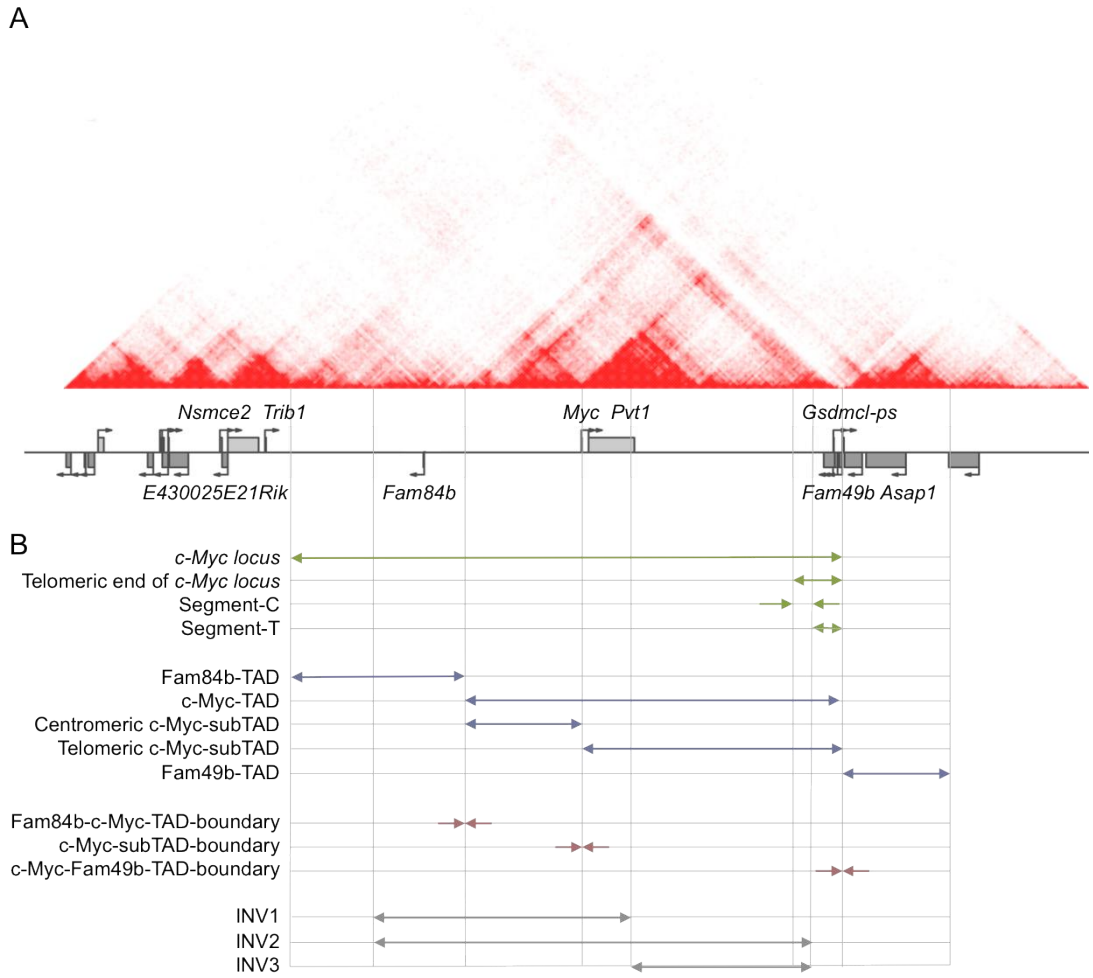


Figure 8 | *c-Myc* and neighboring *loci* topological organization. A) Hi-C interaction heatmap of mouse CH12 cells (mm9 genomic coordinates: chr15:58000000-65500000), available at: http://promoter.bx.psu.edu/hi-c/view.php (Rao SS et al., 2014). Each point shows the interaction between two 25Kbp bins while the intensity refers to the number of reads for each interaction, from ≤1 (white) to ≥80 (red). B) The extension of *c-Myc locus* features is represented by green arrows, the TAD structures by blue arrows, TAD-boundaries by red arrows and the balanced chromosomal inversions (INV1, INV2 and INV3) by grey arrows.

5.1.2 Identification of Important Regions for *c-Myc* Expression in Facial Tissues

In human, GWAS showed that the 8q24.21 is the *locus* with the highest attributable risk for non-syndromic cleft-lip and palate (Birnbaum et al., 2009). In mouse, the ortholog of the human 8q24.21 lies in the chromosome 15 and corresponds to a 575Kbp portion of the telomeric side of *c-Myc locus*. In this region (as well as in the neighboring ones), GROMIT captured transcriptional inputs acting on the MFM and supports the idea that the region contains long-range regulatory elements important for *c-Myc* expression in the developing face.

This tissue in mouse and human is the one forming the anterior structures of the oral cavity, *i.e.* the upper lip and the primary palate. The FM can be further subdivided in two components: the medial facial mesenchyme (MFM) and the lateral facial mesenchyme (LFM), separated by the nasal pits. The two MFN processes fuse with the maxillary process and form the upper lip and primary palate, whereas the LFM give rise to the nasal alae. In mouse, this event occurs between 10 and 15 days of gestation. Growth defects of the MFM at this stage, lead to incomplete merging with the maxillary process and can cause the disease cleft lip and/or cleft palate (CLP) (reviewed in M. J. Dixon, Marazita, Beaty, & Murray, 2011; Mossey, Little, Munger, Dixon, & Shaw, 2009).

Veli Uslu hypothesized that the reduced tissue proliferation causing the incomplete merge of the MFM—observed in the CLP disease—was due to a decrement of *c-Myc* expression, caused in turn by polymorphisms on remote enhancers.

Veli Uslu tested this hypothesis in mouse embryos, and also, restricted the position of the putative regulatory element(s) to smaller windows. He created a panel of chromosomal deletions spanning different parts of the telomeric side of *c-Myc locus*. He did this using the Targeted Meiotic Recombination (TAMERE) method (based on the Cre-LoxP system) taking advantage of the LoxP site present inside each GROMIT insertion (Hérault, Rassoulzadegan, Cuzin, & Duboule, 1998). The analysis of the LacZ reporter expression from E11.5 mouse embryos carrying the deletions showed that MFM and NE reporter activity relied on two independent but close regions. The NE critical region was restricted to about 100Kbp, from ~ 1.4Mbp to ~ 1.5Mbp telomeric to *c-Myc* promoter (Del(14–15), figure 9D), while the one for the MFM was restricted to about 600Kbp, ranging from ~ 800Kbp, to ~ 1.4Mbp telomeric to *c-Myc* promoter (Del(8–14), figure 9D, supplementary figure 3, mm9 genomic coordinates: chr15:62561825-63135562) (Uslu et al., 2014).

Importantly, in the same study it was reported that the deletion of both critical regions, that caused loss of reporter expression in all facial tissues (MFM, LFM and NE, Del(8–17), figure 9D), did not only affect the reporter expression. It also affected the endogenous *c-Myc* expression, specifically in the facial tissues and to a lower extent in the fore limbs (FL), but not in other tissues such as heart and liver. Therefore, the deleted region contained regulatory elements with enhancer feature, affecting selectively *c-Myc* in the face and FL. Their absence sporadically led to cleft lip and palate, and also significantly affected the proportions of craniofacial bones in adult mice, suggesting that these regulatory elements are important for the correct facial morphogenesis (Uslu et al., 2014).

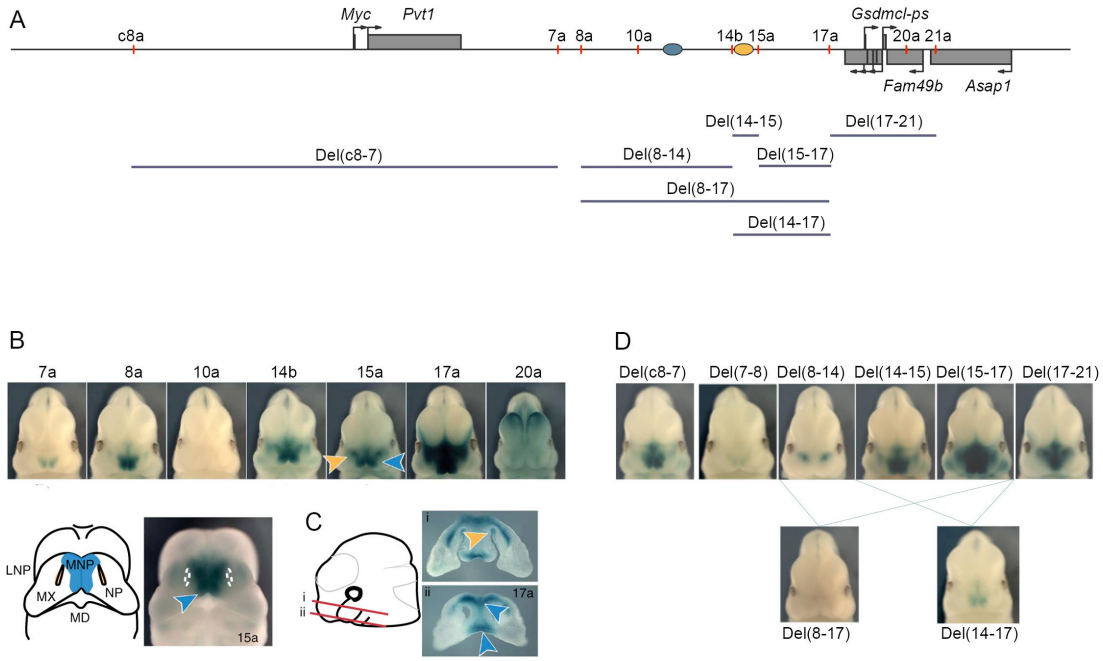


Figure 9 | Identification of facial enhancers in *c-Myc locus*. Adapted from (Uslu et al., 2014). A) Panel representing the disposition of deletions in *c-Myc* and neighboring *loci,* the position of GROMIT insertions, and, the approximate position of the facial regulatory elements identified (blue and yellow ellipses, for facial mesenchyme and nasal epithelia respectively). B) Frontal view of the head of E11.5 mouse embryos from different GROMIT lines subjected to LacZ staining procedure. The staining on medial face mesenchyme (MFM) and nasal epithelia (NE) is indicated by blue and yellow arrows respectively. MNP, medionasal process; MX, maxillary process; MD mandibular process; LNP, lateral nasal process; NP nasal epithelium. C) Vibrotome sections of the face of an E11.5 embryo positive for the 17a GROMIT line. The LacZ staining in the MFM and NE is shown by a blue and a yellow arrow respectively. D) Frontal view of the head from E11.5 mouse embryos carrying selected chromosomal deletions, and subjected to the LacZ staining protocol. The deletions Del(8–14) and Del(14–15) allowed the identification of important regions responsible for the LacZ reporter expression in the MFM and NE respectively.

5.2 Chromatin Composition of *c-Myc locus* in Wt Embryonic Tissues

5.2.1 Enhancer Associated Chromatin

5.2.1.1 Setup

The collocation of the MFM enhancer(s) in a region of 600Kbp needed to be refined. For this reason, in order to restrict their position even further we performed ChIP-seq for the histone modifications enriched at enhancer sites, namely H3K27ac and H3K4me1. These chromatin marks do not provide the exact same readout. H3K4me1 is found at active and inactive/poised enhancers, while H3K27ac comprises mostly active ones (Creyghton et al., 2010b; Xu et al., 2009).

We did these experiments on E11.5 Wt specimens of face, liver, and for H3K27ac, also the fore brain (FB).

The way we dissected the face—common to all the experimental procedures described thereafter, *e.g.* cDNA qPCR and bisulfite-sequencing—included both the FM and the NE tissues (the latter is embedded inside the FM). However, the fact that the FM is much more represented than the NE (in terms of number of cells, data not shown), make us support the idea that the experimental evidences concerning the face specimens are more attributable to the FM than to the NE. However, in order to avoid inaccuracies we will refer to this sample as face.

Regarding the liver, the reason why we included it in this analysis, is that this tissue represents an additional system (to the face) to study *c-Myc* promoter–regulatory elements communication. Despite the fact that no chromosomal deletion was done in order to prove the presence of liver enhancers in the *locus*, the liver is one of the tissues where *c-Myc* expression is higher at this stage, as seen by *in situ* hybridization (figure 7). Two GROMIT insertions, mapped by Veli Uslu, located in the non-coding transcript *Pvt1*, revealed LacZ reporter expression in this tissue. Although GROMIT *per se* does not indicate the position of enhancers, it clearly indicates that in the liver there are regulatory inputs reaching the 194578 (3a) insertion (inside *Pvt1*) as well as *c-Myc* promoter (as seen by *in situ* hybridization) (figure 7). The fact that the deletion of 900Kbp (Del(8–17), figure 9D), which includes the facial enhancers (for MFM and NE), is responsible for *c-Myc* downregulation in the face (and to a lower extent in the FL), but not in the liver, indicates that different sets of regulatory elements coordinate *c-Myc* expression in these two tissues. As the liver enhancers are most likely located centromeric to the

deletion breakpoint 8a (Del(8–17)), and, the LacZ expression in the liver is seen only for insertions integrated inside *Pvt1* transcribed region, it is possible that these enhancers lie inside this region.

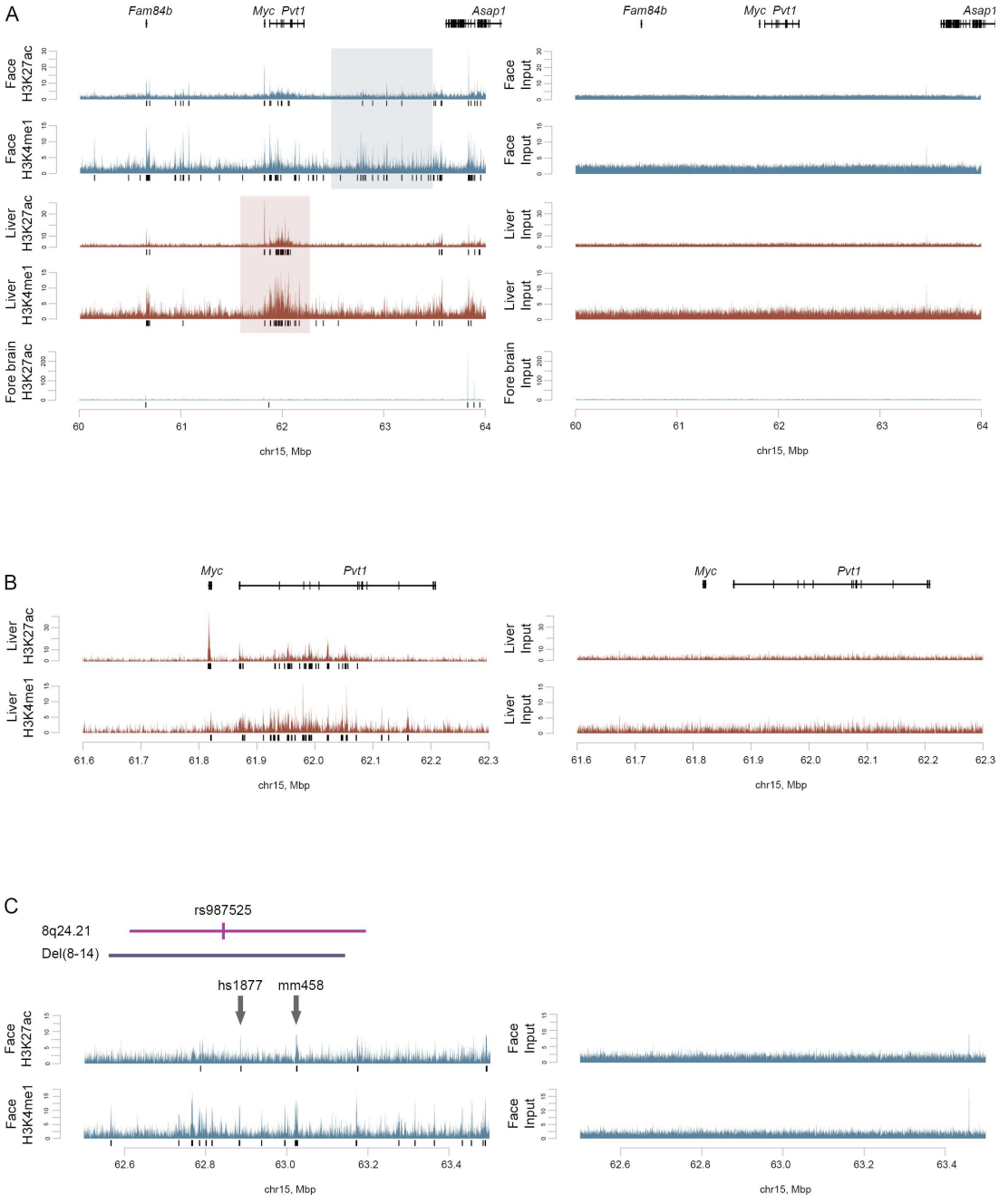
Finally, concerning the FB, we know that in this tissue *c-Myc* expression is kept to the basal level. In addition, no GROMIT insertion (integrated in *c-Myc locus*) captured any regulatory input in the FB. For these reasons, we included the FB in this analysis as a negative control.

5.2.1.2 Results

Focusing on the telomeric side of *c-Myc locus*, in the face we identified several regions with significant enrichment of H3K4me1 compared to the input chromatin (also called "peaks"), and are scattered throughout the *locus*. H3K27ac instead shows fewer enriched regions, clustered in three main parts (figure 10A). One encompasses the whole *Pvt1* transcribed region (about 210Kbp large), another is found at the telomeric end of the *locus*, extending for about 100Kbp around the GROMIT insertion 179039 (17a), and the last one corresponds to the ~ 600Kbp region, delineated by the deletion Del(8–14), containing the MFM enhancer(s). In particular, three out of four H3K27ac peaks of this cluster fall inside this critical region (figure 10C). Interestingly, one of these peaks overlaps with the mouse ortholog of the sequence hs1877, shown by transgenic assays to drive reporter expression in the MFM, as well as in few other tissues (figure 10D and F) (Attanasio et al., 2013). Another peak, also located inside the MFM enhancer(s) critical region, lies close (~ 2.5Kbp) to the sequence mm458 (figure 10E). This sequence was tested by transgenic assay (in the same study as hs1877) and was shown to drive reporter expression in the maxillary process (and a few other tissues, but not the FM or NE) (figure 10F). Although mm458 and the close H3K27ac peak do not overlap, they both map within a broader H3K27ac enriched region, sized about 7Kbp (figure 10E). However, with our peak calling settings, we call significantly enriched just one part of this ~ 7Kbp region, namely, the more centromeric. Importantly, the maxillary process, stained in mm458 transgenic assays, is not included in the portion of facial tissue that we used for the ChIP-seq experiments (unless as a minor contaminant). Therefore it is possible that the H3K27ac peak does not delineate the regulatory elements found on the mm458 sequence.

Regarding the other two clusters of H3K27ac peaks in the face (the one on *Pvt1* transcribed region and the one located at the telomeric end of the *locus*), their

function is unknown. However, chromosomal deletions of both of them show that they are dispensable for driving LacZ transcription in the facial tissues (FM and NE), as reported by Veli Uslu (figure 9D, Del(c8–7), Del(15–17) and Del(17–21)) (Uslu et al., 2014). Therefore, it is possible that MFM enhancer(s) correspond to one or more of the three peaks included in the MFM enhancer(s) critical region found by Veli Uslu.



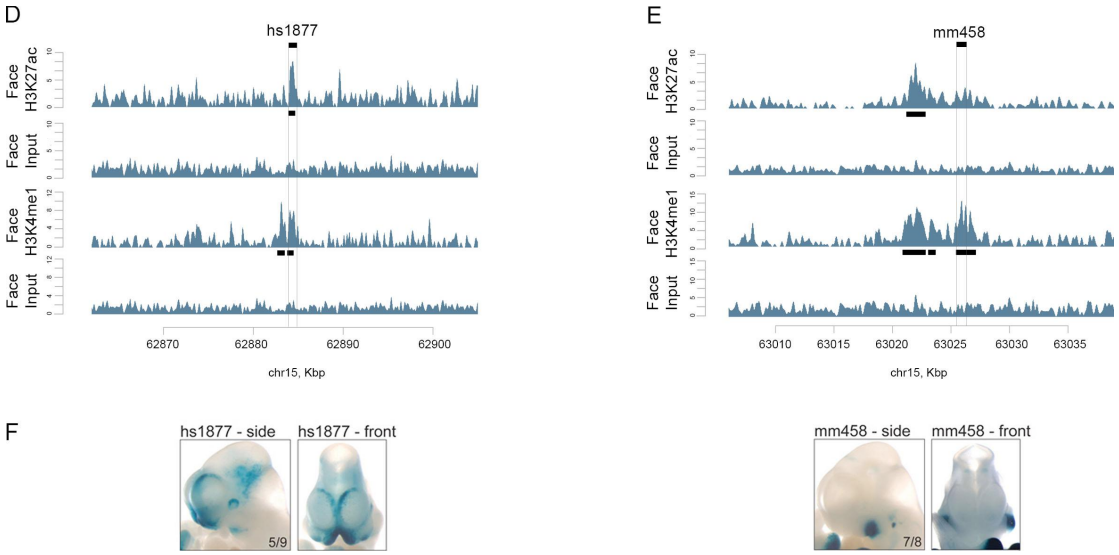


Figure 10 | Enhancer associated chromatin. H3K27ac and H3K4me1 ChIP-seq analysis of E11.5 Wt (C57BL/6J) face (blue), liver (red) and fore brain (FB, green). The y-axis represents the sequencing coverage (different values are due to the slightly different sequencing depth of the individual libraries, as well as differences in signal/noise ratios of the different antibodies and tissues used). The tracks on the left show the actual ChIP-seq, while the ones on the right show the input chromatin control (used to show the biases introduced by the procedure and not due to the antibody pull-down). Peaks (called using MACS14 software) are shown on a dedicated track underneath each ChIP-seq profile. A) Extended view of the *c-Myc locus* (mm9 coordinates: chr15:60000000-64000000). The emphasized regions on the face and liver tracks represent the zoomed regions shown in panel B and C respectively. B) Detailed view of liver ChIP-seq coverage profiles on *Pvt1* transcribed region (region potentially containing liver enhancer(s)). C) Detailed view of face ChIP-seq experiments, representing part of the telomeric side of *c-Myc locus* (mm9 coordinates: chr15:62500000-63500000) which includes the ~ 600Kbp region (Del(8–14)) found by Veli Uslu to be critical for *c-Myc* expression in the MFM (mm9 coordinates: chr15:62561825-63135562). The locations of the mouse orthologous of the human 8q24.21 region (cleft lip and/or palate susceptibility region, including the SNP rs987525) and Del(8–14) are indicated by magenta and blue lines respectively. D) Detailed view of the region overlapping with the mouse ortholog of the (human) sequence hs1877. Attanasio and colleagues showed by transgenic assays that this region is capable of driving reporter expression in the MFM and other tissues, in mouse embryos. Input chromatin controls and peak calling tracks are shown below each ChIP-seq coverage track (same as for panel E). E) View of the region surrounding the sequence mm458, found by Attanasio and colleagues to drive reporter expression in the maxillary process and a few other tissues in mouse embryos. F) Adapted from (Attanasio et al., 2013). Panel showing LacZ staining of E11.5 mouse embryos tested by transgenic enhancer reporter assays (sequences hs1877 an mm458). The numbers indicate the fraction of stained embryos out of the total ones, tested for each sequence.

In the liver, for both H3K4me1 and H3K27ac most of the peaks lie inside the *Pvt1* transcribed region (figure 10B). For both modifications there are few peaks occurring around the insertion 179039 (17a). These data support the idea mentioned above, that the liver enhancer(s) are located inside *Pvt1* transcribed region.

Nevertheless, due to the lack of chromosomal deletions showing any effect on the endogenous *c-Myc* expression, or the abolishment of reporter transcription in the liver, we cannot prove the presence of liver enhancer(s) in this region (there is a deletion that completely removes *c-Myc* and *Pvt1* (figure 9D, Del(c8–7), however, none of the breakpoints show staining in the liver, and therefore, it cannot show any reduction of LacZ expression in this tissue).

In the FB, according to the observations that this tissue has basal *c-Myc* expression, and no regulatory inputs are captured by GROMIT in any location of the *locus* (figure 7B), inside *c-Myc locus,* except for *Fam48b* and *c-Myc* promoters, we find no H3K27ac enriched regions (figure 10A and supplementary figure 4).

The presence of H3K27ac at certain parts of the *locus*, support the observations drawn by Veli Uslu regarding the position of the facial and liver enhancers. In the face, our ChIP-seq datasets show the presence of three regions, sized about 1Kbp that may include the MFM enhancer(s). In the liver, the 21 H3K27ac peaks found in the *Pvt1* transcribed region make it more difficult to de-convolute the subset of regulatory elements (if any in this region) important for *c-Myc* expression in this tissue.

5.2.2 Transcriptionally Active Chromatin

It was recently shown that transcription by RNA Pol II often occurs at enhancer sites other than at protein coding genes (Andersson et al., 2014; Core et al., 2014). Despite the fact that it is not known how universal this mechanism may be, and, what fraction of enhancers undergo active transcription, we decided to take advantage of this feature and analyze the transcriptionally active chromatin inside *c-Myc locus*, with the purpose of refining even further the position of the tissue-specific enhancers.

We considered a few points before performing this analysis. Firstly, there are histone modifications that can be pooled-down (on ChIP-seq experiments) in order to identify actively transcribed genes, or active promoters. However, except for the H3K27ac (already discussed), these marks do not usually cover enhancer sites. Secondly, since the RNA molecules transcribed at enhancer sites are reported to be unstable, being degraded soon after they are produced (probably degraded co-transcriptionally, by the exosome complex), standard transcriptomics methods do not usually allow their identification (Core et al., 2014).

We opted for a RNA Pol II ChIP-seq approach. We performed this experiment on E11.5 face and liver, using an antibody that recognizes the RNA Pol II regardless of any post-translational modification (PTM) (for example the phosphorylation of the serine 2 of the RNA Pol II CTD, present only during elongation).

Inside *c-Myc locus* we observed RNA Pol II occupancy uniquely at annotated transcripts in both face and liver (figure 11A, B and C).

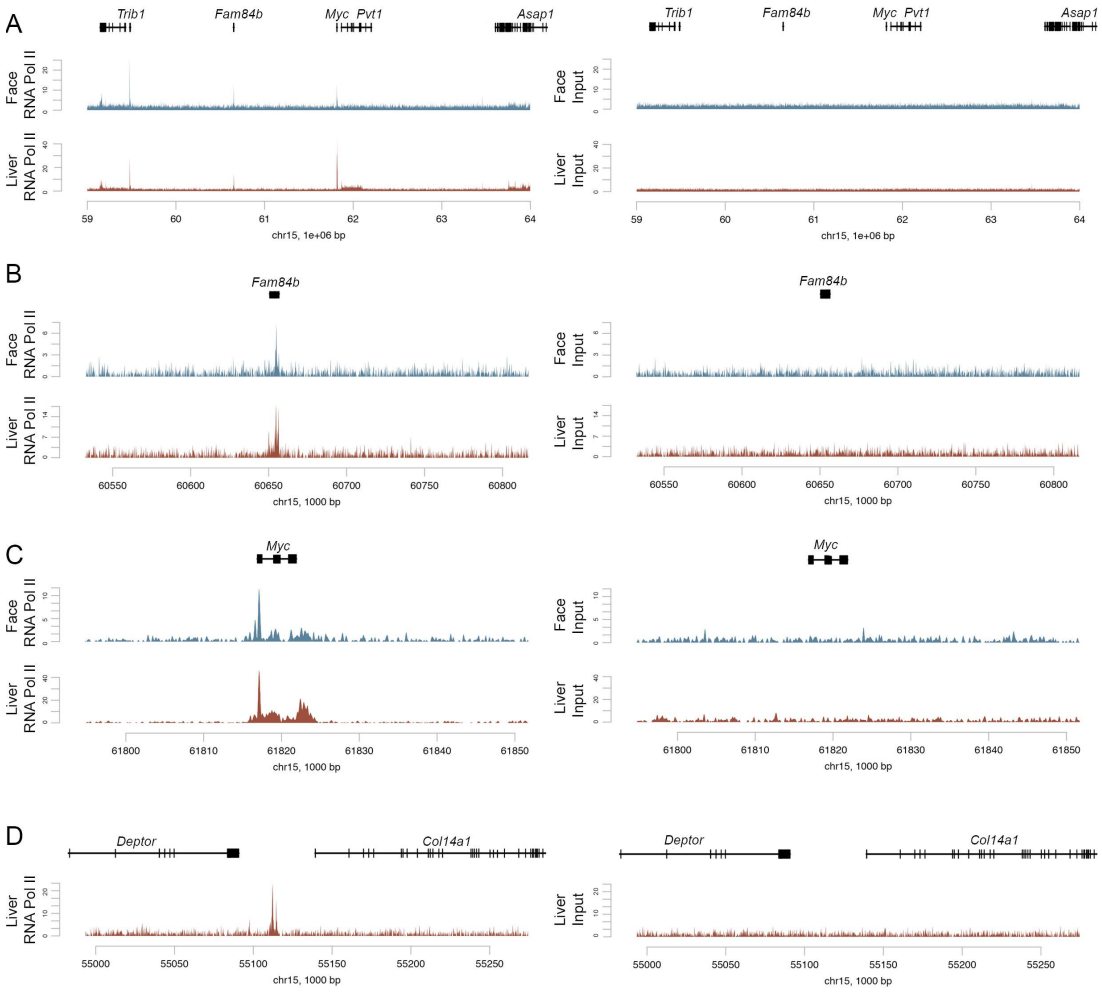


Figure 11 | Transcriptionally active chromatin. RNA Pol II ChIP-seq analysis of E11.5 Wt (C57BL/6J) face (blue) and liver (red). The y-axis represents the sequencing coverage (different values are due to the slightly different sequencing depth of the individual libraries as well as differences in signal/noise ratios characteristics of the different tissues). The tracks on the left show the actual ChIP-seq, the ones on the right, the input chromatin control (used to show the biases introduced by the procedure and not due to the antibody pull-down). A) View of *c-Myc* and surrounding *loci* (mm9 coordinates: chr15:59000000-64000000). B) Detailed view of *Fam84b* (mm9 coordinates: chr15:60530000-60820000). C) Detailed view of *c-Myc* (mm9 coordinates: chr15:61800000-61855000). D) View of an intergenic region showing RNA Pol II enrichment (mm9 coordinates: chr15:55000000-55250000).

In the liver, we saw some enrichment of RNA Pol II on *Pvt1* transcribed region. Since the liver enhancer(s) are likely located inside this region, we cannot tell whether the enrichment we found is the result of the *Pvt1* transcription or enhancer site transcription.

On a genome-wide scale, our data shows some intergenic RNA Pol II occupancy (figure 11D). However, the literature does not provide extensive description regarding the function of non-coding transcription in general. Some reports propose that non-coding transcription may account for a significant portion of gene expression regulation, with a wide spectrum of influence at several layers, including chromatin architecture/epigenetics, RNA splicing, editing and turnover (reviewed by Mattick & Makunin, 2006). Regarding its role on chromatin, it was shown that intergenic transcription can alter nucleosome density and play a role in protein–DNA interactions (Hainer, Pruneski, Mitchell, Monteverde, & Martens, 2011). Since non-coding transcription may have different potential functions, we cannot tell whether the intergenic RNA Pol II occupancy we saw on a few genomic locations underlies enhancer RNAs or other types of non-coding transcription.

In addition, because of the fact that we know from deletion experiments that the telomeric side of the *locus* contains facial enhancers (important for *c-Myc* expression on the FM and NE), but we did not see any intergenic RNA Pol II occupancy, we conclude from this analysis that either the experimental setup we used for this experiment is not suited to identify enhancer RNAs transcription, or that this phenomenon is not a universal feature of the enhancers.

5.2.3 Repressive Chromatin: Heterochromatin and Polycomb Chromatin

The identification of the facial and liver enhancers regions by GROMIT, chromosomal deletions and ChIP-seq allowed us to question why the effect of the enhancer was not reaching any position in the *locus*, but only some parts (as seen by GROMIT remobilization, figure 7). There were a few possibilities that could be tested. The first is that some regions are embedded in repressive chromatin; the second is that the topology of the *locus* allows only certain segments to physically contact the enhancers. We started tackling the question from a chromatin perspective, and for this reason we performed ChIP-seq for two well known repressive histone modifications: H3K9me3, marker of heterochromatin, and

H3K27me3, marker of Polycomb repression. Moreover there is evidence that genes marked by H3K27me3 interact with each other inside nuclear territories with active transcription. The contacts occur both intra and inter-chromosome and form repressive microenvironments inside active territories (Vieux-Rochas, Fabre, Leleu, Duboule, & Noordermeer, 2015). According to this view, Polycomb may also play a role in the 3D chromosomal organization of the genome, which makes the histone modification H3K27me3 particularly relevant for this project.

We performed H3K9me3 and H3K27me3 ChIP-seq experiments using facial and liver tissues from Wt embryos, at the stage E11.5.

In *c-Myc locus,* we observed in both face and liver one H3K9me3 enriched region, occurring about 90Kbp upstream of the GROMIT insertion 179039 (17a) (figure 12). In the liver, we saw an additional peak, located at the beginning of the *Gsdmc* gene cluster. Remarkably, the insertion 179039 (17a), as well as the ones integrated up to 300Kbp more centromeric, had reporter expression at least in the FM.

Regarding H3K27me3, also in this case the data looked very similar for both of the tissues analyzed. In the whole *locus* we saw a single peak, located on the promoter of *Pvt1* (figure 12). This result is consistent with the fact that this gene is expressed to an almost undetectable level in both embryonic and adult tissues.

From these data, we conclude that the histone modifications we analyzed do not play a major role in repressing specific parts of the *locus,* and cannot explain the regulatory landscape seen by GROMIT, described by Veli Uslu.

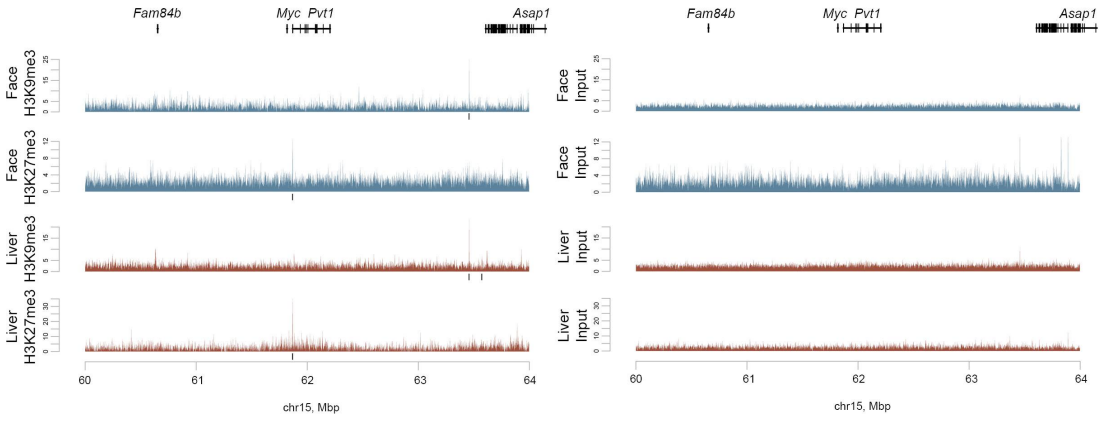


Figure 12 | Repressive chromatin. H3K9me3 and H3K27me3 histone modifications ChIP-seq on E11.5 Wt (C57BL/6J) embryos; face (blue), liver (red). View showing *c-Myc* and neighboring *loci* (mm9 coordinates: chr15:60000000-64000000). Lefthand tracks represent ChIP-seq experiments, while the righthand ones the input chromatin control. Peak-calling results are shown underneath each ChIP-seq track.

Regarding the proposed role of Polycomb to mediate long-range interactions, our data suggest that this is not likely the case of *c-Myc locus*. Inside the *locus* the only H3K27me3 enriched region is separated from the closest ones, on the *Fam49b-Asap1 locus*, by a known boundary region (probably overlapping with the c-Myc–Fam49b–TAD-boundary and the *Gsdmcl-ps* gene). GROMIT insertions integrated telomeric to this boundary element, show a very distinct LacZ staining pattern than the one observed in *c-Myc locus*. For these reasons, we rule out the possibility that remote H3K27me3 enriched regions are involved in the topological organization of *c-Myc locus*.

Despite the scarce presence of heterochromatin and Polycomb mark in *c-Myc locus*, compared to the rest of the genome, it is possible that another kind of repressive chromatin helps shaping the observed RDs. It is reported that the histone H1 positive chromatin covers almost half of the *Drosophila* genome and is the most represented chromatin type in this model, particularly common at gene poor regions such as *c-Myc locus* (Filion et al., 2010). Whether this kind of repressive chromatin covers some parts of the *locus*, remains to be proven, however, we cannot completely exclude the role of repressive chromatin to shape the regulatory landscape of the *locus*.

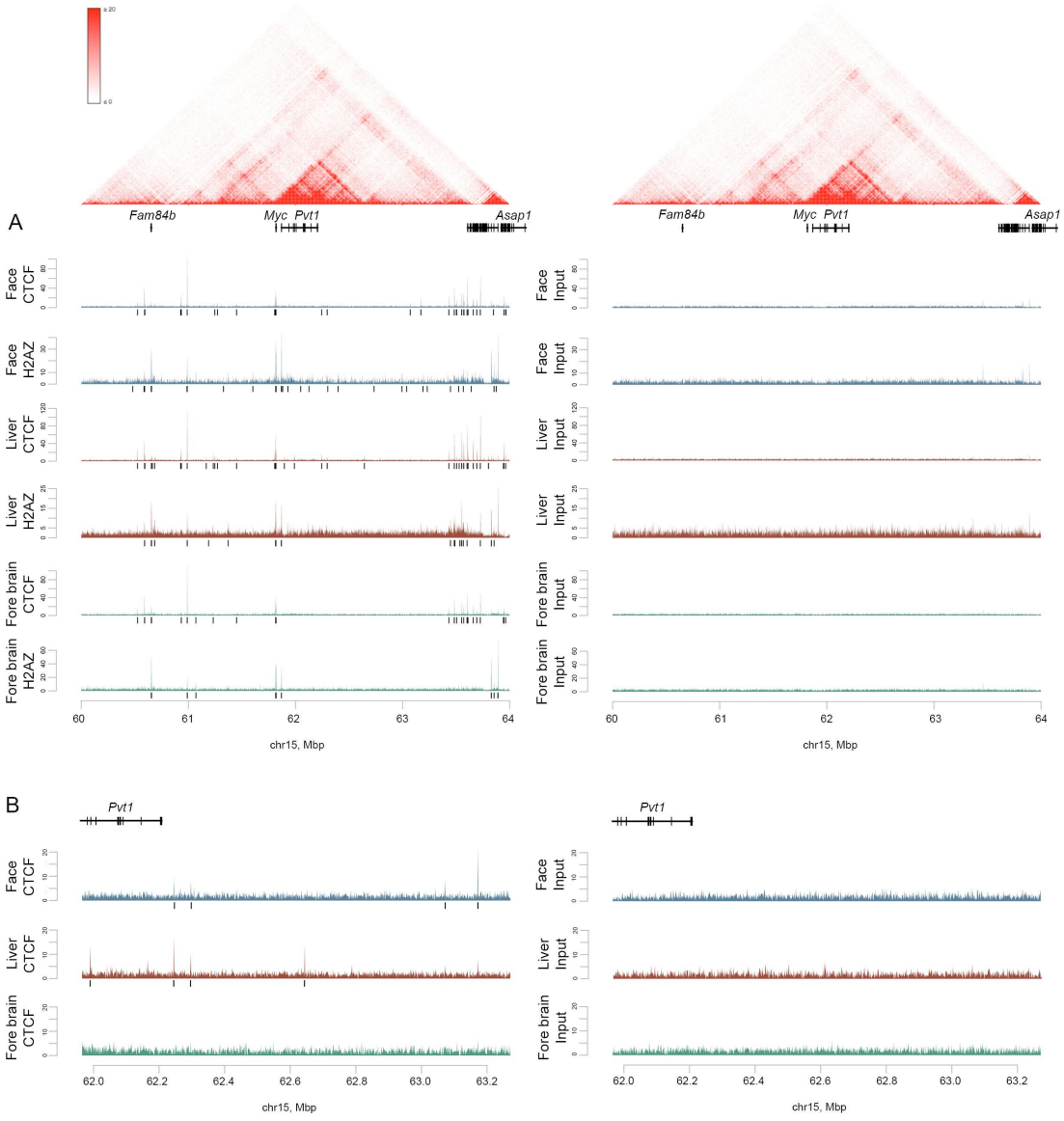
5.3 Binding of Architectural Proteins in *c-Myc locus* in Wt Embryonic Tissues

5.3.1 CTCF

The role of CTCF in mediating long-range interactions has been demonstrated in a few cases, the *H19-Igf2* and *Hbb loci* (figure 2) (Kurukuti et al., 2006; Splinter et al., 2006). Other than these few *locus* based cases, its general role in maintaining chromosomal architecture has been shown by genome-wide studies (Seitan et al., 2013; Sofueva et al., 2013; Solovei et al., 2013). Therefore, we hypothesized that CTCF may be important for promoter–enhancer long-range interactions in *c-Myc locus*.

We performed CTCF ChIP-seq on E11.5 face, liver, and FB as a control tissue. In these three tissues, we observed few qualitative changes in CTCF binding.

Starting from the common features, we saw an enrichment of CTCF at the four TAD-boundaries of the *locus* (figure 13), extending from the one overlapping with *Trib1*, to the one overlapping with *Gsdmcl-ps* (figure 8). We saw at the Trib1–TAD-boundary four distinct CTCF peaks; at the Fam84b–c-Myc–TAD-boundary, two; and at c-Myc–subTAD-boundary, three. Regarding the last TAD-boundary, overlapping with the *Gsdmcl-ps* gene—despite that we saw no CTCF enrichment—we saw a cluster of 10 binding sites at its flanking region, namely, the telomeric end of *c-Myc locus* (figure 13 A and C).



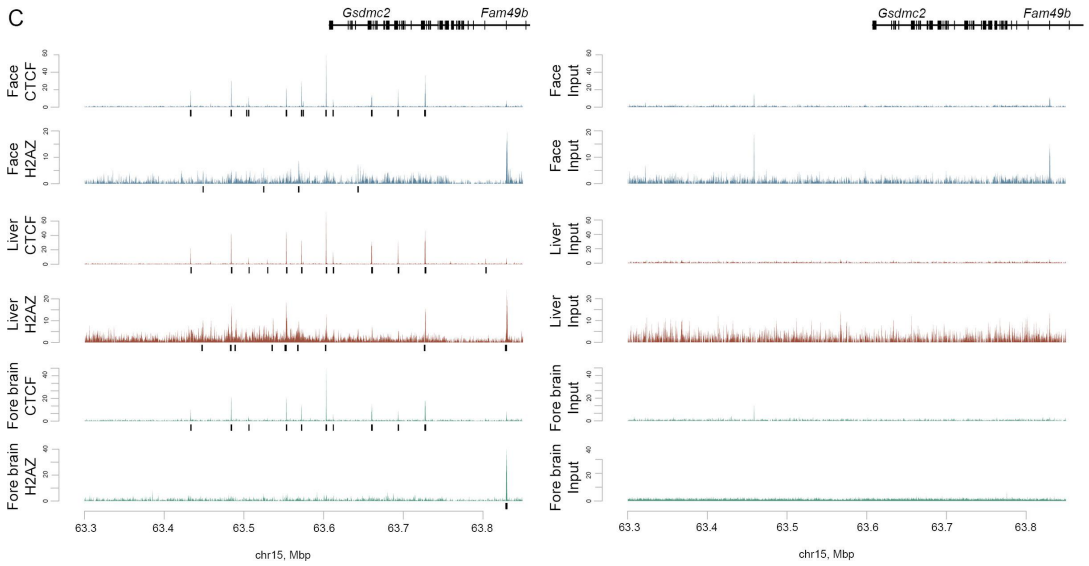


Figure 13 | Architectural proteins occupancy. H2AZ and CTCF ChIP-seq on E11.5 Wt (C57BL/6J) face (blue), liver (red) and fore brain (FB, green). y-axis shows the sequencing coverage (slightly different for each library). ChIP-seq experiments are shown on the left, the input controls, on the right. Peaks are shown underneath ChIP-seq tracks. A) View of *c-Myc* and neighboring *loci* (mm9 coordinates: chr15:60000000-64000000) showing enrichment of CTCF at the TAD-boundaries (Fam48b–*c-Myc*–TAD-boundary, c-Myc–subTAD-boundary) and at the telomeric end of *c-Myc locus* (Hi-C data is shown at the top of the figure (Rao et al., 2014)). Hi-C data is shown at the top of the panel (available at: http://promoter.bx.psu.edu/hi-c/view.php) (Rao SS et al., 2014). Similarly to CTCF, H2AZ shows enrichment on all the mentioned regions, except for the telomeric end of the *locus* in the FB (tissue with basal *c-Myc* expression). B) Detailed view of part of the telomeric c-Myc-TAD (chr15:62000000-63200000) showing only CTCF ChIP-seq tracks. The region contains both tissue-specific and tissue-invariant CTCF peaks. C) Detailed view of the surroundings of c-Myc–Fam49b–TAD-boundary (chr15:63300000-63850000) and telomeric end of *c-Myc locus* (mm9 coordinates: chr15:634200000-637500000). This region contains 10 CTCF binding sites shared in the three tissues analyzed. In the face and liver H2AZ is enriched at some of these CTCF sites. In the same tissues H2AZ displays a broader enrichment, spanning the entire end of *c-Myc locus,* and not necessarily localized into defined peaks.

These data obtained *in vivo* confirm what was observed on cell lines, proposing an enrichment of CTCF at TAD-boundaries (often occurring as clusters of binding sites), and suggest that CTCF occupancy at these regions (in *c-Myc locus*) is fairly constant across the three tissues analyzed (Zuin et al., 2013).

We detected a few differences on CTCF occupancy in the three tissues. The whole gene-desert contains two TADs, the more centromeric, including *Fam84b* (or Fam84b-TAD) and c-Myc-TAD. In both cases we observed CTCF binding sites occurring intra-TAD. Focusing on c-Myc-TAD, precisely on the centromeric subTAD, face, liver and FB shared two out of the six peaks present in the area (figure 13A). In the telomeric c-Myc-subTAD, with exception of the telomeric end of the *locus*, there were no CTCF peaks in the FB (figure 13A and B). Here, in the

other two tissues, face and liver, we saw altogether seven peaks, but only two of them were shared (figure 13A and B). Two out of the three liver specific CTCF peaks were located inside the *Pvt1* transcribed region (figure 13A). The two face specific peaks, lay more on the telomeric side of the subTAD, 400Kbp centromeric to the cluster of CTCF binding sites at the telomeric end of the *locus*.

We speculate that these few differences in intra-TAD CTCF binding could, in principle, be functional and reflect the fact that the MFM and liver coordinate *c-Myc* expression *via* different regulatory elements, and perhaps achieve so by implementing distinct intra-TAD sub-topologies.

5.3.2 Histone Variant H2AZ

H2AZ is a histone variant of the histone H2A. It is found preferentially in nucleosomes containing another histone variant, H3.3 (a variant replacing the histone H3). The nucleosomes containing this couple of histone variants are reported to be particularly unstable, and enriched at gene promoters, regulatory regions, transcription factor binding sites, as well as at CTCF binding sites (Jin & Felsenfeld, 2007; Jin et al., 2009; Yukawa et al., 2014). Despite the fact that for technical reasons it is not possible to perform H3.3 ChIP-seq with the standard procedure, the analysis of H2AZ alone is relevant to identify these unstable regions.

ChIP-seq experiments on E11.5 face, liver and FB revealed the presence of several tissue-invariant peaks located at both gene promoters and intergenic regions.

The same experiment also showed an important difference between the three tissues: the telomeric end of *c-Myc locus* was enriched for this histone variant in the face and liver, but not in the FB (figure 13C). All the other TAD-boundaries flanking regions in the *locus* showed similar H2AZ occupancy amongst the three tissues. Consistent with this observation, at the telomeric end of the *locus* we detected enrichment of the acetylation mark H3K27ac only in the face and liver, but not in the FB (figure 10A). Importantly, this part of the *locus*, which contains a tissue-invariant cluster of CTCF peaks, is also a TAD-boundary flanking sequence. The tissue-specific differences in enrichment of both H2AZ and H3K27ac that we observed could underlie different properties of this region in tissues where *c-Myc* expression is high, compared to the tissues where *c-Myc* expression is kept at a basal level.

5.4 Balanced Chromosomal Inversions in *c-Myc locus*

5.4.1 Inversion 1

One of the main purposes of the project is to gain insights into the influence the *locus* topology has on transcription. The reason why we decided to tackle this question by making three balanced chromosomal inversions is that these rearrangements provide us a way to reshuffle the genetic elements in *c-Myc locus*, without perturbing qualitatively and quantitatively the genetic context. On the basis of the position of the tissue specific enhancers relative to *c-Myc* promoter, as well as considering the TADs present in the *locus*, each inversion is designed to affect the *locus* topology in a specific way.

The inversion 1 (INV1) extends from Fam84b-TAD (GROMIT insertion 196231 (-17a or c17a)) to the end of *Pvt1* transcribed region (GROMIT insertion 194578 (3a)). Considering the effect of this perturbation in the MFM, we know that the enhancer(s) active in this tissue are located outside the inverted region, precisely on the telomeric side. The inversion brings *c-Myc* and *Pvt1* on the centromeric end of the *locus*, further from the facial enhancers compared to the Wt situation. In the liver, as we believe that the relevant enhancer(s) lie inside *Pvt1* transcribed region, the relationship between *c-Myc* promoter and these enhancer(s) should remain unaltered upon inversion. Concerning the topological organization, the two breakpoints of the inversion are in two different TADs: the centromeric one in Fam84b-TAD, and the telomeric one, in c-Myc-TAD. INV1 therefore mixes the two TADs together, and more importantly, it introduces a TAD-boundary (Fam84b–c-Myc–TAD-boundary) between *c-Myc* promoter and the telomeric side of the *locus* (figure 8). This configuration is particularly interesting for the facial tissues as in this case *c-Myc* promoter and the enhancers, other than being further away than in the Wt, are also separated by such a boundary.

The analysis of the LacZ reporter expression showed that the centromeric breakpoint 196231 (-17a) remained unresponsive to any transcriptional input even after inversion. Regarding the telomeric breakpoint, insertion 194578 (3a), the staining on the liver and the weak one on MFM, somites and PLM disappeared once this region was moved on the centromeric end of the *locus* (figure 14B). This is somehow expected for MFM, PLM and somites, as we know that the relevant enhancers are located telomeric to the insertion 194578 (3a). After inversion, the communication between LacZ promoter and enhancers may be affected by what is discussed above, namely a greater distance that separates them and/or the

interposition of a TAD-boundary in between. The case of the liver is less clear. According to our model, for which the liver enhancer(s) may be found inside *Pvt1* transcribed region, nothing changes between the insertion 194578 (3a) and these enhancer(s). What changes instead is the context in which these sequences operate. We believe that this context is crucial for any regulatory element to function correctly.

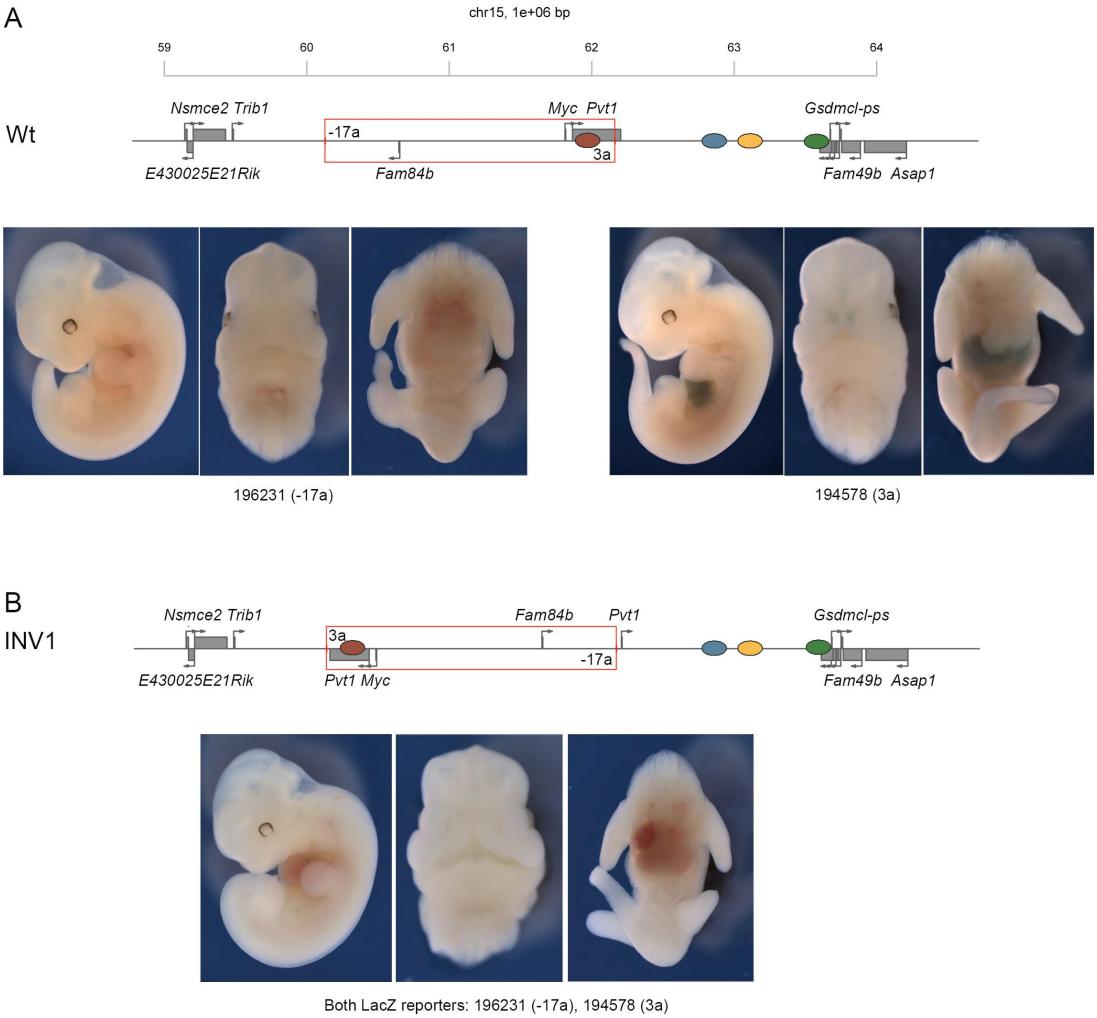


Figure 14 | INV1 regulatory activity at the inversion breakpoints. A) LacZ reporter staining in E11.5 196231 (-17a, or c17) and 194578 (3a) GROMIT lines (paternal allele). The approximate location of the tissue-specific enhancers is depicted by ellipses, red for the liver, blue for the medial face mesenchyme (MFM) and yellow for the nasal epithelia (NE). The telomeric end of the *locus,* which contains a cluster of CTCF binding sites is represented by a green ellipse (same color coding as for the panel B). B) LacZ reporter staining on an INV1 E11.5 embryo, carrying LacZ reporter at both inversion breakpoints (paternal allele). INV1 causes the complete loss of reporter activity on liver, MFM, limbs and somites (very weak on the last three tissues) from the 194578 (3a) breakpoint.

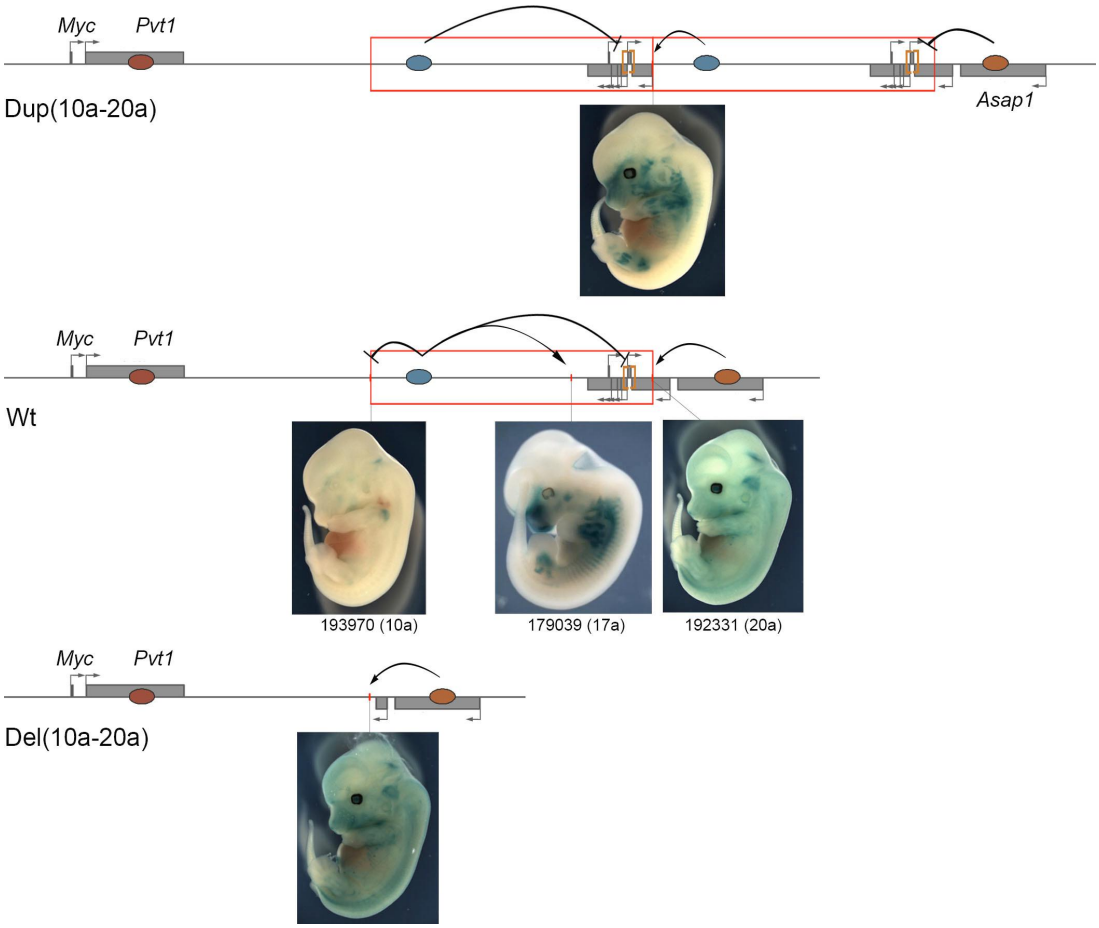
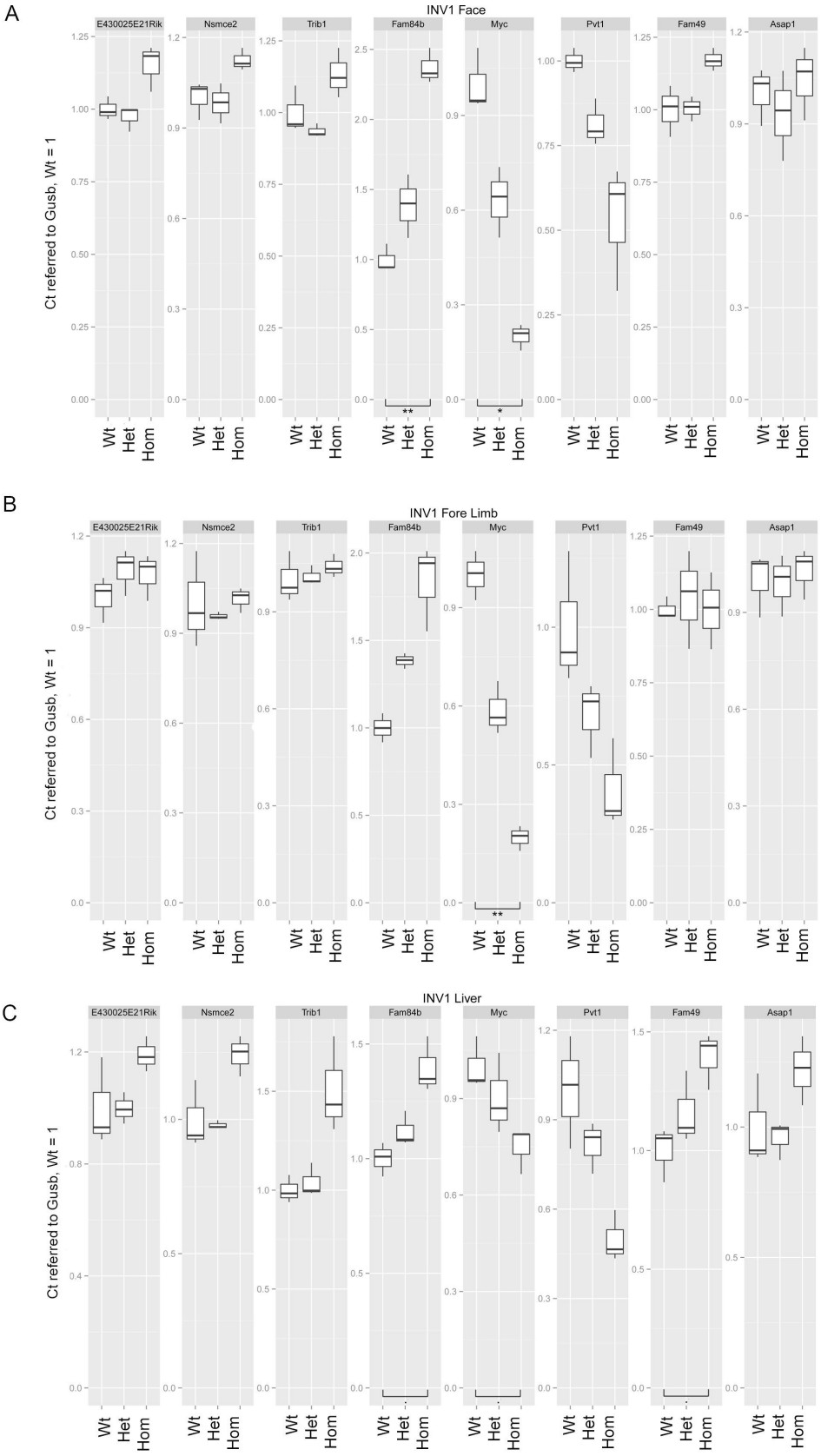


Figure 15 | Context dependency of promoter–long-range enhancers communication. Material shown thanks to kind concession of Veli Uslu, representing LacZ reporter staining in E11.5 embryos from Wt (GROMIT insertions 179039 (17a), 193970 (10a) and 192331 (20a)), Dup(10a–20a) and Del(10a–20a). In the Wt, facial, nasal epithelia, limbs and somites enhancer(s) (just shown the facial and liver enhancer(s), as blue and red ellipses respectively) reach the GROMIT insertion 179039 (17a), but do not act on the blank spot 193970 (10a). Outside of *c-Myc locus*, on the insertion 192331 (20a), other enhancer(s) (*Fam49b-Asap1* enhancer(s), depicted as a brown ellipse) give a widespread reporter expression scattered throughout the embryo. *c-Myc* and *Fam49b-Asap1* enhancers act within their regulatory domain likely because of the presence of a boundary element, probably located nearby the gene *Gsdmcl-ps* (inferred from the TAD disposition observed on Hi-C data, represented by sand color square-brackets). Upon duplication Dup(10a–20a), the LacZ staining acquires the typical pattern of the insertion 179039 (17a), likely because of the effect of the enhancers located on the telomeric segment of the duplication. In this instance however, the relationship between *c-Myc* enhancer(s) and the promoter of the reporter is identical to the Wt scenario for the insertion 193970 (10a), which gives no staining in any tissue. It is possible that the duplication altered the overall genetic context, changing the enhancer(s) range of action. Another possibility is that the duplication displaced the LacZ reporter from a hypothetical repressor, located centromeric to the insertion 193970 (10a). However, the deletion Del(10a–20a) gives a LacZ staining pattern comparable to the one of the insertion 192331 (20a) and rules out this possibility.

Veli Uslu observed a similar case before. In this instance, the GROMIT insertion 193970 (10a), integrated in the middle of the telomeric side of *c-Myc locus,* did not respond to any enhancer (figure 15) (it is one of the few blank spots inside the telomeric side of the *locus*). Veli Uslu made a chromosomal duplication between this insertion and 192331 (20a), occurring inside the *Fam49b* transcribed region (telomeric to the *Gsdmcl-ps* boundary). After the duplication, the remaining GROMIT insertion captured transcriptional inputs recapitulating the characteristic *c-Myc* pattern seen at the insertion 179039 (giving the typical staining in FM, NE, PLM, BA, and somites). In this duplication the enhancers were also duplicated, having the LacZ reporter in the middle. However, in this configuration the enhancers located on the centromeric segment of the duplication were separated from the LacZ reporter by the c-Myc–Fam49b–TAD-boundary, and therefore, it is likely that the transcriptional inputs reaching the reporter came uniquely from the enhancers on the telomeric segment of the duplication. Importantly, the sequence in between the LacZ reporter and the enhancers present in this telomeric segment remained exactly the same. Also in this case, what changed was the context surrounding them. Veli Uslu hypothesized that it was the different context that allowed the blank spot 193970 (10a) to reach the enhancers and acquire reporter expression in FM, NE, PLM, BA and somites upon duplication.

We analyzed the expression changes of the transcripts inside the *locus*: *c-Myc*, *Pvt1* and *Fam84b*, as well as the ones at the neighboring *loci E430025E21Rik*, *Nsmce2*, *Trib1*, *Fam49b* and *Asap1*. We used E11.5 face, liver, proximal fore limb mesoderm (PFLM) and heart as control tissue (tissue where *c-Myc* is expressed at basal level). We only observed significant changes in expression of *c-Myc* and *Fam84b*. *c-Myc* expression in INV1 homozygote face and PFLM was reduced to approximately the basal *c-Myc* transcription, being ~ 20% of the Wt level (equivalent to the deletion of the facial enhancers, seen on Del(8–17)) (figure 16A and B). This data supports the hypothesis of *c-Myc* promoter being separated from the facial (and limb) enhancers by the introduction of a TAD-boundary in between them. The disappearance of the LacZ reporter expression from the insertion 194578 (3a) upon INV1 is consistent with this observation. On the other hand we observed in the face, a concomitant upregulation of *Fam84b* of 2.3 times the Wt level in INV1 homozygotes (figure 16A). The PFLM had a similar trend although it was not statistically significant (figure 16B). We believe that this effect was consequent to the hijacking of the facial (and probably also the FL) enhancer(s), usually acting on *c-Myc*, by *Fam84b*.


A
INV1 Face
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom
B
INV1 Fore Limb
C
INV1 Liver

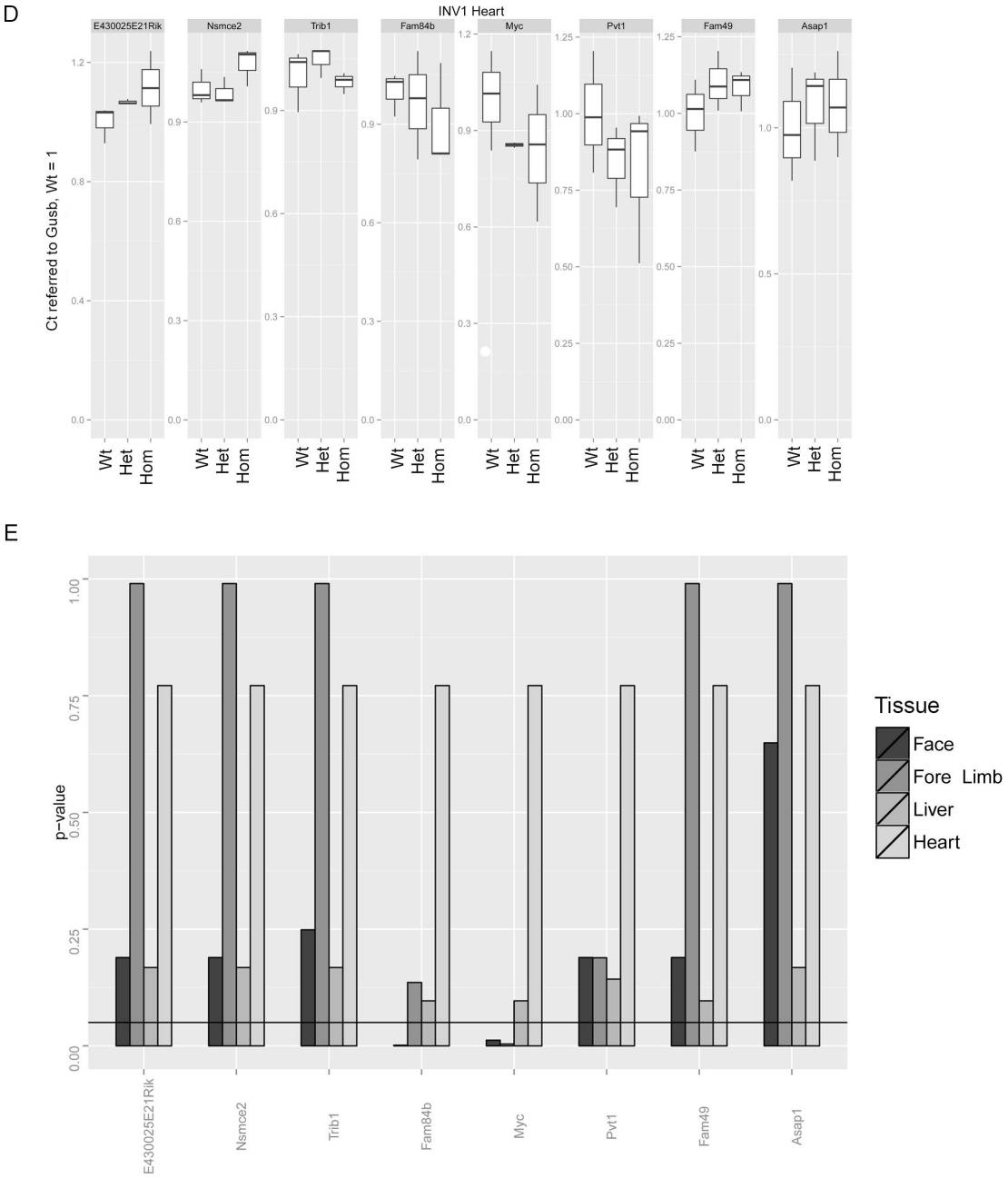


Figure 16 | INV1 *c-Myc* and neighboring transcripts expression. cDNA-qPCR on E11.5 INV1$^{-/-}$, INV1$^{+/-}$, INV1$^{+/+}$ facial mesenchyme, proximal fore limb mesoderm, liver and heart (sample size *n* = *3*, 46–48 somite pairs stage, genotypes balanced for litter provenience within comparison groups). A–D) Individual panels representing the four tissues analyzed. The transcripts Ct are referred to the housekeeper gene *Gusb* and divided by the median Wt value (y-axis). Significance codes: *p*-value > 0.1 ns; 0.05 ≤ *p*-value < 0.1 " . "; 0.01 ≤ *p*-value < 0.5 " * "; 0.001 ≤ *p*-value < 0.01 " ** "; *p*-value < 0.001 " *** ". E) Benjamini-Hochberg corrected *p*-values.

In INV1, *Fam84b* is brought closer to the facial enhancers compared to the Wt, and with this arrangement, it is no longer affected by the presence of a TAD-boundary between them. Heart and liver had minor and not significant changes in *c-Myc* expression (figure 16C, D and E).

Finally, INV1 is lethal in homozygote animals. We believe the death occurs during embryonic development, between 12.5 and 18.5 days of gestation. Although INV1 led to both gene upregulation and downregulation (in the tissues we analyzed), the death of the embryos limited to homozygote animals led us to hypothesize a loss-of-function mutation. We tested this possibility by crossing INV1 heterozygotes with Del(c17–17) (deletion of the whole *c-Myc locus*, GROMIT insertions 196231 (-17a)–192857 (17a)). Out of the 24 pups we screened, we genotyped no double positive animal. We inferred a significant deviation from mendelian proportions (*p*-value 0.046, Pearson's Chi-squared test with Yates' continuity correction). The other three genotypes instead appeared with the same frequency, confirming the lethality in INV1 homozygotes to be due to a loss-of-function mutation of *c-Myc* and/or *Pvt1*. Although we do not know which tissue(s) are dysfunctional in INV1 homozygotes embryos and cause the lethality, we saw downregulation of *c-Myc* in the face and PFLM, and for this reason we speculate that the lethality we saw may have to do with a downregulation of *c-Myc* itself.

5.4.2 Inversion 2

Inversion 2 (INV2) was designed to the test the importance of the telomeric end of the *locus*, *i.e.*, the ~ 200Kbp ranging from the insertion 179039 (17a) to the gene *Gsdmcl-ps* (location of the c-Myc–Fam49b–TAD-boundary), termed segment-T (segment telomeric to the insertion 179039 (17a), while the region encompassing the ~ 130Kbp centromeric to this insertion is termed segment-C). We hypothesized from chromatin data that the whole telomeric end of the *locus* may be important for *c-Myc* regulation in tissues where its expression is higher than the basal level. The region contains a cluster of 10 CTCF binding sites, present in all the analyzed tissues, however it has H3K27ac and H2AZ enrichment only in tissues with high *c-Myc* expression (FM and liver). GROMIT insertions integrated in this region, give LacZ staining patterns similar to the one of 179039 (17a), resembling (with exception of the liver) the endogenous *c-Myc* expression pattern (observed by *in-situ* hybridization, figure 7). Hi-C experiments performed on cell lines show that this

region contacts other TAD edges inside the *locus*. Similarly, it is also possible that in embryonic tissues the telomeric end of *c-Myc locus* interacts with other sequences inside the *locus*, including tissue specific enhancers and *c-Myc*. INV2 uses as the centromeric breakpoint the insertion 196231 (-17a or c17a, same as for INV1), located at the centromeric side of the *locus*, inside Fam84b-TAD.

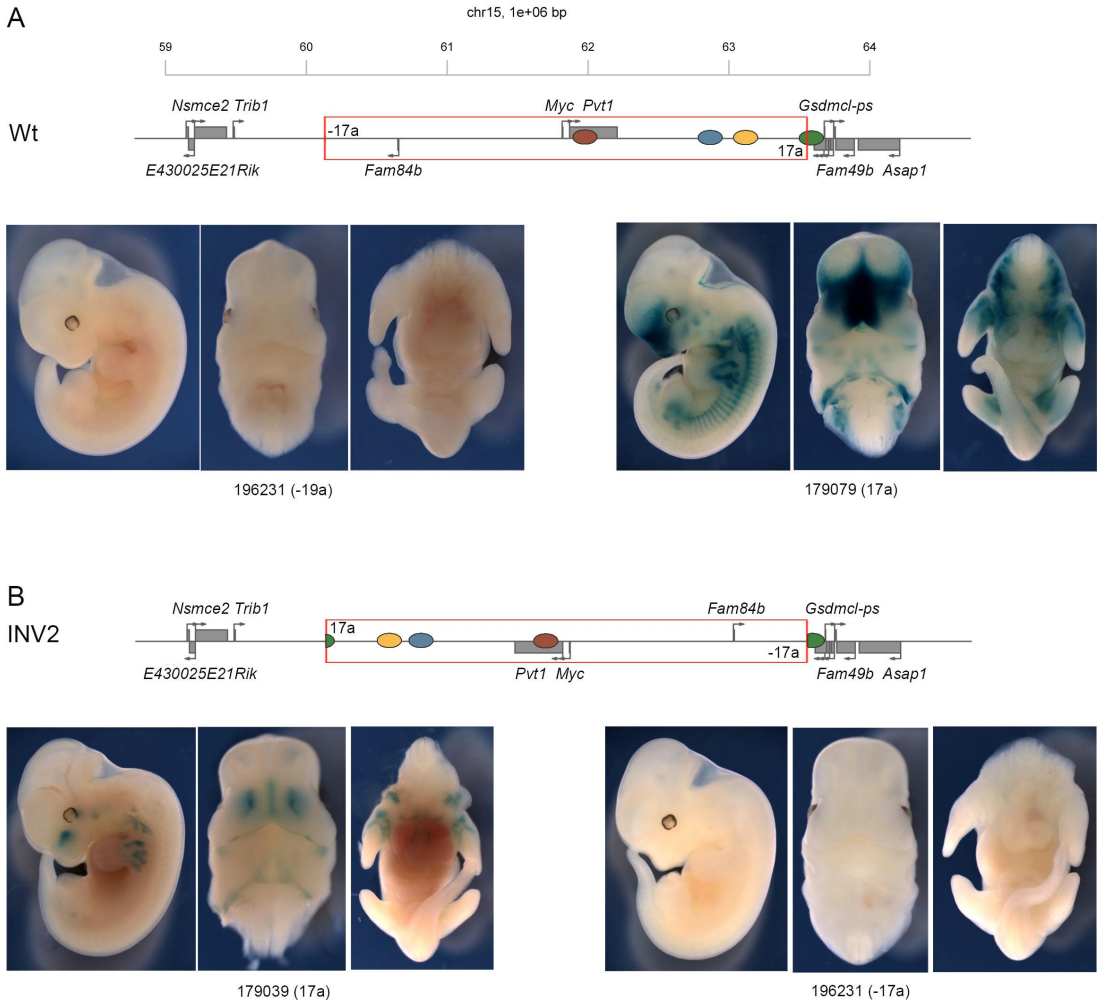
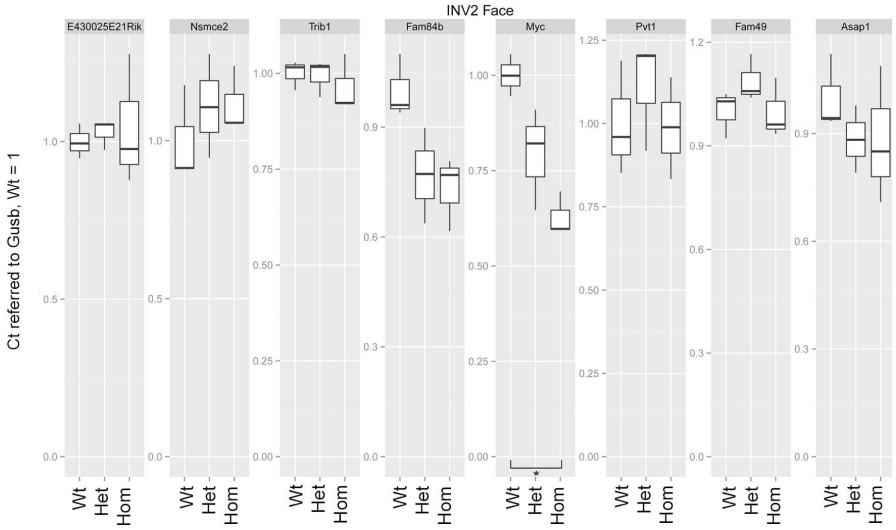
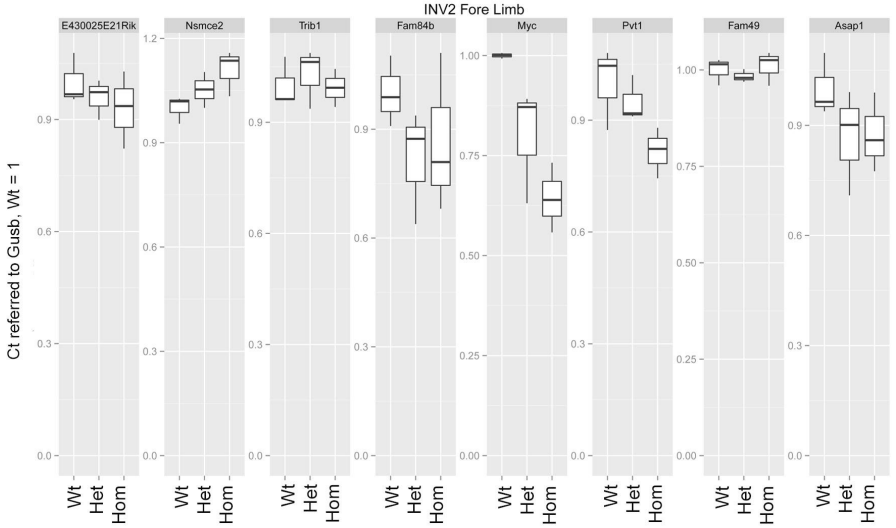


Figure 17 | INV2 regulatory activity at the inversion breakpoints. A) LacZ reporter staining of E11.5 196231 (-17a, or c17) and 179039 (17a) GROMIT lines (paternal allele). The approximate location of the main cis-acting DNA elements of the *locus* is depicted on the gene model: red, liver enhancer(s); blue, facial enhancer(s); yellow, nasal epithelia enhancer(s); green, telomeric end of the *locus* (represented also on the B panel). B) LacZ reporter staining on an INV2 E11.5 embryo, carrying a single LacZ reporter at either of the inversion breakpoints (paternal allele). This analysis shows no changes in LacZ staining at the breakpoint 196231 (-16a, or c16), silent in both Wt and INV2. The breakpoint 179039 (17a) shows a reduction of LacZ staining, compared to the Wt, in medial face mesenchyme, lateral face mesenchyme, nasal epithelia, proximal fore/hind limb mesoderm, somites, rhombic lip and branchial arches.

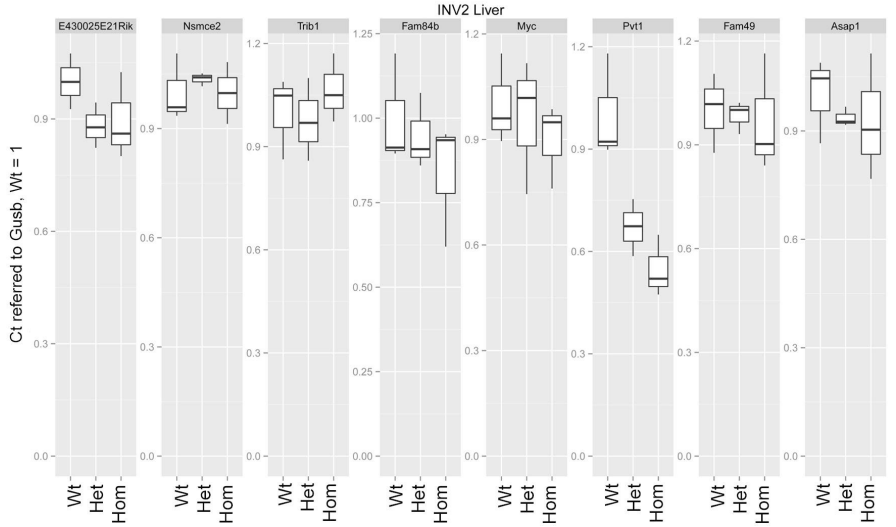
A


INV2 Face
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom
*

B


INV2 Fore Limb
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom

C


INV2 Liver
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom

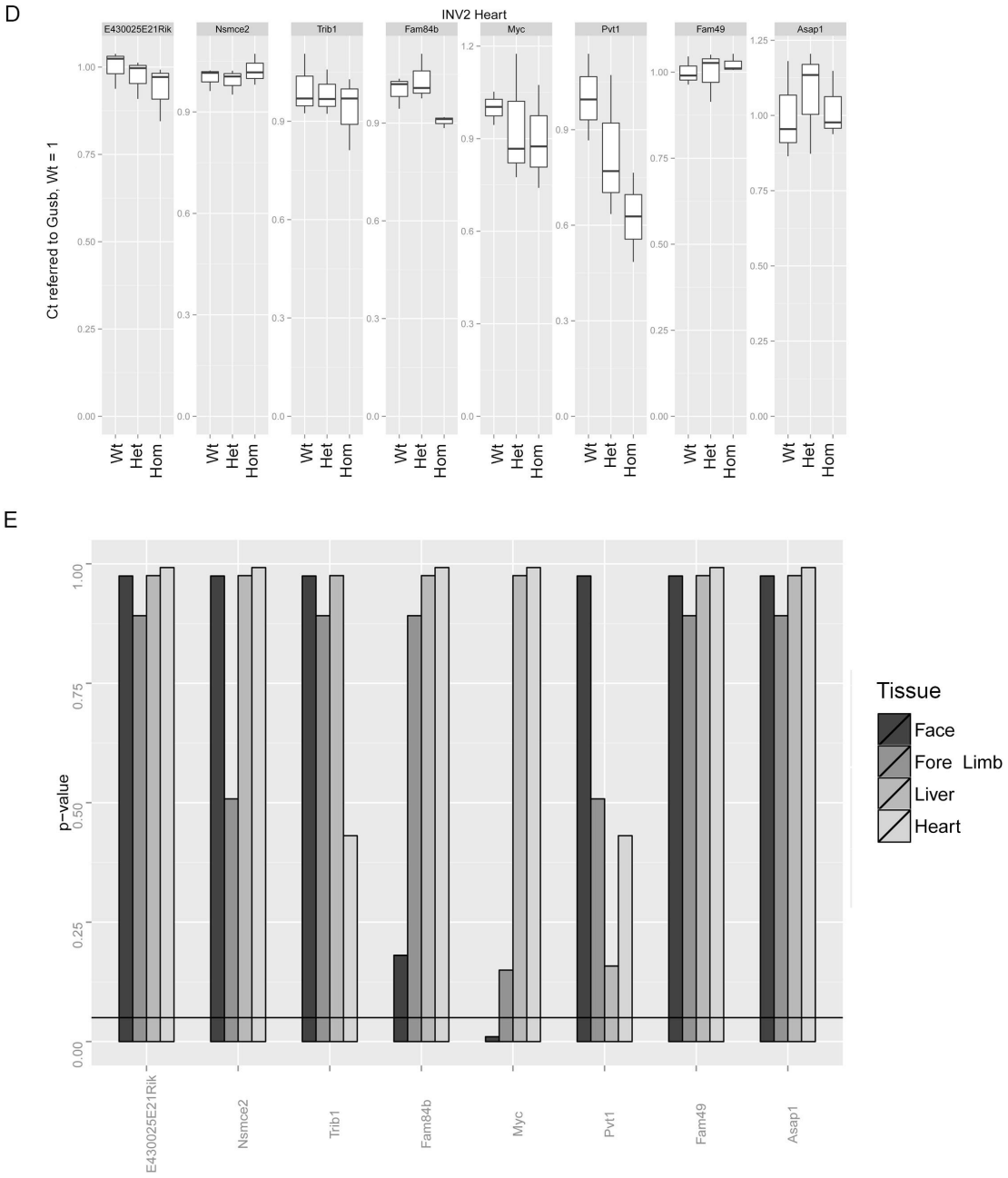


Figure 18 | INV2 *c-Myc* and neighboring transcripts expression. cDNA-qPCR on E11.5 INV2$^{-/-}$, INV2$^{+/-}$, INV2$^{+/+}$ facial mesenchyme, proximal fore limb mesoderm, liver and heart (sample size *n* = *3*, 46–48 somite pairs stage, genotypes balanced for litter provenience within comparison groups). A–D) Individual panels representing the four tissues analyzed. The transcripts Ct are referred to the housekeeper gene *Gusb* and divided by the median Wt value (y-axis). Significance codes: *p*-value > 0.1 ns; 0.05 ≤ *p*-value < 0.1 " . "; 0.01 ≤ *p*-value < 0.5 " * "; 0.001 ≤ *p*-value < 0.01 " ** "; *p*-value < 0.001 " *** ". E) Benjamini-Hochberg corrected *p*-values.

As for the telomeric breakpoint, INV2 uses the insertion 179039 (17a), located between the third and fourth CTCF binding site of the cluster (for both face and liver). INV2 splits the telomeric end of the c-Myc-TAD in two parts, segment-C and segment-T. Except for the segment-T, which is not included in the inversion, INV2 leaves c-Myc-TAD primary sequence almost unaltered. On the other hand, this inversion may have some effect on the Fam84b-TAD, as it breaks it into two parts, fusing the centromeric one with c-Myc-TAD; except for this one, INV2 does not introduce other perturbations. In contrast to INV1, INV2 does not reshuffle the position of promoters and enhancers relative to each other, and therefore, it preserves their arrangement inside the *locus*.

We analyzed the LacZ reporter expression coming from both breakpoints of the inversion. Similarly to the Wt, the insertion 196231 (-17a) gave no LacZ staining upon inversion (figure 17B). The insertion 179039 (17a) instead, showed a marked reduction of reporter expression in all the responsive tissues typical of 17a (FM, NE, PLM, BA, RL and somites). Conversely to the case of INV1, where the general genetic context affected the activity of the liver enhancer(s) on the LacZ reporter, in INV2, the widespread reduction of LacZ staining from the insertion 179039 (17a), suggests that the segment-T, which remains at the telomeric end of c-Myc-TAD, in particular, is important for the tissue specific enhancers to reach the LacZ promoter.

RT-qPCR assay showed a significant reduction of *c-Myc* expression in the face to 60% of the Wt level (figure 18A). We observed no other significant changes of expression of other transcripts in *c-Myc* and neighboring *loci* in the tissues analyzed (face, PFLM, liver and heart, figure 18B, C, D and E). We conclude from this experiment that the segment-T is important for the facial enhancers to act throughout the *locus*, on the 17a position and on *c-Myc*.

5.4.3 Inversion 3

Inversion 3 (INV3) uses as the telomeric breakpoint the insertion 192857 (17a), occurring at the same genomic position as 179039 (17a), and in this respect, it has some similarities with INV2. As for INV2, INV3 was designed to separate the first three CTCF binding sites (on the segment-C) from the rest of the cluster, remaining at the telomeric end of the *locus*. However, the centromeric breakpoint is the insertion 194578 (3a), situated at the end of *Pvt1* transcribed region, and therefore, conversely to INV2, this inversion occurs intra-TAD.

LacZ staining assay in INV3 revealed a weakening of reporter transcription from the insertion 192857 (17a) in FM, NE, BA, RL, somites and PLM, and an increment in the liver (figure 19B). Similar to INV2, we interpret the weakening of the staining pattern of the insertion 179039/192857 (17a) as the effect of the separation of the reporter from the segment-T, not included in the inversion.

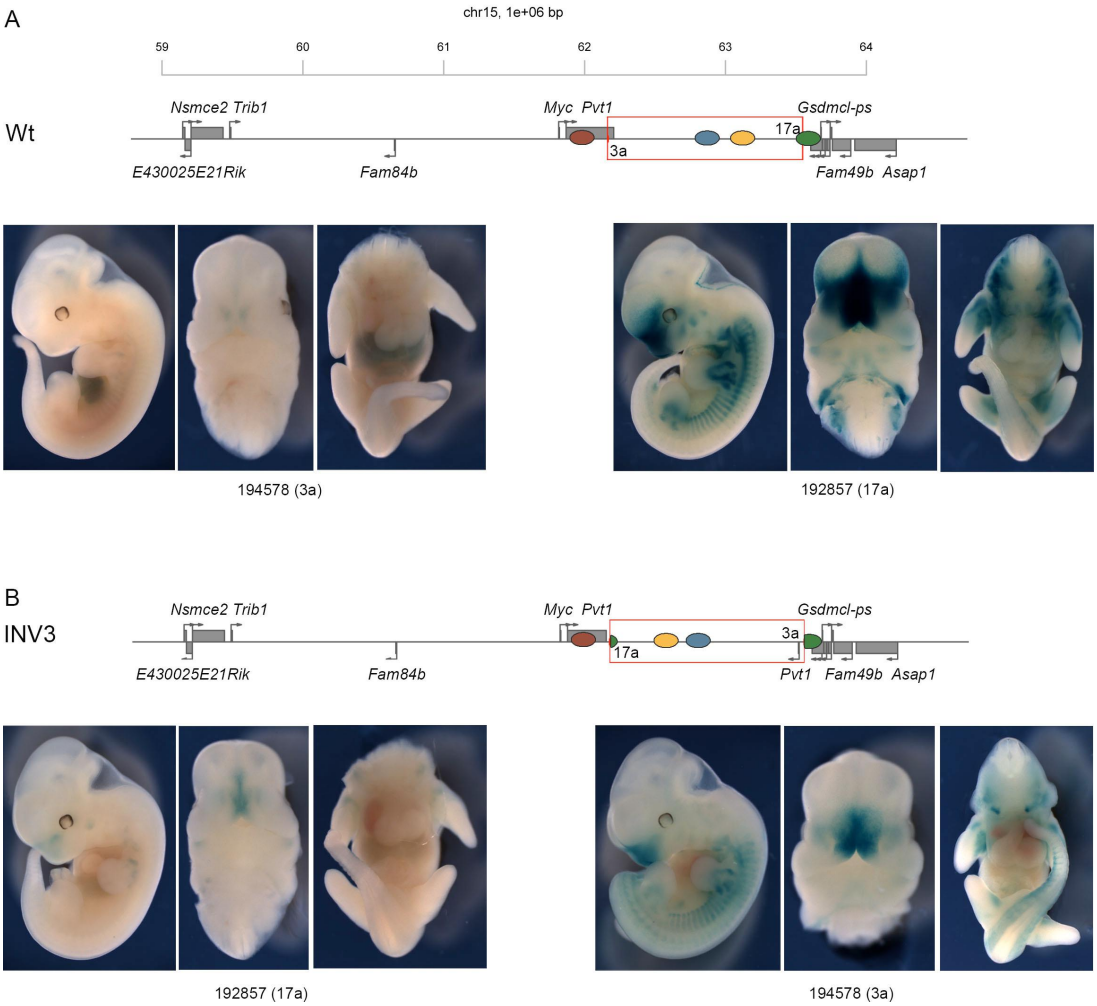
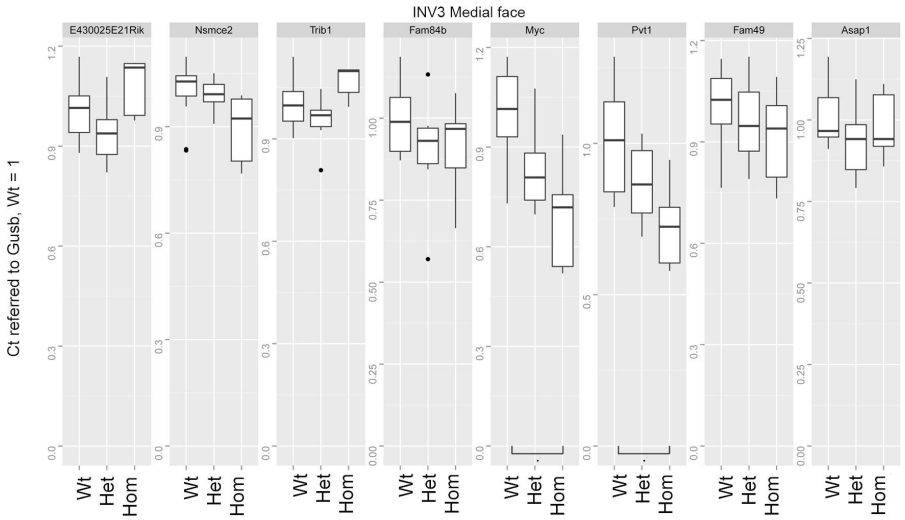
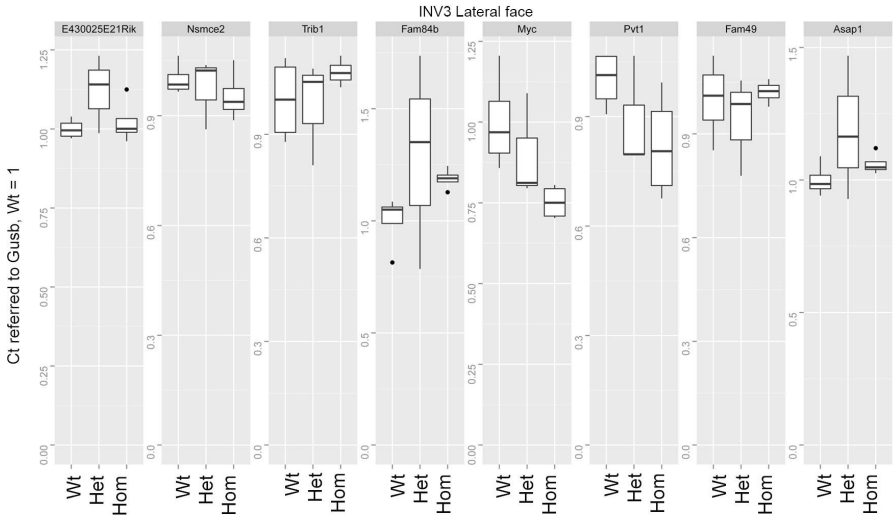


Figure 19 | INV3 regulatory activity at the inversion breakpoints. A) LacZ reporter staining of E11.5 194578 (3a) and 192857 (17a) GROMIT lines (paternal allele). Cis-acting DNA elements (approximate location): red, liver enhancer(s); blue, facial enhancer(s); yellow, NE enhancer(s); green, telomeric end of the *locus* (represented also on the B panel). B) LacZ reporter staining on an INV3 E11.5 embryo, carrying a single LacZ reporter at either of the inversion breakpoints, 194578 (3a) and 192857 (17a), on the paternal chromosome. The GROMIT insertion 192857 (17a) upon INV3 shows an increment of LacZ staining in the liver, as well as a marked reduction in all the 192857 (17a) characteristic tissues, medial/lateral face mesenchyme, nasal epithelia, proximal fore/hind limb mesoderm, somites, rhombic lip and branchial arches. The insertion 194578 (3a) shows a gain of LacZ staining in medial/lateral face mesenchyme, nasal epithelia, proximal fore/hind limb mesoderm, somites and branchial arches and a complete loss of staining in the liver.

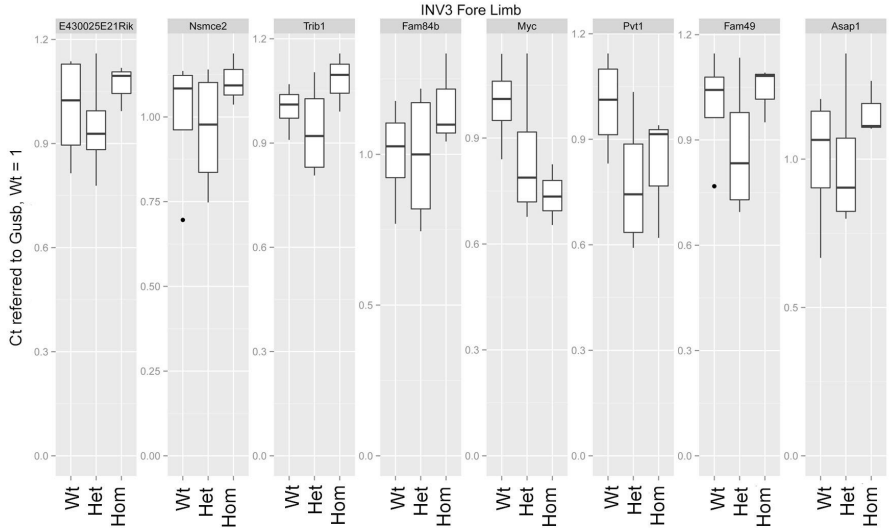
A


INV3 Medial face
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom

B


INV3 Lateral face
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom

C


INV3 Fore Limb
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom

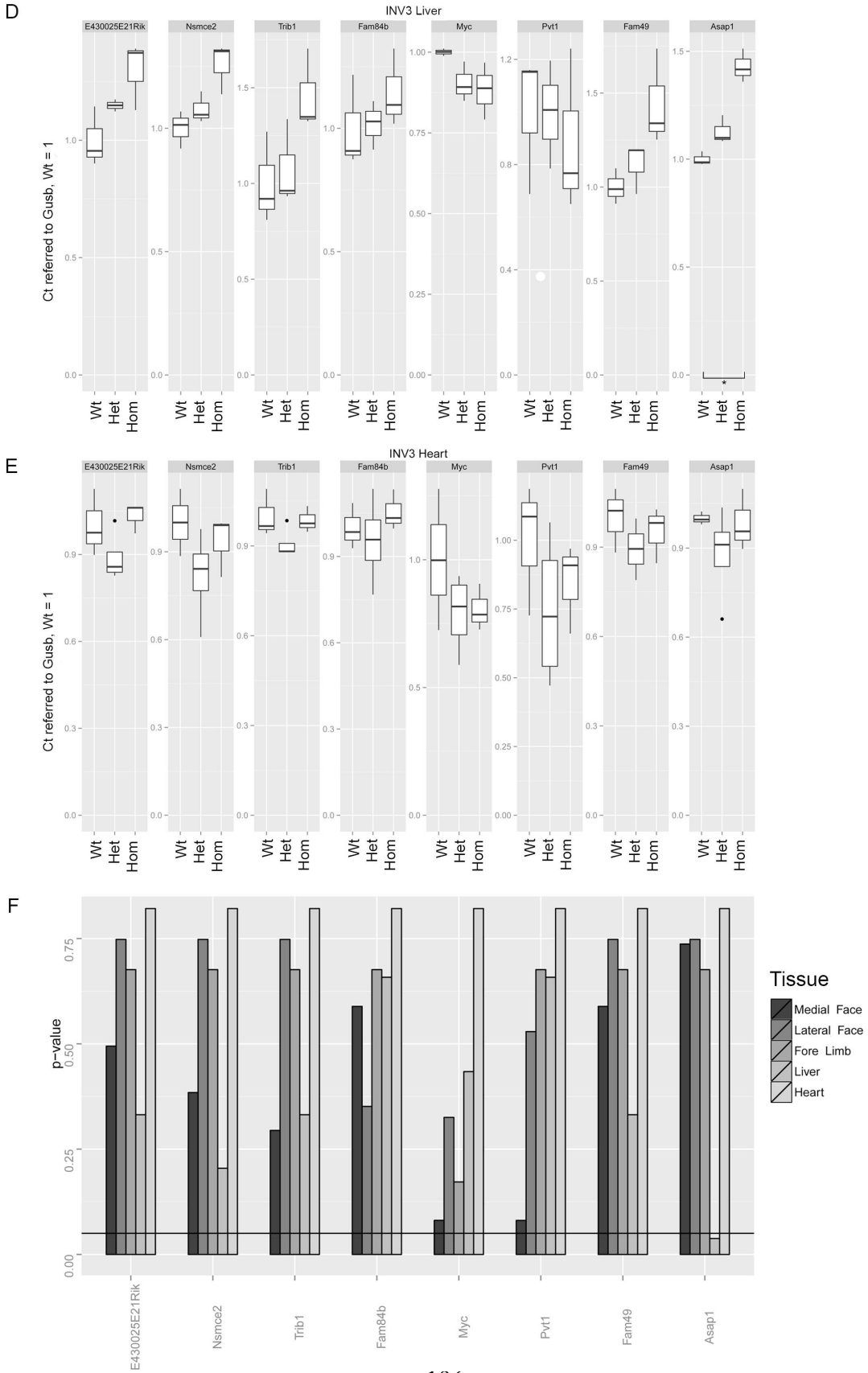

D
INV3 Liver
E430025E21Rik
Nsmce2
Trib1
Fam84b
Myc
Pvt1
Fam49
Asap1
Ct referred to Gusb, Wt = 1
Wt
Het
Hom
E
INV3 Heart
F
p−value
Tissue
Medial Face
Lateral Face
Fore Limb
Liver
Heart

Figure 20 | INV3 *c-Myc* and neighboring transcripts expression. cDNA-qPCR on E11.5 $INV3^{-/-}$, $INV3^{+/-}$, $INV3^{+/+}$ medial/lateral face, proximal fore limb mesoderm (PFLM), liver and heart (sample size $n = 3$ for PFLM, liver and heart, $n \geq 4$ for lateral face and $n \geq 5$ for medial face, 46–48 somite pairs stage, genotypes balanced for litter provenience within comparison groups for PFLM, heart and liver, unbalanced for lateral face and medial face). A–E) Individual panels representing the five tissues analyzed. The transcripts Ct are referred to the housekeeper gene *Gusb* and divided by the median Wt value (y-axis). Significance codes: p-value > 0.1 ns; 0.05 ≤ p-value < 0.1 " . "; 0.01 ≤ p-value < 0.5 " * "; 0.001 ≤ p-value < 0.01 " ** "; p-value < 0.001 " *** ". F) Benjamini-Hochberg corrected p-values.

In this case, the activity of the segment-C does not seem to be sufficient to compensate for the function of the whole telomeric end of the *locus*. On the other hand, the increment of reporter activity in the liver may be explained by an increased proximity to the liver enhancer(s). However, this increment did not reach the level of the Wt insertion 194578 (3a).

Regarding the insertion 194578 (3a), after inversion the reporter gained expression in all the typical tissues of 179039/192857 (17a), and lost the expression in the liver (figure 19B) completely. The increment in FM, PLM, BA and somites, despite not reaching the level of 179039/192857 (17a), confirms the general importance of the segment-T to direct enhancers to the LacZ reporter inserted at the native 17a position.

We analyzed the expression levels of the transcripts inside *c-Myc locus* and in the neighboring *loci*, in face, PFLM, liver and heart. We saw no significant changes, except for an upregulation to 1.4 times the Wt level of *Asap1* in the liver (figure 20D). We saw similar trends, although not reaching statistical significance, for *E43000E21Rik, Nsmce2, Trib1* and *Fam49b* in this tissue, but not in the face, PFLM or heart. This suggests that upon INV3, the effect of liver enhancer(s) may leak to other genes, on both centromeric and telomeric sides. The fact that this effect is limited to the liver, supports the idea that the communication between promoters and enhancers may vary from tissue to tissue.

Regarding the face, we performed this experiment twice (subdividing the face into medial face and lateral face (the NE is split between these two tissues), in case these two tissues have effects not seen as a whole face, figure 20A and B), and we observed a reproducible, although non-significant, downregulation of *c-Myc* to 75% of the Wt level in both tissues. We think that this result may confirm the observation drawn from INV2, that the telomeric end of the *locus*—even if not equipped with enhancers—plays a role in establishing *c-Myc* promoter–enhancers communication. Moreover it suggests that genetic reshuffling occurring intra-TAD may lead to small changes in gene expression, probably due to perturbations in long-range interactions.

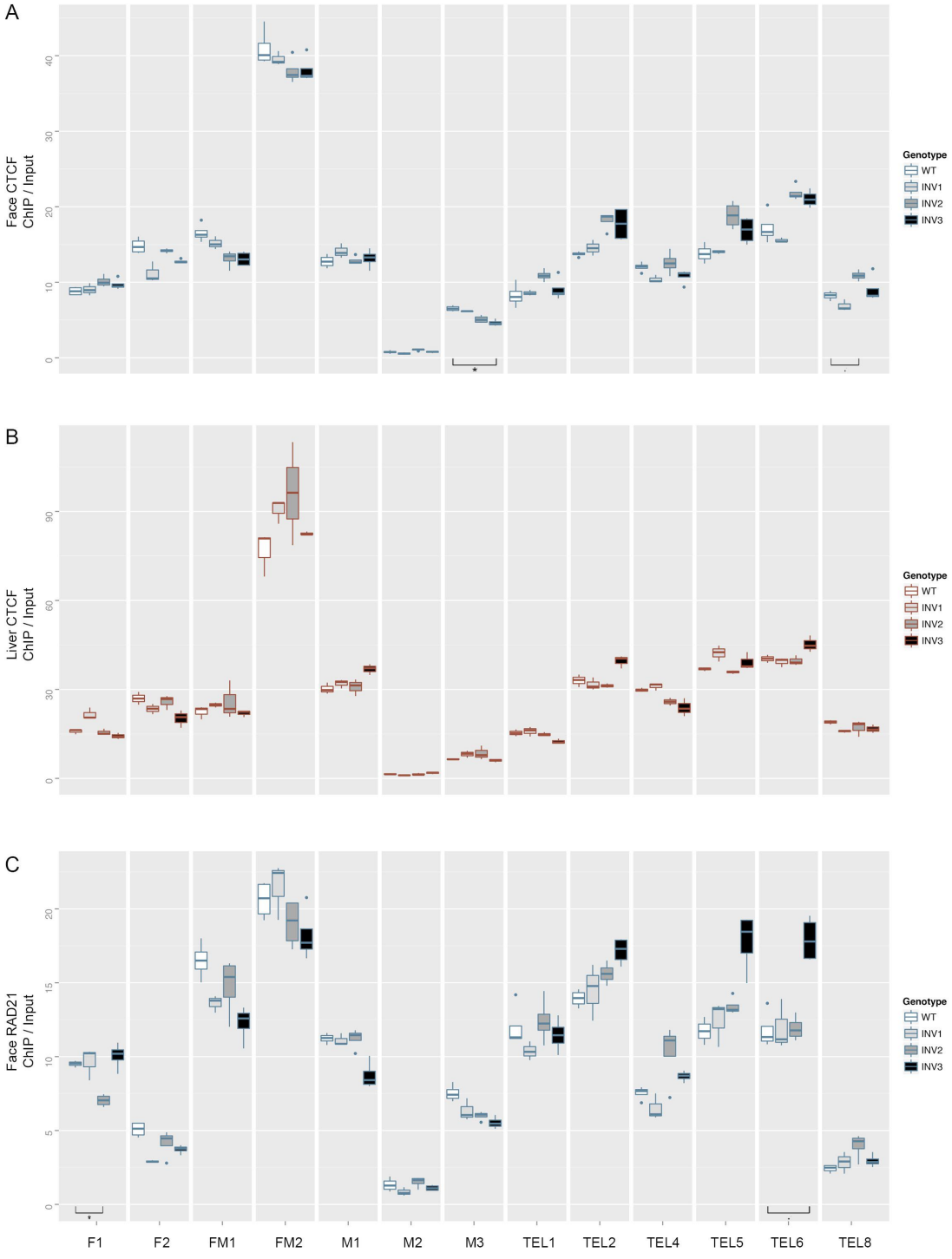
5.5 Redistribution of Architectural Proteins in Inversion Lines

After analyzing the functional effects of the chromosomal inversion on *c-Myc locus*, we wanted to test whether the effects we observed were accompanied by changes in the occupancy of architectural proteins inside the *locus*. This experiment performed on both Wt and inversion lines, aimed to question the importance of the architecture of *c-Myc locus* to determine the binding of architectural proteins.

As we knew from the ChIP-seq experiments the location of CTCF binding sites in face and liver, we opted for a targeted approach, ChIP-qPCR, that offered us several technical advantages compared to the ChIP-seq. We assessed 13 main CTCF and RAD21 (Cohesin subunit) binding sites of the *locus*. We performed the experiment in face and liver from Wt (C57BL/6J), INV1, INV2 and INV3 homozygotes E11.5 embryos. For each line we performed at least three biological replicates and at least four qPCR replicates.

We saw significant changes in CTCF occupancy only on the M3 binding site in the face INV3, with a reduction to 74% of the Wt level (figure 21A). We also detected a reduction to 70% of the Wt level of RAD21 binding on the site F1 in INV2 face (figure 21C).

For this kind of procedure, similar changes are commonly seen also on the control regions (designed in order to be unaffected (in theory) by the experimental treatments applied). This effect probably reflects a natural variability between different lines (supplementary figure 2). Other than the mentioned cases, we observed no other significant changes. Therefore, we conclude from this experiment that the occupancy of CTCF and the Cohesin subunit do not show major changes upon genetic engineering.


A
Face CTCF
ChIP / Input
0
10
20
30
40
B
Liver CTCF
ChIP / Input
0
30
60
90
C
Face RAD21
ChIP / Input
0
5
10
15
20
F1
F2
FM1
FM2
M1
M2
M3
TEL1
TEL2
TEL4
TEL5
TEL6
TEL8
Genotype
WT
INV1
INV2
INV3

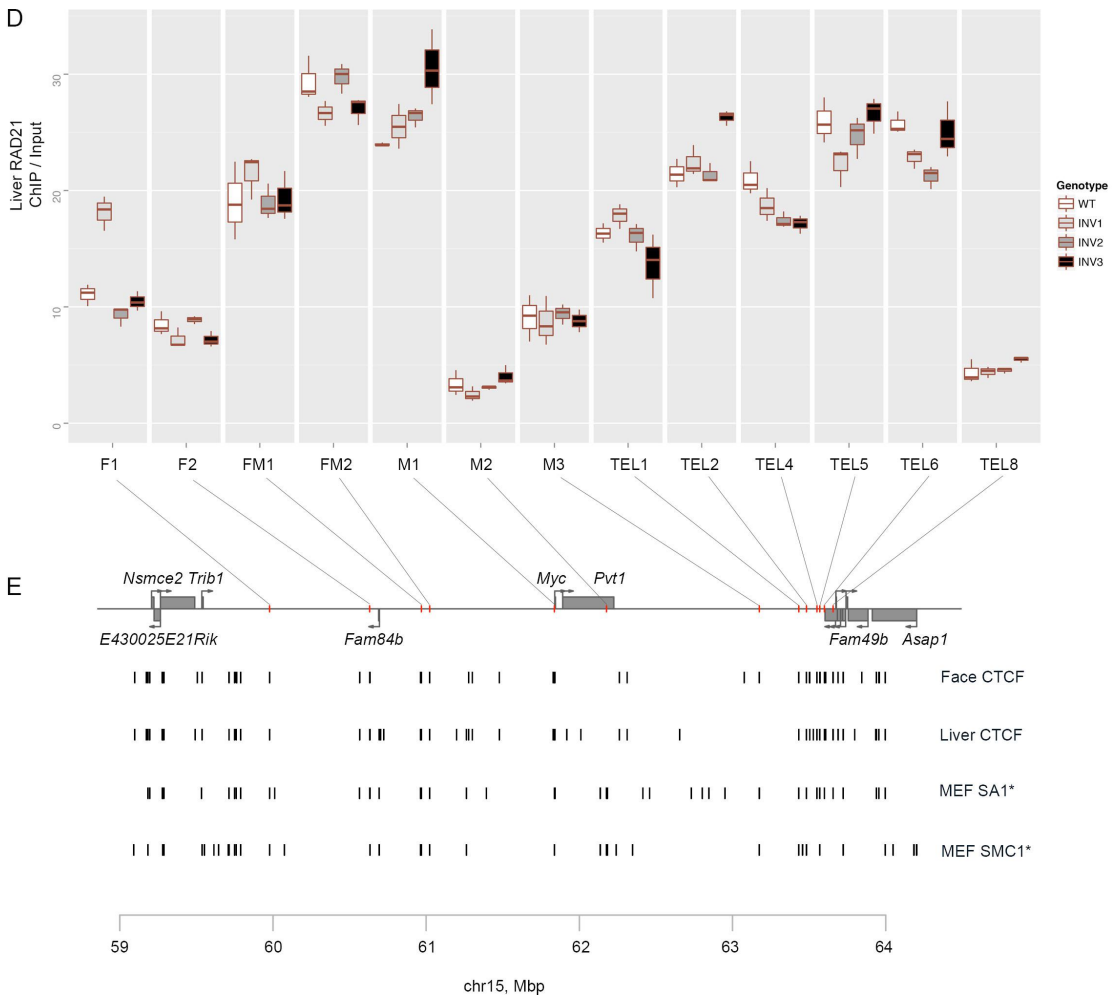


Figure 21 | Architectural proteins occupancy in inversion lines. A–D) CTCF and RAD21 (cohesin subunit) occupancy at selected regions of the *locus* (see panel E for details) in Wt, $INV1^{+/+}$, $INV2^{+/+}$, $INV3^{+/+}$ E11.5 face and liver (sample size $n \geq 3$). The y-axis represents the fold enrichment of ChIP samples to the input chromatin control (background = 1). Significance codes: p-value > 0.1 ns; 0.05 ≤ p-value < 0.1 " . "; 0.01 ≤ p-value < 0.5 " * "; 0.001 ≤ p-value < 0.01 " ** "; p-value < 0.001 " *** ". The p-values are corrected for family-wise error rate using the Benjamini-Hochberg method. F = Fam84b-TAD; FM = Fam84b–c-Myc–TAD-boundary; M = c-Myc-TAD; TEL = Telomeric end of *c-Myc locus* (the numbers represent the CTCF peak number of figure 23A). E) Gene model showing in red the location of the individual PCR amplicons (inside *c-Myc locus*) and the peaks of CTCF and cohesin ChIP-seq experiments. The analyzed regions were selected for containing CTCF and/or cohesin binding sites. They represent the closest peaks to the inversions breakpoints, TAD-boundaries and most of the tissue invariant peaks at the telomeric end of *c-Myc locus.* The two upper peak calling tracks refer to CTCF ChIP-seq in Wt E11.5 face and liver, while the ones labeled with the " * " are obtained from the publically available datasets SA1 and SMC1 ChIP-seq (cohesin subunits) done in mouse embryonic fibroblasts (MEF) (Cuadrado et al., 2015).

5.6 Imprinting in *c-Myc locus*

5.6.1 Discovery and Characterization, Performed by Veli Uslu

Veli Uslu observed that inside *c-Myc locus*, three close GROMIT insertions (including 179039 (17a)) had reporter expression only if inherited from the father (figure 22). Therefore, these insertions were maternally imprinted.

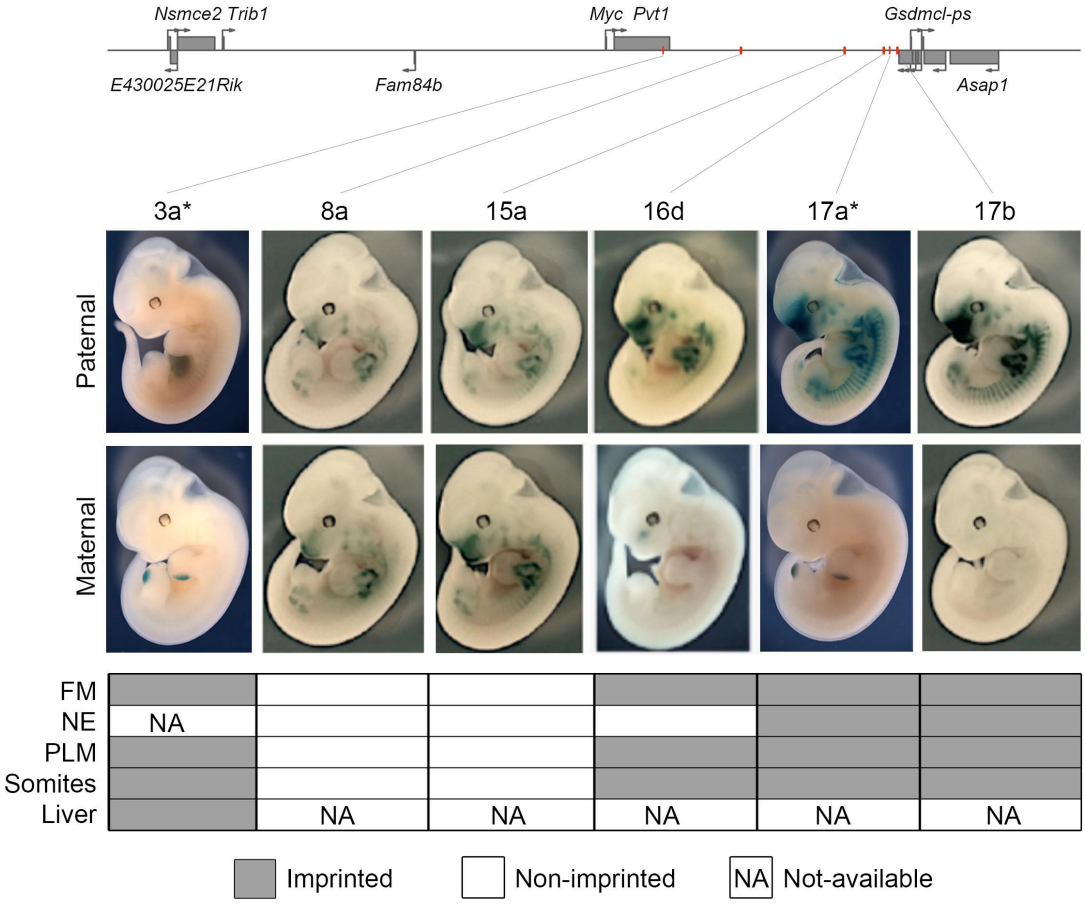


Figure 22 | Imprinting in *c-Myc locus*. Data shown thanks to kind concession of Veli Uslu. LacZ reporter staining on E11.5 mouse embryos carrying paternally or maternally inherited GROMIT insertions The picture shows that the LacZ reporter gene from the insertion 179039 (17a) (as well as two other close ones) is silent if inherited from the mother, but is expressed if it is inherited from the father, suggesting a maternal imprinting effect (the imprinting at the insertion 194578 (3a) was determined later on by myself). The insertions 194578 (3a) and 179039 (17a) are marked with " * " to indicate that the pictures were taken in a different batch from the others. For these two lines, the embryos carrying GROMIT insertions in the maternal allele were obtained from a cross where the mother was positive for the indicated GROMIT insertion (in *c-Myc locus*), and the father had a GROMIT insertion located in an unrelated *locus.* This second insertion is not under imprinting effect and gives LacZ staining in tissues (terminal limb) not overlapping with any stained tissue for any insertion located inside *c-Myc locus*; this insertion was used as an internal positive control.

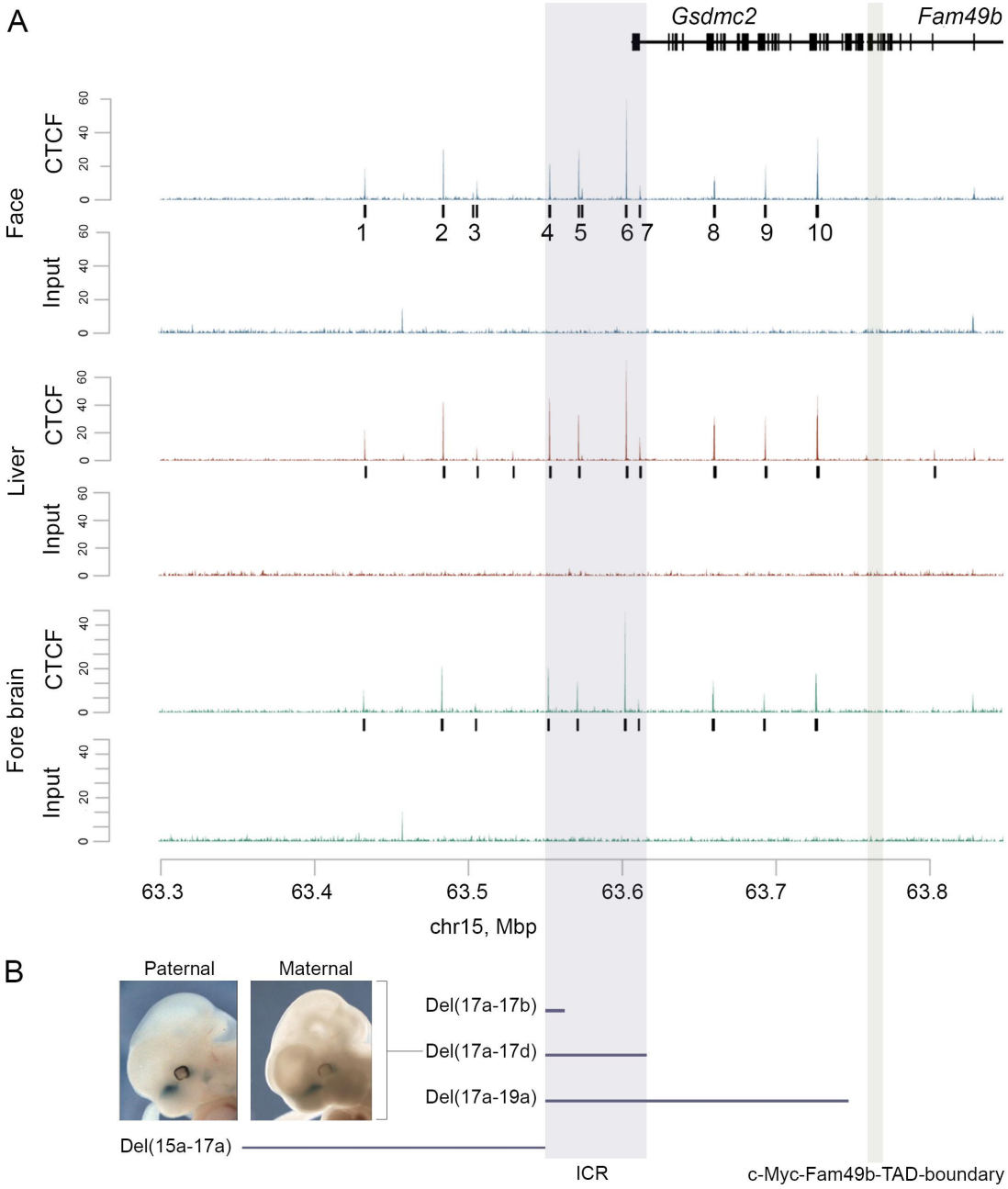


Figure 23 | Identification of the imprinting control region in *c-Myc locus.* A) CTCF ChIP-seq in mouse E11.5 face, fore brain and liver showing the 10 CTCF binding sites clustered at the telomeric end of *c-Myc locus* (their numeration is shown below the face CTCF dataset). Peak-calling and input controls tracks are shown underneath each ChIP-seq track. B) Material shown thanks to Veli Uslu's concession. In this experiment, four chromosomal deletions, covering the complete CTCF cluster, were created in order to identify the imprinting control region (ICR) of the *locus*. Upon Del(17a–17d), the LacZ staining reappeared on the maternal allele (on nasal epithelia and branchial arches, E11.5 embryos), being comparable to the reporter staining on the paternal allele, indicating that the imprinting effect was lost.

The dissection of the region by small chromosomal deletions showed that the ICR lies telomeric to the insertion 179039 (17a) (Del(17a–17d), figure 23).

The analysis of other GROMIT insertions (8a, 14c and 15a) revealed that already 200Kbp centromeric to the affected region, the reporter expression from the maternal allele was restored, being indistinguishable from the paternal allele (figure 22, the imprinting on the insertion 194578 (3a) was determined later by myself, and is treated in the discussion). For this reason, the imprinting effect was reported to be local, limited to 50Kbp around 179039 (17a). Remarkably, the reporter expression observed from the maternal allele for the insertions located outside the imprinted region, rules out the possibility that the imprinting inactivates the facial enhancers.

Bisulfite assays on both face and heart tissues (stage E11.5) from 179039 (17a) GROMIT positive embryos, showed that the promoter of the LacZ had DNA methylation only if it was maternally inherited (figure 24).

In order to test whether the imprinting was affecting the activity of the facial enhancers on *c-Myc*, a mouse line carrying their deletion (Del(8–17)) was used. E11.5 embryonic facial tissues from animals carrying the deletion on the maternal or paternal chromosome were compared. This experiment showed that the paternal and maternal enhancers contribute to 50% and 30% of the total *c-Myc* expression respectively, suggesting that the imprinting was responsible for a mild reduction of the enhancers' activity on *c-Myc* promoter in the maternal allele.

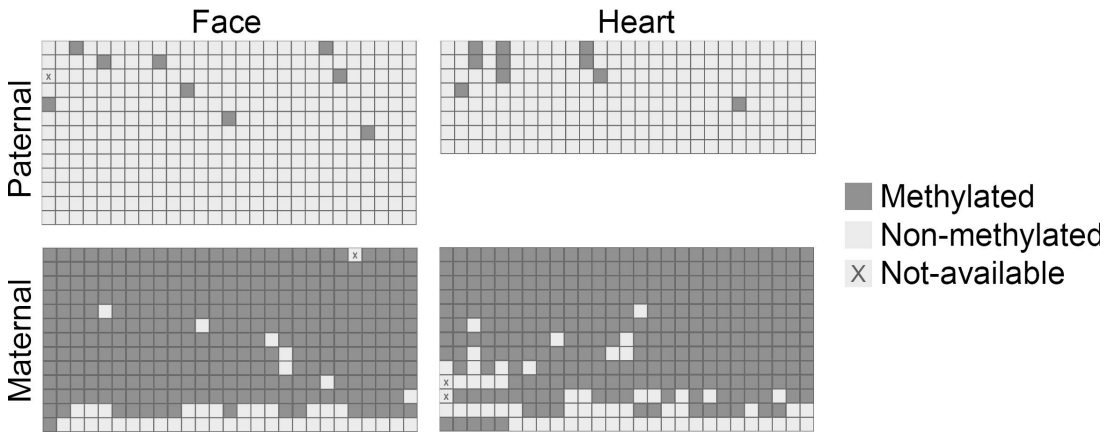


Figure 24 | Methylation status of the 179039 (17a) LacZ promoter. Data shown thanks to kind concession of Veli Uslu representing bisulfite sequencing assays on E11.5 facial mesenchyme and heart from mouse embryos carrying the 179039 (17a) insertions on the paternal or maternal chromosome. Each panel shows different CpG di-nucleotides present in the LacZ promoter (columns), and different sequencing clones (rows).

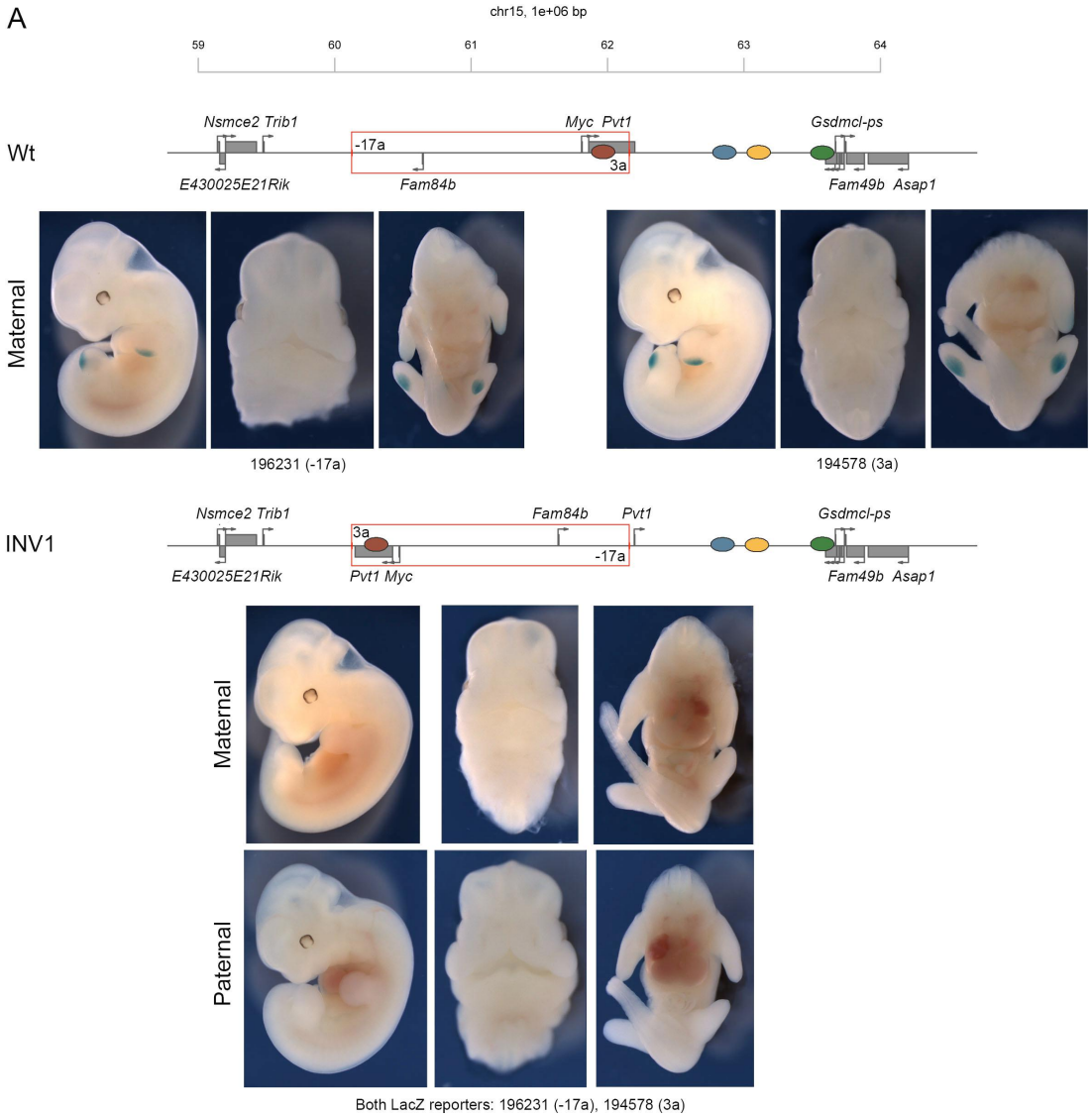
5.6.2 Effects of Chromosomal Inversion on the Imprinted Regions

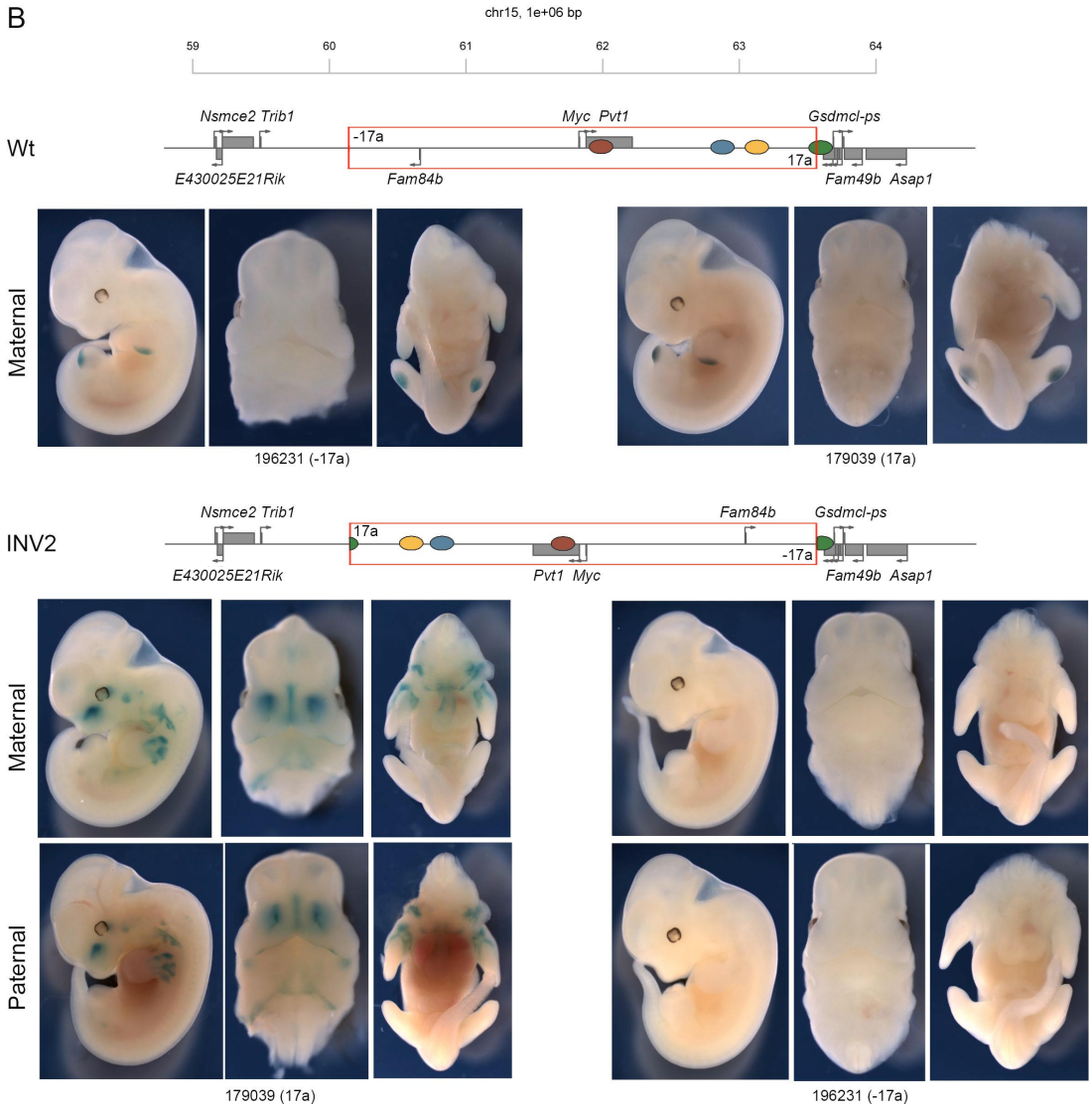
The inversion mouse lines were useful to confirm the results obtained by Veli Uslu and also to test whether the extent of the imprinted region may be affected by the genetic reshuffling. We addressed these points by analyzing the LacZ reporter staining at the inversions breakpoints.

In INV1, the lack of reporter staining on the paternal allele, from both breakpoints, does not allow us to infer whether or not the imprinting in the maternal allele is still reaching the telomeric breakpoint (insertion 196231 (-17a) upon INV1) (figure 25A).

In INV2, the insertion 179039 (17a) in the maternal allele gave a LacZ staining pattern comparable to the one observed on the paternal allele (figure 25B). The lack of LacZ staining from the Wt 179039 (17a) on the maternal allele shows that at this site the imprinting is lost upon INV2. Regarding the insertion 196231 (-17a), as it is silenced on both paternal and maternal alleles in both Wt and INV2, we cannot tell if it is still under imprinting effect upon inversion (figure 25B).

In INV3, GROMIT insertions located at both inversion breakpoints gave LacZ staining from the paternal allele (although with different patterns, figure 25C). In the maternal allele, the same locations are completely silenced, as in the Wt scenario, showing that the imprinting still reaches the insertions 179039 (17a) and 194578 (3a) (figure 25C). This experiment supports the idea that the function of the ICR is unaffected by the genetic reshuffling introduced by INV3.


A
chr15, 1e+06 bp
59
60
61
62
63
64
Wt
Nsmce2 Trib1
-17a
Myc Pvt1
3a
Gsdmcl-ps
E430025E21Rik
Fam84b
Fam49b Asap1
Maternal
196231 (-17a)
194578 (3a)
INV1
Nsmce2 Trib1
3a
Fam84b
Pvt1
Gsdmcl-ps
-17a
E430025E21Rik
Pvt1 Myc
Fam49b Asap1
Maternal
Paternal
Both LacZ reporters: 196231 (-17a), 194578 (3a)


B
chr15, 1e+06 bp
59
60
61
62
63
64
Wt
Nsmce2 Trib1
E430025E21Rik
-17a
Fam84b
Myc Pvt1
17a
Gsdmcl-ps
Fam49b Asap1
Maternal
196231 (-17a)
179039 (17a)
INV2
Nsmce2 Trib1
E430025E21Rik
17a
Pvt1 Myc
Fam84b
-17a
Gsdmcl-ps
Fam49b Asap1
Maternal
Paternal
179039 (17a)
196231 (-17a)

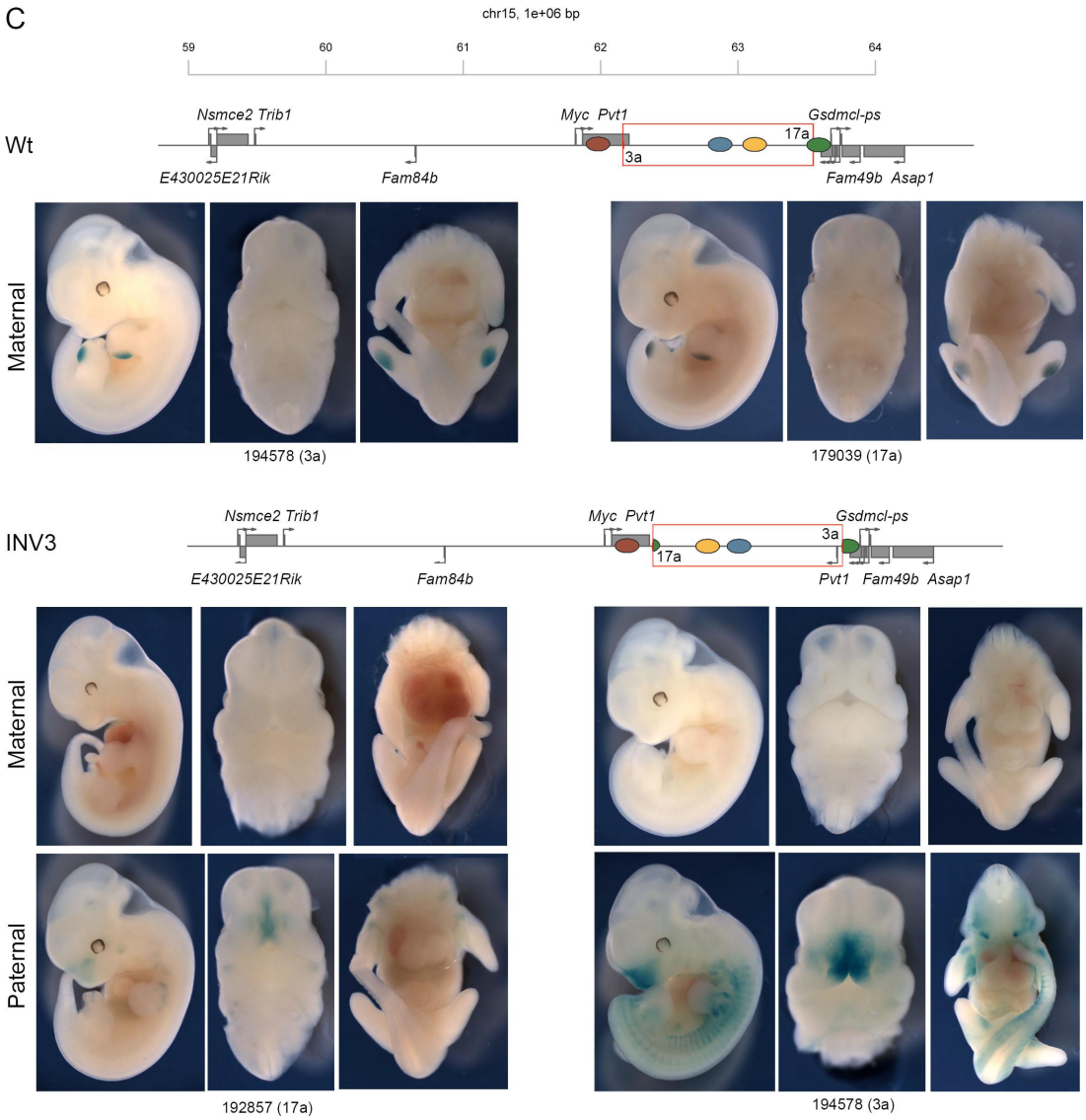


Figure 25 | Effect of the chromosomal inversions on the imprinted regions. LacZ staining of E11.5 mouse embryos carrying GROMIT on the maternal and (only for the inversion lines) on the paternal allele. The pictures refer to insertions located at the inversions breakpoints. In order to prove that the lack of LacZ staining in the Wt controls was not due to a technical failure of the staining procedure, the Wt controls embryos shown here were obtained from a cross where the paternal genome contained a GROMIT insertion located in an unrelated *locus,* giving a staining pattern (on the terminal limbs) not overlapping with the analyzed *c-Myc locus* insertions. Wt controls are shown at the top of each panel. Engineered lines are shown at the bottom. Tissue-specific enhancers are represented on the gene models: blue, MFM; yellow, NE; red, liver. The telomeric end of the *locus* is depicted in green. A) In INV1 the imprinting at the telomeric breakpoint 194578 (3a) cannot be verified because of the lack of reporter staining also on the paternal allele. B) In INV2 the presence of LacZ staining in several tissues from the insertion 179039 (17a) (comparable to the LacZ staining from the paternal allele) shows that the imprinting is lost upon INV2. C) In INV3 both inversion breakpoints, silent in the Wt, remain silent also upon inversion, showing that the imprinting effect is still present at these regions.

6 Discussion

6.1 Roles of Chromatin in *c-Myc locus*

The identification of different regulatory domains in *c-Myc locus* and adjacent *loci,* as well as the observation that even inside the same regulatory domain there can be significant differences in transcriptional capability of the individual regions analyzed, led us to question whether the chromatin composition of the *locus* has some correlation with this phenomenon. Moreover, the characterization of different regions containing tissue-specific enhancers, sized in the range of hundreds of Kbp, offered us an important guideline to restrict their position even further, by investigating the enhancer-associated chromatin. For these reasons we performed different ChIP-seq analysis in Wt animals, focusing on our model tissues, *i.e.* face and liver, at the embryonic stage E11.5.

ChIP-seq for histone marks enriched at enhancer sites confirmed previous information regarding the position of facial and liver enhancers acting on *c-Myc* during mouse embryonic development, and helped us to refine their possible genomic locations. Concerning the H3K27ac ChIP-seq on the facial tissue, we saw three peaks located inside the ~ 600Kbp critical region previously shown to contain important MFM enhancer(s) (figure 10C) (Uslu et al., 2014). One of these peaks in particular, overlapped with the mouse ortholog of the sequence hs1877, shown by transgenic assays to promote reporter expression in the MFM and support the idea that the small regions identified by this approach may contain facial enhancers (figure10D and F) (Attanasio et al., 2013). In the liver, the analysis is complicated by the fact that inside the candidate region for containing the liver enhancers, we saw 21 distinct H3K27ac peaks. It is therefore not possible to know at this stage (and without further experiments) which region(s) could contain the liver enhancer(s).

We also assessed the distribution of two kinds of repressive chromatin, with the aim of finding clues possibly explaining the regulatory landscape of the *locus.* We investigated two well-known chromatin types, namely the heterochromatin, positive for the H3K9me3 histone mark, and the Polycomb chromatin, positive for the H3K27me3 histone mark. We found just sporadic representation of these histone modifications in *c-Myc locus.* The only enriched region for H3K9me3 shared

between the face and liver lay close to the telomeric end of the *locus*. Importantly, the closest GROMIT insertion to this peak (insertion 193315 (16a)), found just 3.5Kb away, is not silenced. Similarly, other close insertions, including 179039 (17a) do show LacZ staining. Concerning the H3K27me3 mark, we believe that the enrichment at *Pvt1* promoter we observed in both face and liver, may be linked to the fact that this gene is expressed at an extremely low level in these two tissues.

Regarding the enhancer-associated mark H3K27ac on the face, the peak overlapping with the mouse ortholog of the sequence hs1877 (capable of driving reporter expression in the MFM, mentioned above) and included in the facial enhancer(s) critical region, occurs 30Kbp away from the GROMIT insertion 193970 (10a). This insertion, despite being the closest to this H3K27ac peak, is silent in all the tissues, and does not respond to any enhancer. On the contrary, insertions located much further away from the facial enhancer(s) critical region (the case of 179039 (17a) is a clear example) do show LacZ expression in the facial tissues.

According to these observations, our data seems to suggest that the chromatin composition of the *locus* cannot explain the regulatory landscape (and therefore the regulatory domains) observed from the GROMIT screening performed by Veli Uslu. We therefore tend to favor the idea that the *locus* topology and long-range interactions play a crucial role in *c-Myc locus* to connect regulatory elements to their target regions.

6.2 Importance of Chromosome Architecture

6.2.1 Roles of DNA Topology on *c-Myc* Transcriptional Regulation

In mammalian genomes, regulatory elements and target promoters often occur far away from each other, commonly at distances greater than 500Kbp (Martin et al., 2015). Sometimes these long-range interactions can even skip genes, indicating that the proximity is not the only player arranging promoter–regulatory element interactions (Marinić, Aktas, Ruf, & Spitz, 2013; Martin et al., 2015). A growing body of evidence suggest that metazoan genomes are condensed into the nucleus following a determined topology (J. R. Dixon et al., 2012; Rao et al., 2014). It has been shown in a few (although increasing number of) cases that the topology of the

loci has a functional relevance in gene expression, and if perturbed can lead to pathological phenotypes (Lupiáñez et al., 2015; Montavon et al., 2011; Spitz et al., 2005). In this project we addressed this question using chromosomal inversions in *c-Myc locus* to rearrange the relative distribution of enhancers, TAD-boundaries and promoters. We probed the resulting effects at the level of expression of *c-Myc* and neighboring genes, as well as by reporter assays.

With our genetic engineering setup we aimed to alter the genetic context between the tissue-specific enhancers and *c-Myc* promoter. In the case of INV1, we saw a significant reduction of *c-Myc* expression in the face and PFLM, as well as a reproducible drop of reporter expression at the inversion breakpoint 194578 (3a).

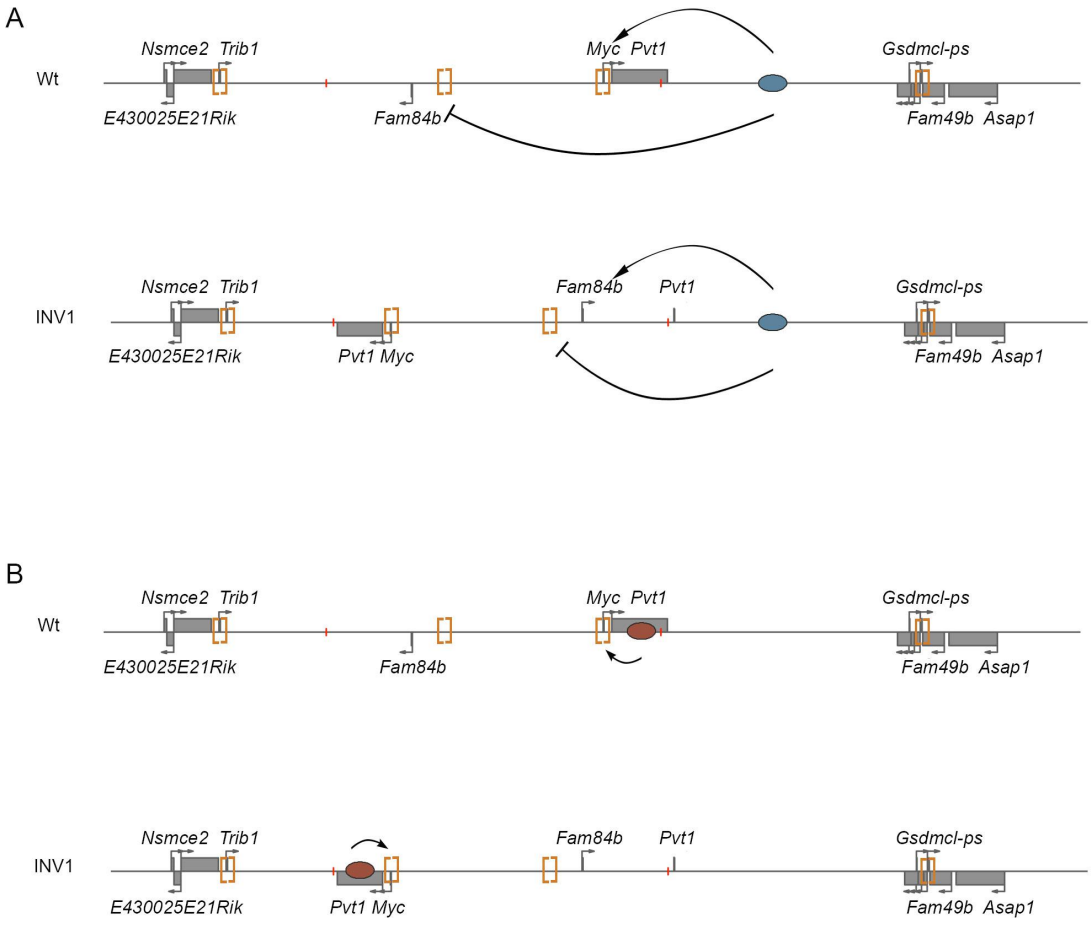


Figure 26 | Perturbations of long-range interactions in INV1. A) Model proposing how the facial enhancers (for the FM and NE; here shown only the FM one(s) as a blue ellipse) upon INV1 are redirected from *c-Myc* to *Fam84b*. TAD-boundaries are represented by sand color square-brackets (shown also on panel B). B) In the liver the activity of long-range enhancer(s) (probably located inside the *Pvt1* transcribed region) on *c-Myc* promoter, may be just mildly affected by the inversion.

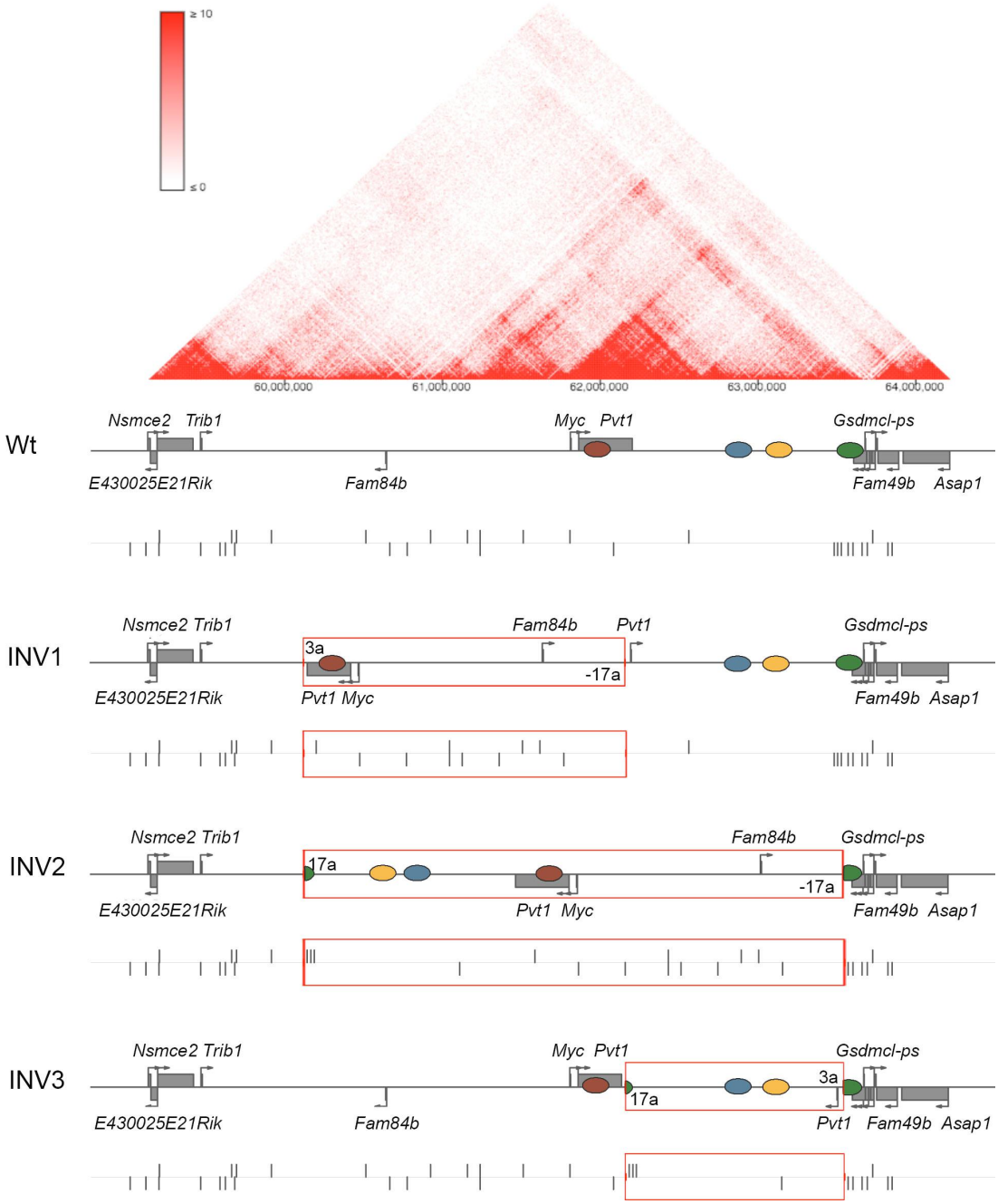


Figure 27 | CTCF motifs distribution in *c-Myc* and adjacent *loci*. The motif positions are obtained by BSgenome.Mmusculus.UCSC.mm9 R package (Bioconductor version: Release (2.12)), using the JASPAR position weighed matrices (publically available at: http://jaspar.genereg.net), using the default parameters (minimum score = 85% identity). The CTCF motifs distribution is shown for the Watson and Crick strands separately. The top of the figure shows the Hi-C interaction matrix of *c-Myc* and neighboring *loci* (Rao et al., 2014). Cis-acting DNA elements are shown on the gene models: blue, facial enhancer(s); yellow, nasal epithelia enhancer(s); red, liver enhancer(s); green, telomeric end of *c-Myc locus*.

Along with the downregulation of *c-Myc*, we observed a concomitant upregulation of *Fam84b* in the same tissues, compatible with a hijacking of the enhancers from *c-Myc* to *Fam84b* (figure 26A).

We believe that the marked downregulation of *c-Myc* observed in the face and PFLM could in principle affect also other tissues, and cause the embryonic lethality seen on INV1 homozygotes animals.

One possible explanation for the transcriptional changes observed in the face and PFLM could be that INV1 inverts the orientation of important CTCF binding sites, compatible with recent evidence obtained in different model *loci,* proposing an orientation dependent CTCF function (reviewed in (Zlotorynski, 2015). In the *Pcdh locus*, the HS51 enhancer lies between *Pcdha* and *Pcdhg* gene clusters, interacting only with *Pcdha* promoters in both mouse and human cell lines. Importantly, the HS51 enhancer contains a CTCF binding site. In this study the whole HS51 enhancer was inverted using the CRISPR-Cas9 technology. The inversion caused the HS51 enhancer to gain interactions with the *Pcdhg* cluster promoters and lose interactions with the *Pcdha* cluster promoters. Concomitantly, *Pcdhg* genes expression increased, while *Pcdha* genes expression decreased. Similar data obtained on the *β-globin locus* strengthen the hypothesis that CTCF binding sites orientation contributes to organize promoter–enhancer long-range interactions.

In *c-Myc locus,* the closest CTCF motifs to *c-Myc* and *Fam84b* promoters lie on opposite strands. INV1 inverts the orientation of these CTCF motifs, in reference to the CTCF motif located on the region containing the facial and limb enhancers (not included in the inversion) (figure 27). With INV1 arrangements, CTCF motif orientation does correlate with the redirection of the facial and liver enhancers from *c-Myc* to *Fam84b*. However, looking at the actual CTCF binding on these motifs in the Wt, out of the three of them, CTCF binds only the motif at *c-Myc* promoter. Other data (discussed further on) supports the idea that CTCF occupancy does not change upon inversions. Moreover, the part of the *locus* included in INV1, contains many other CTCF motifs other than the ones mentioned. The case described for the *Pcdh locus* instead was much simpler, as the inversion involved just one CTCF binding site. It is therefore not possible to extrapolate their conclusions to *c-Myc* case. We believe instead that the effect we observed on face and fore limbs was mostly due to the interposition of a TAD-boundary between *c-Myc* and the facial/limb enhancers, more than the increased distance separating them or the inverted orientation of the CTCF motifs.

Still with INV1, in the liver we saw a similar trend as in the face and PFLM, showing a reduction of *c-Myc* and *Pvt1* expression and an increase of *Fam84b* (as

well as *E4300021Rik, Nsmce2, Trib1, Fam49b* and *Asap1,* in all the cases being not statistically significant). However, in the liver, this effect was not as pronounced as in the face and PFLM, suggesting that the INV1 did not alter the relationship between *c-Myc* and the liver enhancer(s) as much as for the face and PFLM (figure 26B). In addition, INV1 caused the disappearance of the liver staining from the insertion 194578 (3a). Importantly, INV1 preserves the Wt disposition relative to each other of the insertion 194578 (3a), *c-Myc* promoter and the liver enhancer(s) (likely located inside *Pvt1* transcribed region), suggesting the possibility that the changed context in which these sequences operate may be responsible for the effects we observed at the level of both reporter expression and endogenous transcript expression. Veli Uslu in the past drew similar conclusions using a different genetic engineering model (figure 15). In this specific case, we speculate that the effects we saw have to do with the importance of the telomeric side of the *locus* to allow the liver enhancers to reach *c-Myc* promoter and the insertion 194578 (3a).

6.2.2 Influence of DNA Topology to Determine the Binding of Architectural Proteins to the Genome

According to the current models proposed in the literature, the architectural transcription factor CTCF and the Cohesin complex are involved in DNA-DNA looping and long-range interactions. Their function led us to question to what extent their binding to the genome depends on the DNA sequence itself or on the *locus* architecture.

We addressed this question by analyzing their binding inside *c-Myc locus* upon inversions. Despite the fact that with our experimental setup we could only probe a limited number of regions inside the *locus*, in INV1 we observed just minor changes in CTCF and RAD21 enrichments. As their binding did not seem to have rearranged together with the inversion, adapting to the new configuration, we tend to think that in this instance, CTCF and the Cohesin complex disposition was not significantly constrained by the architecture of the rest of the *locus* and perhaps more influenced by the DNA sequence.

This experiment also suggests that the changes in *c-Myc* expression that we saw may be more explainable by a rewiring of long-range interactions, occurring at the same CTCF and Cohesin binding sites than in the Wt, rather than by changes in

CTCF and Cohesin occupancy inside the *locus* (provided that architectural proteins do play an active role in promoting long-range interactions).

Similarly to INV1, in INV2 and INV3 the changes in *c-Myc* expression were not accompanied by changes of CTCF and RAD21 occupancy, suggesting that these conclusions may not represent a particular case limited to INV1, but could reflect a more general scenario, in which the binding to the DNA of these architectural proteins may be determined to a large extent by the DNA sequence itself, rather than by the surrounding genetic context or the *locus* architecture, and therefore may not be affected (to some extent) by the sequences reshuffling introduced by the inversions, or by other kinds of genetic engineering.

6.3 Roles of the Telomeric End of *c-Myc locus*

6.3.1 Long-Range Interactions

The first evidence showing that the telomeric end of *c-Myc locus* had some remarkable properties comes from the first GROMIT insertion mapped inside *c-Myc locus*, namely 179039 (17a), obtained by Sandra Ruf (Ruf et al., 2011). The analysis of the LacZ staining pattern given by this insertion revealed that *c-Myc* and the 17a region might respond to the same regulatory elements. At the time it was not clear how this was possible, as the two of them were separated by a considerable genetic distance, and in addition to this, the insertion 179039 (17a) was closer to other genes than it was to *c-Myc*.

Now, we know from ChIP-seq experiments that the telomeric end of *c-Myc locus* is enriched for H2AZ and H3K27ac in face and liver, but not in the FB (control tissue with basal *c-Myc* expression).

Unpublished data, obtained by Veli Uslu and Peter Krijger (from Wouter de Laat laboratory) using the 4C method on the embryonic face, showed that this region contacted *c-Myc* promoter as well as the TAD-boundary separating Fam48b-TAD from c-Myc-TAD. We do not have this kind of data for liver and FB as control, therefore we cannot state that the telomeric end of *c-Myc locus* engages the same contacts in tissues where *c-Myc* transcriptional boost relies on other long-range enhancers, or in tissues with basal *c-Myc* expression. However, some recently published Hi-C data showed similar findings to the ones described by Veli Uslu for the face, in other cell-types. This study performed with higher sequencing depth and

a few technical improvements of the standard protocol, managed to capture in human and mouse cells not only the previously described topological domains, but also individual loops (Rao et al., 2014). It was proposed that a large fraction of these loops occur between TAD edges, namely, the sequences flanking TAD-boundaries. In CH12 cells (mouse B-cell lymphoma) c-Myc-TAD also show this behavior, namely the 330Kbp at the telomeric end of the *locus* (corresponding to the telomeric end of c-Myc-TAD, mm9 coordinates: ~ chr15:63420000-63750000), extending up to the gene *Gsdmcl-ps*, engage extensive contacts with the centromeric end of c-Myc-TAD, as well as with *c-Myc* transcribed region, defining the centromeric edge of the telomeric c-Myc-subTAD (figure 8).

The ChIP-seq data on face, liver and FB (as well as publically available data (T. H. Kim et al., 2007)) confirms the presence of multiple CTCF binding sites at each of these TAD-edges. In particular, at the telomeric end of c-Myc-TAD, we saw ten CTCF peaks. In INV2 and INV3, the telomeric breakpoint lies inside this CTCF cluster, splitting it into two blocks and for this reason these lines were useful to ask important questions regarding the functions as a contact point of the telomeric end of the *locus*.

In both cases, we observed a reduction of reporter responsiveness from the insertion 17a, upon inversion (on FM, NE, somites, PLM, BA and RL), being more pronounced for INV2 (figure 17 and figure 19). These lines show that the centromeric (relative to the insertion 17a) segment (segment-C), moved with the inversion, is not affected by the enhancer(s) (for the MFM in particular) as much as it is in the Wt, likely because of the strengthening action of the other segment (~ 200Kbp telomeric to the insertion 17a, segment-T) not included in the inversion. Supporting this view, in INV3, the insertion 194578 (3a) acquires an expression pattern resembling the insertion 192857 (17a) in Wt configuration. Therefore the segment-T may have an important role to capture transcriptional input in all the characteristic tissues of the insertions 179039 and 192857 (17a).

In INV2 we saw a reduction of *c-Myc* expression in the face, while in INV3 this effect was reproducible, however not statistically significant (figure 18 and figure 20). Remarkably, INV2 preserves the *locus* structure (with exception of the segment-T), therefore, it is likely that the downregulation of *c-Myc*, may be due to an insufficient function of the telomeric end of the *locus* once it is depleted of the segment-T.

Also in this case, there is the possibility that changes in the orientation of CTCF binding sites are responsible for the effect we observed on *c-Myc* expression. Indeed, 5 out of 8 CTCF motifs are not included in INV2, and therefore their orientation, compared to the rest of *c-Myc locus*, is changed (figure 27). However, the

fact that the two parts of the telomeric end of *c-Myc locus* are separated from each other—being at the opposite sides of the *locus*—with two TAD-boundaries in between, supports the idea that this physical separation rather than the changes in CTCF motifs orientation are responsible for the insufficient activity of the telomeric end of the *locus*. A good control experiment, in order to rule out the possibility that the CTCF motif orientation is important for the topological organization of the *locus*, could be to perform a targeted inversion (using the CRISPR-Cas9 technology) of the segment-T.

Despite the importance of this segment to promote the communication between *c-Myc* promoter and the facial enhancers, in INV3 the relationship between the two is unaltered, yet we saw a mild (and non-significant) downregulation of *c-Myc* in the face. There is the possibility that this effect may be due to an improper function of the segment-C. However, this segment is brought closer to *c-Myc* promoter than it is in the Wt and moreover it is still present inside the same subTAD. Alternatively, it is possible that the mild effect INV3 has on *c-Myc* expression may be caused by the general reshuffling of the sequences inside the subTAD rather than by a single sequence in particular. This model does not fully agree with the current ones present in the literature, proposing a TAD-centric view of the long-range interactions, however, our data support the idea that even inside TADs there may be sub-topologies whose alteration (for example by genetic engineering), could provoke small although perceptible perturbations of gene expression.

6.3.2 Genomic Imprinting as a Way to Modulate Enhancer Functions

The telomeric end of *c-Myc locus* was shown by Veli Uslu to be imprinted in the maternal allele. Veli Uslu mapped the location of the ICR to be telomeric to the insertion 179039 (17a) (figure 23). INV2 was useful to confirm this observation. Upon INV2, the imprinting is lost on the breakpoint 179039 (17a). We think that in this case, the separation of this insertion from the ICR (with the presence of two TAD-boundaries in between) is likely what causes the loss of imprinting (figure 17 and figure 25B). Interestingly, the small deletions that Veli Uslu used to identify the ICR (Del(17a–17d) and Del(17a–19a)) had a very similar staining pattern to the INV2 breakpoint 179039 (17a).

Veli Uslu reported that the extent of the imprinting was limited to a short region around the 17a insertion site. However, in this project we determined that the insertion 194578 (3a), which gives LacZ expression in multiple tissues if inherited from the father, is silent if inherited from the mother (figure 22). Despite the fact that we lack bisulfite-sequencing data regarding the methylation status of the LacZ promoter of the insertion 194578 (3a) (a required experiment to undoubtedly prove imprinting effects), our data is compatible with the idea that the insertion 194578 (3a) is imprinted in the maternal allele. This suggests (similarly to the regulatory landscape) that the imprinting spreading throughout the *locus* may be discontinuous. Considering the case of the MFM, PLM and somites, the imprinting is present at the telomeric end of the *locus*, close to the ICR, it is absent from the insertion 8a to 15a, and it reappears more centromeric, on the insertion 194578 (3a). We speculate that probably the topology of the telomeric c-Myc-subTAD affects which regions contact the ICR, determining the extent of the imprinting. INV3 that includes most of the telomeric c-Myc-subTAD was useful to test whether the genetic reshuffling that it introduces might have some effect on the distribution of the imprinted regions. In INV3, the LacZ reporter at both breakpoints remains silenced upon inversion in the maternal allele, indicating that the imprinting is still acting at these sites. This data suggests (despite not having information regarding other positions, like 8a and 15a, inside the inverted region) that the genetic and topological changes did not alter the imprinting distribution.

One last remark, probably the most important, regards the physiological meaning of the imprinting on *c-Myc locus*. According to the models proposed in the literature, the DNA methylation, at least for the case of maternal imprinting, should occur at genes promoters, or close by. This, as consequence, leads to the complete silencing of some genes and expression of others, with an on/off effect. In *c-Myc locus* the ICR lies 1.7Mbp away from *c-Myc* promoter. The effect of the DNA methylation on the maternal allele does not cause the complete silencing of *c-Myc,* but rather has a milder effect, reducing its expression of ~ 20% (compared to the paternal allele). Veli Uslu observed that this difference could be attributed to a different use of the enhancers present in the telomeric c-Myc-subTAD.

In addition, the effect of the imprinting does not seem to affect all the tissues in the same manner, for example the GROMIT insertion 16d (figure 22) from the maternal allele, gives a mild, although clearly perceptible LacZ staining in the NE, although it is imprinted in MFM, PLM and somites. According to the current view, in most of the *loci,* the imprinting does not show tissue-variability.

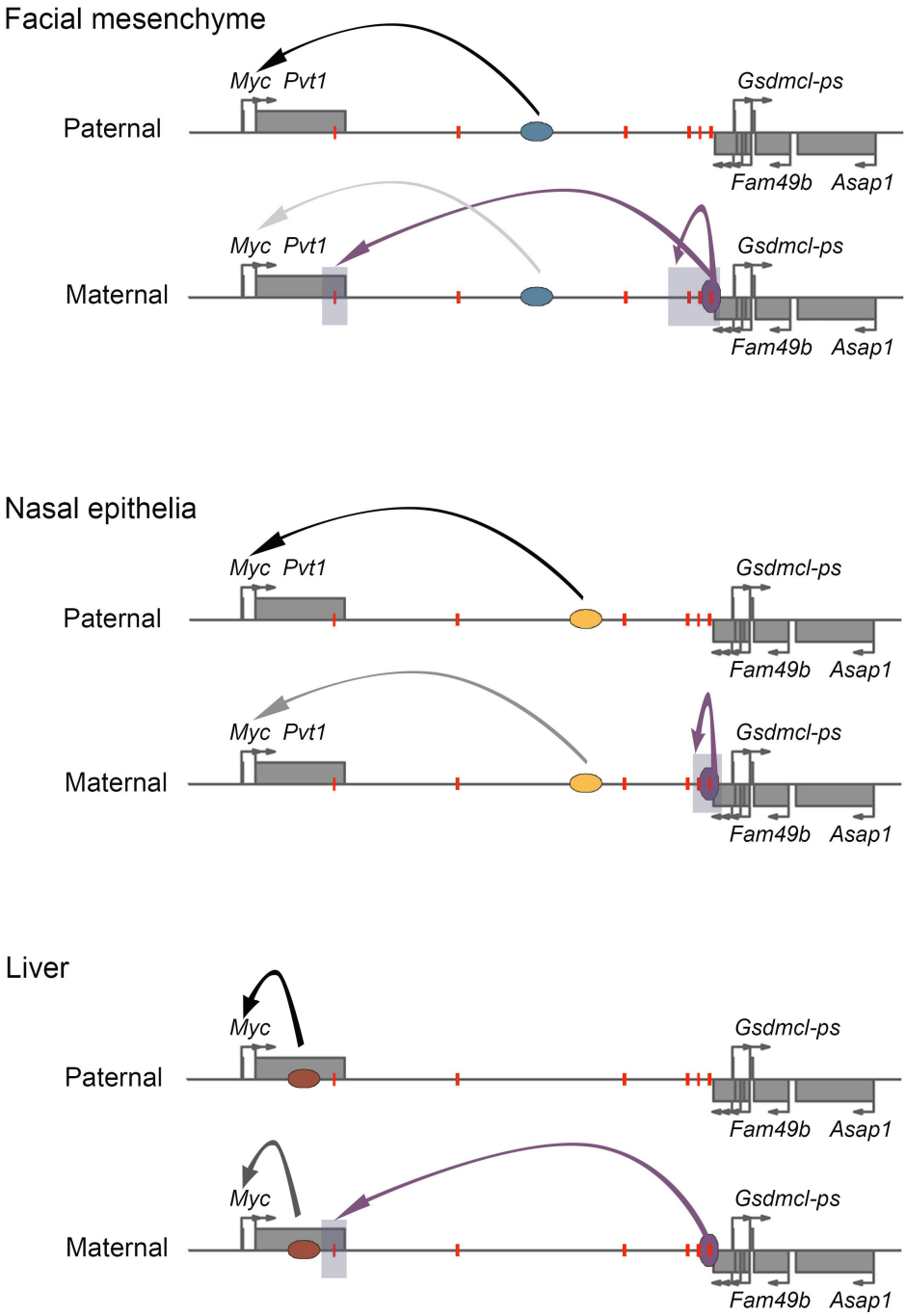


Figure 28 | Tissue-specific imprinted regions in *c-Myc locus*. Hypothetical model showing how the different extent of the imprinting—observed in different tissues—could affect the activity of the tissue-specific enhancers. The enhancers are represented by ellipses, blue for the facial mesenchyme, yellow for the nasal epithelia and red for the liver. The imprinting control region is depicted as a lilac ellipse. The imprinted regions are shown as lilac semi-transparent rectangles. The red lines on the gene model show the location of the GROMIT insertions on which the imprint has been tested.

On the other hand, there are more and more reported cases describing tissue-specific imprinting that could support our observations obtained in *c-Myc locus* (Babak et al., 2015; J. T. Lee & Bartolomei, 2013; Prickett & Oakey, 2012). In the NE (differently from MFM, PLM and somites) the imprinting is not seen at any position centromeric to 179039 (17a). Because of the fact that the insertion 194578 (3a) does not give any staining on the NE, regardless of the parental origin, we cannot say whether it is also imprinted on the NE, in addition to MFM, PLM and somites. Therefore we cannot tell whether in this tissue the imprinting is local, involving only the ICR, or it is discontinuous.

In conclusion, we observed several differences in *c-Myc locus*, compared to the general models of imprinting described in the literature. These differences include that *c-Myc* is not embedded in a gene-cluster, that the imprinting does not have an on/off effect on *c-Myc* expression, that its extension is discontinuous inside the *locus* and that likely there are differences in the imprinted regions comparing different tissues. Therefore, it is possible that the imprinting in *c-Myc locus* may not be established in order to attain mono-allelic expression of *c-Myc* but rather to modulate the activity of the tissue-specific enhancers in different ways for the paternal and maternal *loci* (figure 28).

Our data suggest that the telomeric end of *c-Myc locus*, and in particular the segment extending from the insertion 179039 (17a) to the gene *Gsdmcl-ps* may be important for the communication between *c-Myc* and its facial enhancers. As the ICR lies on the telomeric end of the *locus*, we speculate that the imprinting may be a way to affect the function of the telomeric end of the *locus* in the maternal allele, which as a consequence, would slightly reduce the effect of the facial enhancers (and possibly other ones) on *c-Myc.* Supporting this view, in INV2 (with which we force a reduced function of the telomeric end of the *locus*), the displacement of the segment telomeric to the insertion 179039 (17a) from the rest of the *locus*, resulted in a reduced *c-Myc* expression in the face. We think it is possible that the imprinting at the telomeric end of the *locus* may cause a relaxation of intra-TAD interactions in the maternal *locus*, causing the reduced effect of the facial enhancers on *c-Myc*. Regarding the mechanisms that could explain this relaxation of interactions, we know that the ICR is located somewhere inside a ~ 50Kbp critical region telomeric to the insertion 179039 (17a), and that most of the CpG di-nucleotides on the LacZ promoter in this insertion are methylated in the maternal chromosome. As the ICR critical region contains four tissue-invariant CTCF binding sites (number 4, 5, 6 and 7, figure 23), out of which, only one (number 6, figure 23) possesses the canonical CTCF motif and a CpG dinucleotide. It is possible that if it is methylated in the

maternal chromosome, there may be (similarly to the ICR of the *Igf2/H19 locus*) a drop in CTCF binding. If this is the case—and whether it causes a reduction (or just a rearrangement) of interactions inside the TAD—must be determined experimentally. In order to test CTCF binding on maternal and paternal chromosomes, the deletion line Del(17a–17d) used by Veli Uslu to identify the ICR critical region can be used. Males and females Del(17a–17d) can be crossed against Wt and the allele specific CTCF binding can be tested by ChIP-qPCR.

7 References

Akhtar, M. M., Scala, G., Cocozza, S., Miele, G., & Monticelli, A. (2013). CpG islands under selective pressure are enriched with H3K4me3, H3K27ac and H3K36me3 histone modifications. *BMC Evolutionary Biology*, *13*, 145. http://doi.org/10.1186/1471-2148-13-145

Akhtar, W., de Jong, J., Pindyurin, A. V, Pagie, L., Meuleman, W., de Ridder, J., … van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*, *154*(4), 914–27. http://doi.org/10.1016/j.cell.2013.07.018

Albert, T., Wells, J., Funk, J. O., Pullner, A., Raschke, E. E., Stelzer, G., … Eick, D. (2001). The chromatin structure of the dual c-myc promoter P1/P2 is regulated by separate elements. *The Journal of Biological Chemistry*, *276*(23), 20482–90. http://doi.org/10.1074/jbc.M100265200

Alexander, R. D., Innocente, S. A., Barrass, J. D., & Beggs, J. D. (2010). Splicing-dependent RNA polymerase pausing in yeast. *Molecular Cell*, *40*(4), 582–93. http://doi.org/10.1016/j.molcel.2010.11.005

Ali, S. A., Dobson, J. R., Lian, J. B., Stein, J. L., van Wijnen, A. J., Zaidi, S. K., & Stein, G. S. (2012). A RUNX2-HDAC1 co-repressor complex regulates rRNA gene expression by modulating UBF acetylation. *Journal of Cell Science*, *125*(Pt 11), 2732–9. http://doi.org/10.1242/jcs.100909

Ambros, P. F., & Karlic, H. I. (1987). Chromosomal insertion of human papillomavirus 18 sequences in HeLa cells detected by nonisotopic in situ hybridization and reflection contrast microscopy. *Human Genetics*, *77*(3), 251–254. http://doi.org/10.1007/BF00284479

Amendola, M., & van Steensel, B. (2015). Nuclear lamins are not required for lamina-associated domain organization in mouse embryonic stem cells. *EMBO Reports*, *16*(5), 610–617. http://doi.org/10.15252/embr.201439789

Amundadottir, L. T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B. A., … Stefansson, K. (2006). A common variant associated with prostate cancer in European and African populations. *Nature Genetics*, *38*(6), 652–8. http://doi.org/10.1038/ng1808

Anderson, D. M., Beres, B. J., Wilson-Rawls, J., & Rawls, A. (2009). The homeobox gene Mohawk represses transcription by recruiting the sin3A/HDAC co-repressor complex. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *238*(3), 572–80. http://doi.org/10.1002/dvdy.21873

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M.,

… Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–61. http://doi.org/10.1038/nature12787

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)*, *339*(6123), 1074–7. http://doi.org/10.1126/science.1232542

Arvey, A., Agius, P., Noble, W. S., & Leslie, C. (2012). Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Research*, *22*(9), 1723–34. http://doi.org/10.1101/gr.127712.111

Askew, D. S., Ashmun, R. A., Simmons, B. C., & Cleveland, J. L. (1991). Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. *Oncogene*, *6*(10), 1915–1922.

Atchley, W. R., & Fitch, W. M. (1995). Myc and Max: molecular evolution of a family of proto-oncogene products and their dimerization partner. *Proceedings of the National Academy of Sciences*, *92*(22), 10217–10221. http://doi.org/10.1073/pnas.92.22.10217

Attanasio, C., Nord, A. S., Zhu, Y., Blow, M. J., Li, Z., Denise, K., … Visel, A. (2014). Fine Tuning of Craniofacial Morphology by Distant-Acting Developmental Enhancers. *NIH Public Access*, *342*(6157), 1–20. http://doi.org/10.1126/science.1241006.Fine

Attanasio, C., Nord, A. S., Zhu, Y., Blow, M. J., Li, Z., Liberton, D. K., … Visel, A. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. *Science (New York, N.Y.)*, *342*(6157), 1241006. http://doi.org/10.1126/science.1241006

Babak, T., DeVeale, B., Tsang, E. K., Zhou, Y., Li, X., Smith, K. S., … Fraser, H. B. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature Genetics*, *47*(5), 544–9. http://doi.org/10.1038/ng.3274

Ball, S. T., Kelly, M. L., Robson, J. E., Turner, M. D., Harrison, J., Jones, L., … Peters, J. (2013). Gene Dosage Effects at the Imprinted Gnas Cluster. *PloS One*, *8*(6), e65639. http://doi.org/10.1371/journal.pone.0065639

Banerji, J., Olson, L., & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, *33*(3), 729–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6409418

Barutcu, A. R., Fritz, A. J., Zaidi, S. K., vanWijnen, A. J., Lian, J. B., Stein, J. L., … Stein, G. S. (2015). C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. *Journal of Cellular Physiology*. http://doi.org/10.1002/jcp.25062

Baudino, T. A., McKay, C., Pendeville-Samain, H., Nilsson, J. A., Maclean, K. H., White, E. L., … Cleveland, J. L. (2002). c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes and*

Development, *16*(19), 2530–2543. http://doi.org/10.1101/gad.1024602
Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–5. http://doi.org/10.1038/35013100
Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, *98*(3), 387–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10458613
Bian, Q., Khanna, N., Alvikas, J., & Belmont, A. S. (2013). β-Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *The Journal of Cell Biology*, *203*(5), 767–83. http://doi.org/10.1083/jcb.201305027
Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.-Y., Lubkowska, L., Kashlev, M., & Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. *Cell*, *151*(4), 738–49. http://doi.org/10.1016/j.cell.2012.10.009
Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., … Mangold, E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nature Genetics*, *41*(4), 473–7. http://doi.org/10.1038/ng.333
Bishop, J. M. (1982). Retroviruses and cancer genes. *Advances in Cancer Research*, *37*, 1–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6763842
Blackwood, E. M., & Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science (New York, N.Y.)*, *251*(4998), 1211–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2006410
Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., … Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology*, *3*(5), e157. http://doi.org/10.1371/journal.pbio.0030157
Bothma, J. P., Magliocco, J., & Levine, M. (2011). The snail repressor inhibits release, not elongation, of paused Pol II in the Drosophila embryo. *Current Biology : CB*, *21*(18), 1571–7. http://doi.org/10.1016/j.cub.2011.08.019
Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., … Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, *132*(2), 311–22. http://doi.org/10.1016/j.cell.2007.12.014
Brooks, T. A., & Hurley, L. H. (2009). The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nature Reviews. Cancer*, *9*(12), 849–61. http://doi.org/10.1038/nrc2733
Brutzer, H., Luzzietti, N., Klaue, D., & Seidel, R. (2010). Energetics at the DNA supercoiling transition. *Biophysical Journal*, *98*(7), 1267–76. http://doi.org/10.1016/j.bpj.2009.12.4292
Bulger, M., & Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes & Development*, *13*(19), 2465–2477.

http://doi.org/10.1101/gad.13.19.2465
Butler, J. E., & Kadonaga, J. T. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & Development*, *15*(19), 2515–9. http://doi.org/10.1101/gad.924301
Calhoun, V. C., Stathopoulos, A., & Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9243–7. http://doi.org/10.1073/pnas.142291299
Camozzi, D., Capanni, C., Cenni, V., Mattioli, E., Columbaro, M., Squarzoni, S., & Lattanzi, G. (2014). Diverse lamin-dependent mechanisms interact to control chromatin dynamics. Focus on laminopathies. *Nucleus (Austin, Tex.)*, *5*(5), 427–40. http://doi.org/10.4161/nucl.36289
Campbell, K. J., & White, R. J. (2014). MYC regulation of cell growth through control of transcription by RNA polymerases I and III. *Cold Spring Harbor Perspectives in Medicine*, *4*(5). http://doi.org/10.1101/cshperspect.a018408
Campos, E. I., Stafford, J. M., & Reinberg, D. (2014). Epigenetic inheritance: histone bookmarks across generations. *Trends in Cell Biology*, *24*(11), 664–74. http://doi.org/10.1016/j.tcb.2014.08.004
Cannon, B., & Nedergaard, J. (2004). Brown adipose tissue: function and physiological significance. *Physiological Reviews*, *84*(1), 277–359. http://doi.org/10.1152/physrev.00015.2003
Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., … Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, *38*(6), 626–35. http://doi.org/10.1038/ng1789
Charalambous, M., da Rocha, S. T., & Ferguson-Smith, A. C. (2007). Genomic imprinting, growth control and the allocation of nutritional resources: consequences for postnatal life. *Current Opinion in Endocrinology, Diabetes, and Obesity*, *14*(1), 3–12. http://doi.org/10.1097/MED.0b013e328013daa2
Chen, H., Tian, Y., Shu, W., Bo, X., & Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PloS One*, *7*(7), e41374. http://doi.org/10.1371/journal.pone.0041374
Chen, M., Berger, A., Kablan, A., Zhang, J., Gavrilova, O., & Weinstein, L. S. (2012). Gsα deficiency in the paraventricular nucleus of the hypothalamus partially contributes to obesity associated with Gsα mutations. *Endocrinology*, *153*(9), 4256–65. http://doi.org/10.1210/en.2012-1113
Chen, M., Gavrilova, O., Liu, J., Xie, T., Deng, C., Nguyen, A. T., … Weinstein, L. S. (2005). Alternative Gnas gene products have opposite effects on glucose and lipid metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7386–91. http://doi.org/10.1073/pnas.0408268102
Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., … Price, D.

H. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. *Molecular Cell*, *45*(1), 38–50. http://doi.org/10.1016/j.molcel.2011.10.022

Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P., & Schedl, P. (2014). Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *36*(2), 163–72. http://doi.org/10.1002/bies.201300125

Clark, S. S. (1997). Recurring proviral integration suggests a role for proto-oncogene activation in thymomas induced with Mo-MuLV-rescued BCR/ABL virus. *Leukemia*, *11*(7), 1026–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9204986

Conrad, S. E., & Botchan, M. R. (1982). Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. *Molecular and Cellular Biology*, *2*(8), 949–65. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=369883&tool=pmcentrez&rendertype=abstract

Constância, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., … Reik, W. (2002). Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature*, *417*(6892), 945–8. http://doi.org/10.1038/nature00819

Conte, C., Dastugue, B., & Vaury, C. (2002). Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. *Molecular and Cellular Biology*, *22*(6), 1767–77. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135603&tool=pmcentrez&rendertype=abstract

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014a). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–20. http://doi.org/10.1038/ng.3142

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014b). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–1320. http://doi.org/10.1038/ng.3142

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, *322*(5909), 1845–8. http://doi.org/10.1126/science.1162228

Cowell, I. G., Aucott, R., Mahadevaiah, S. K., Burgoyne, P. S., Huskisson, N., Bongiorni, S., … Singh, P. B. (2002). Heterochromatin, HP1 and methylation at lysine 9 of histone H3 in animals. *Chromosoma*, *111*(1), 22–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12068920

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010a). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National*

Academy of Sciences of the United States of America, *107*(50), 21931–21936. http://doi.org/10.1073/pnas.1016071107

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010b). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–6. http://doi.org/10.1073/pnas.1016071107

Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., & Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of Cell Biology*, *145*(6), 1119–31. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2133153&tool=pmcentrez&rendertype=abstract

Cuadrado, A., Remeseiro, S., Graña, O., Pisano, D. G., & Losada, A. (2015). The contribution of cohesin-SA1 to gene expression and chromatin architecture in two murine tissues. *Nucleic Acids Research*, *43*(6), 3056–67. http://doi.org/10.1093/nar/gkv144

Cukier, C. D., Hollingworth, D., Martin, S. R., Kelly, G., Díaz-Moreno, I., & Ramos, A. (2010). Molecular basis of FIR-mediated c-myc transcriptional control. *Nature Structural & Molecular Biology*, *17*(9), 1058–64. http://doi.org/10.1038/nsmb.1883

Curtin, P. T., Liu, D. P., Liu, W., Chang, J. C., & Kan, Y. W. (1989). Human beta-globin gene expression in transgenic mice is enhanced by a distant DNase I hypersensitive site. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(18), 7082–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=297998&tool=pmcentrez&rendertype=abstract

Dai, M.-S., Sun, X.-X., & Lu, H. (2010). Ribosomal protein L11 associates with c-Myc at 5 S rRNA and tRNA genes and regulates their expression. *The Journal of Biological Chemistry*, *285*(17), 12587–94. http://doi.org/10.1074/jbc.M109.056259

Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., … Kraus, W. L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Molecular Cell*, *50*(2), 212–22. http://doi.org/10.1016/j.molcel.2013.02.015

Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R., & Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes & Development*, *7*(4), 671–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8458579

de Villiers, J., Olson, L., Tyndall, C., & Schaffner, W. (1982). Transcriptional "enhancers" from SV40 and polyoma virus show a cell type preference. *Nucleic Acids Research*, *10*(24), 7965–76. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=327063&tool=pmcentrez&rendertype=abstract

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–22. http://doi.org/10.1101/gad.2037511

Dekker, J., & Heard, E. (2015). Structural and Functional Diversity of Topologically Associating Domains. *FEBS Letters*. http://doi.org/10.1016/j.febslet.2015.08.044

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., & West, A. G. (2010). VEZF1 elements mediate protection from DNA methylation. *PLoS Genetics*, *6*(1), e1000804. http://doi.org/10.1371/journal.pgen.1000804

Dijkwel, P. A., & Hamlin, J. L. (1988). Matrix attachment regions are positioned near replication initiation sites, genes, and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. *Molecular and Cellular Biology*, *8*(12), 5398–409. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=365642&tool=pmcentrez&rendertype=abstract

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–80. http://doi.org/10.1038/nature11082

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nature Reviews. Genetics*, *12*(3), 167–78. http://doi.org/10.1038/nrg2933

Dornan, A. J., Gailey, D. A., & Goodwin, S. F. (2005). GAL4 enhancer trap targeting of the Drosophila sex determination gene fruitless. *Genesis (New York, N.Y. : 2000)*, *42*(4), 236–46. http://doi.org/10.1002/gene.20143

Doyle, H. J., Kraut, R., & Levine, M. (1989). Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. *Genes & Development*, *3*(10), 1518–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2612903

Drabent, B., Saftig, P., Bode, C., & Doenecke, D. (2000). Spermatogenesis proceeds normally in mice without linker histone H1t. *Histochemistry and Cell Biology*, *113*(6), 433–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10933220

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D. P., Thompson, D., Ballinger, D. G., … Ponder, B. A. J. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, *447*(7148), 1087–93. http://doi.org/10.1038/nature05887

Edelmann, P., Bornfleth, H., Zink, D., Cremer, T., & Cremer, C. (2001). Morphology and dynamics of chromosome territories in living cells. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1551*(1), M29–M39. http://doi.org/10.1016/S0304-419X(01)00023-3

Engineer, C. B., Fitzsimmons, K. C., Schmuke, J. J., Dotson, S. B., & Kranz, R. G. (2005). Development and evaluation of a Gal4-mediated LUC/GFP/GUS enhancer trap system in Arabidopsis. *BMC Plant Biology*, *5*, 9.

http://doi.org/10.1186/1471-2229-5-9
Ernst, J., & Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology*, *28*(8), 817–25. http://doi.org/10.1038/nbt.1662
Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., … Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43–49. http://doi.org/10.1038/nature09906
Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., … Bickmore, W. A. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Molecular Cell*, *38*(3), 452–64. http://doi.org/10.1016/j.molcel.2010.02.032
Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., … Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Molecular Cell*, *31*(3), 337–46. http://doi.org/10.1016/j.molcel.2008.06.021
Euskirchen, G. M., Auerbach, R. K., Davidov, E., Gianoulis, T. A., Zhong, G., Rozowsky, J., … Snyder, M. (2011). Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genetics*, *7*(3), e1002008. http://doi.org/10.1371/journal.pgen.1002008
Facchini, L. M., & Penn, L. Z. (1998). The molecular role of Myc in growth and transformation: recent discoveries lead to new insights. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *12*(9), 633–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9619443
Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., … van Steensel, B. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, *143*(2), 212–224. http://doi.org/10.1016/j.cell.2010.09.009
Finlan, L. E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., … Bickmore, W. A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics*, *4*(3), e1000039. http://doi.org/10.1371/journal.pgen.1000039
Forrester, W. C., van Genderen, C., Jenuwein, T., & Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. *Science (New York, N.Y.)*, *265*(5176), 1221–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8066460
Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., & Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology*, *15*(5), R69. http://doi.org/10.1186/gb-2014-15-5-r69
Fuchs, G., Voichek, Y., Rabani, M., Benjamin, S., Gilad, S., Amit, I., & Oren, M. (2015). Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with

4sUDRB-seq. *Nature Protocols*, *10*(4), 605–18. http://doi.org/10.1038/nprot.2015.035

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews. Genetics*, *7*(9), 703–13. http://doi.org/10.1038/nrg1925

Gdula, D. A., Gerasimova, T. I., & Corces, V. G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(18), 9378–83. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=38435&tool=pmcentrez&rendertype=abstract

Georgatos, S. D., Meier, J., & Simos, G. (1994). Lamins and lamin-associated proteins. *Current Opinion in Cell Biology*, *6*(3), 347–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7917324

Gerasimova, T. I., & Corces, V. G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annual Review of Genetics*, *35*, 193–208. http://doi.org/10.1146/annurev.genet.35.102401.090349

Gerasimova, T. I., Lei, E. P., Bushey, A. M., & Corces, V. G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Molecular Cell*, *28*(5), 761–72. http://doi.org/10.1016/j.molcel.2007.09.024

Geyer, P. K., & Corces, V. G. (1992). DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes & Development*, *6*(10), 1865–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1327958

Ghoussaini, M., Song, H., Koessler, T., Al Olama, A. A., Kote-Jarai, Z., Driver, K. E., … Dunning, A. M. (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. *Journal of the National Cancer Institute*, *100*(13), 962–6. http://doi.org/10.1093/jnci/djn190

Gillies, S. D., Morrison, S. L., Oi, V. T., & Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, *33*(3), 717–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6409417

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, *17*(6), 877–85. http://doi.org/10.1101/gr.5533506

Goldberg, M., Harel, A., Brandeis, M., Rechsteiner, T., Richmond, T. J., Weiss, A. M., & Gruenbaum, Y. (1999). The tail domain of lamin Dm0 binds histones H2A and H2B. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 2852–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=15858&tool=pmcentrez&rendertype=abstract

Gong, Q., & Dean, A. (1993). Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-

1/NF-E2. *Molecular and Cellular Biology*, *13*(2), 911–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=358974&tool=pmcentrez&rendertype=abstract

González, V., Guo, K., Hurley, L., & Sun, D. (2009). Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of Biological Chemistry*, *284*(35), 23622–35. http://doi.org/10.1074/jbc.M109.018028

González, V., & Hurley, L. H. (2010). The c-MYC NHE III(1): function and regulation. *Annual Review of Pharmacology and Toxicology*, *50*, 111–29. http://doi.org/10.1146/annurev.pharmtox.48.113006.094649

Gourdon, G., Morlé, F., Roche, J., Tourneur, N., Joulain, V., & Godet, J. (1992). Identification of GATA-1 and NF-E2 binding sites in the flanking regions of the human alpha-globin genes. *Acta Haematologica*, *87*(3), 136–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1642094

Grandori, C., Cowley, S. M., James, L. P., & Eisenman, R. N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annual Review of Cell and Developmental Biology*, *16*, 653–699. http://doi.org/10.1146/annurev.cellbio.16.1.653

Gromak, N., West, S., & Proudfoot, N. J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and Cellular Biology*, *26*(10), 3986–96. http://doi.org/10.1128/MCB.26.10.3986-3996.2006

Gross, D. S., & Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry*, *57*, 159–97. http://doi.org/10.1146/annurev.bi.57.070188.001111

Gruenbaum, Y., Goldman, R. D., Meyuhas, R., Mills, E., Margalit, A., Fridkin, A., … Enosh, A. (2003). The nuclear lamina and its functions in the nucleus. *International Review of Cytology*, *226*, 1–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12921235

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., … Stefansson, K. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature Genetics*, *39*(5), 631–7. http://doi.org/10.1038/ng1999

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., … van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, *453*(7197), 948–51. http://doi.org/10.1038/nature06947

Guibert, S., Zhao, Z., Sjölinder, M., Göndör, A., Fernandez, A., Pant, V., & Ohlsson, R. (2012). CTCF-binding sites within the H19 ICR differentially regulate local chromatin structures and cis-acting functions. *Epigenetics*, *7*(4), 361–9. http://doi.org/10.4161/epi.19487

Guillemot, F., Caspary, T., Tilghman, S. M., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., … Nagy, A. (1995). Genomic imprinting of Mash2, a mouse

gene required for trophoblast development. *Nature Genetics*, *9*(3), 235–42. http://doi.org/10.1038/ng0395-235

Haiman, C. A., Patterson, N., Freedman, M. L., Myers, S. R., Pike, M. C., Waliszewska, A., … Reich, D. (2007). Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature Genetics*, *39*(5), 638–44. http://doi.org/10.1038/ng2015

Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M., & Martens, J. A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. *Genes & Development*, *25*(1), 29–40. http://doi.org/10.1101/gad.1975011

Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N., & Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, *205*(1-2), 73–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9461381

Hargreaves, D. C., Horng, T., & Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell*, *138*(1), 129–45. http://doi.org/10.1016/j.cell.2009.05.047

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, *405*(6785), 486–9. http://doi.org/10.1038/35013106

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., & Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17507–12. http://doi.org/10.1073/pnas.1111941109

Heidemann, M., Hintermair, C., Voß, K., & Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta*, *1829*(1), 55–62. http://doi.org/10.1016/j.bbagrm.2012.08.013

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–12. http://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., … Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–8. http://doi.org/10.1038/ng1966

Henckel, A., Nakabayashi, K., Sanz, L. A., Feil, R., Hata, K., & Arnaud, P. (2009). Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals. *Human Molecular Genetics*, *18*(18), 3375–83. http://doi.org/10.1093/hmg/ddp277

Henriksson, M., & Lüscher, B. (1996). Proteins of the Myc network: essential

regulators of cell growth and differentiation. *Advances in Cancer Research*, *68*, 109–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8712067

Hérault, Y., Rassoulzadegan, M., Cuzin, F., & Duboule, D. (1998). Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nature Genetics*, *20*(4), 381–4. http://doi.org/10.1038/3861

Hirose, Y., & Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *Journal of Biochemistry*, *141*(5), 601–8. http://doi.org/10.1093/jb/mvm090

Hou, C., Dale, R., & Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3651–6. http://doi.org/10.1073/pnas.0912087107

Hsu, F.-M., Cheng, J. C.-H., Chang, Y.-L., Lee, J.-M., Koong, A. C., & Chuang, E. Y. (2015). Circulating mRNA Profiling in Esophageal Squamous Cell Carcinoma Identifies FAM84B As A Biomarker In Predicting Pathological Response to Neoadjuvant Chemoradiation. *Scientific Reports*, *5*, 10291. http://doi.org/10.1038/srep10291

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., … Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, *46*(2), 205–12. http://doi.org/10.1038/ng.2871

Ikegami, K., Egelhofer, T. A., Strome, S., & Lieb, J. D. (2010). Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biology*, *11*(12), R120. http://doi.org/10.1186/gb-2010-11-12-r120

Iqbal, K., Jin, S.-G., Pfeifer, G. P., & Szabó, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3642–7. http://doi.org/10.1073/pnas.1014033108

Isles, A. R., Davies, W., & Wilkinson, L. S. (2006). Genomic imprinting and the social brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2229–37. http://doi.org/10.1098/rstb.2006.1942

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & Development*, *28*(24), 2679–2692. http://doi.org/10.1101/gad.253443.114

Izzo, A., Kamieniarz-Gdula, K., Ramírez, F., Noureen, N., Kind, J., Manke, T., … Schneider, R. (2013). The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Reports*, *3*(6), 2142–54. http://doi.org/10.1016/j.celrep.2013.05.003

Jin, C., & Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development*, *21*(12), 1519–29. http://doi.org/10.1101/gad.1547707

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., & Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature Genetics*, *41*(8), 941–5. http://doi.org/10.1038/ng.409

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., … Roeder, R. G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. *Molecular Cell*, *45*(1), 51–63. http://doi.org/10.1016/j.molcel.2011.12.014

John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A., … Stamatoyannopoulos, J. A. (2011a). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics*, *43*(3), 264–8. http://doi.org/10.1038/ng.759

John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A., … Stamatoyannopoulos, J. A. (2011b). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics*, *43*(3), 264–8. http://doi.org/10.1038/ng.759

Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N., & Gallant, P. (1999). Drosophila myc regulates cellular growth during development. *Cell*, *98*(6), 779–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10499795

Jonkers, I., Kwak, H., & Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife*, *3*, e02407. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4001325&tool=pmcentrez&rendertype=abstract

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews. Molecular Cell Biology*, *16*(3), 167–177. http://doi.org/10.1038/nrm3953

Joos, S., Falk, M. H., Lichter, P., Haluska, F. G., Henglein, B., Lenoir, G. M., & Bornkamm, G. W. (1992). Variable breakpoints in Burkitt lymphoma cells with chromosomal t(8;14) translocation separate c-myc and the IgH locus up to several hundred kb. *Human Molecular Genetics*, *1*(8), 625–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1301171

Joos, S., Haluska, F. G., Falk, M. H., Henglein, B., Hameister, H., Croce, C. M., & Bornkamm, G. W. (1992). Mapping chromosomal breakpoints of Burkitt's t(8;14) translocations far upstream of c-myc. *Cancer Research*, *52*(23), 6547–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1330296

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., … Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430–5. http://doi.org/10.1038/nature09380

Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Müller, C. W., Vermeulen, M., & Müller, J. (2014). Histone H2A monoubiquitination

promotes histone H3 methylation in Polycomb repression. *Nature Structural & Molecular Biology*, *21*(6), 569–71. http://doi.org/10.1038/nsmb.2833

Katsani, K. R., Arredondo, J. J., Kal, A. J., & Verrijzer, C. P. (2001). A homeotic mutation in the trithorax SET domain impedes histone binding. *Genes & Development*, *15*(17), 2197–202. http://doi.org/10.1101/gad.201901

Kelly, M. L., Moir, L., Jones, L., Whitehill, E., Anstee, Q. M., Goldin, R. D., … Cox, R. D. (2009). A missense mutation in the non-neural G-protein alpha-subunit isoforms modulates susceptibility to obesity. *International Journal of Obesity (2005)*, *33*(5), 507–18. http://doi.org/10.1038/ijo.2009.30

Keng, V. W., Yae, K., Hayakawa, T., Mizuno, S., Uno, Y., Yusa, K., … Takeda, J. (2005). Region-specific saturation germline mutagenesis in mice using the Sleeping Beauty transposon system. *Nature Methods*, *2*(10), 763–9. http://doi.org/10.1038/nmeth795

Kiemeney, L. A., Thorlacius, S., Sulem, P., Geller, F., Aben, K. K. H., Stacey, S. N., … Stefansson, K. (2008). Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genetics*, *40*(11), 1307–12. http://doi.org/10.1038/ng.229

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., & Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, *432*(7016), 517–22. http://doi.org/10.1038/nature03041

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., … Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, *128*(6), 1231–45. http://doi.org/10.1016/j.cell.2006.12.048

Kim, T. K., Ebright, R. H., & Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science (New York, N.Y.)*, *288*(5470), 1418–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827951

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S. S., Janssen, H., … van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, *153*(1), 178–92. http://doi.org/10.1016/j.cell.2013.02.028

Kirkland, J. G., Raab, J. R., & Kamakaka, R. T. TFIIIC bound DNA elements in nuclear organization and insulation. *Biochimica et Biophysica Acta*, *1829*(3-4), 418–24. http://doi.org/10.1016/j.bbagrm.2012.09.006

Kirkland, J. G., Raab, J. R., & Kamakaka, R. T. (2013). TFIIIC bound DNA elements in nuclear organization and insulation. *Biochimica et Biophysica Acta*, *1829*(3-4), 418–24. http://doi.org/10.1016/j.bbagrm.2012.09.006

Kirov, N. C., Lieberman, P. M., & Rushlow, C. (1996). The transcriptional corepressor DSP1 inhibits activated transcription by disrupting TFIIA-TBP complex formation. *The EMBO Journal*, *15*(24), 7079–87. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452533&tool=pmcentrez&rendertype=abstract

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., … Andrau, J.-C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology*, *18*(8), 956–63. http://doi.org/10.1038/nsmb.2085

Köhler, C., Wolff, P., & Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annual Review of Plant Biology*, *63*, 331–52. http://doi.org/10.1146/annurev-arplant-042811-105514

Korn, R., Schoor, M., Neuhaus, H., Henseling, U., Soininen, R., Zachgo, J., & Gossler, A. (1992). Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. *Mechanisms of Development*, *39*(1-2), 95–109. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1283314

Krogan, N. J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O. W., Ding, H., … Greenblatt, J. F. (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Molecular Cell*, *12*(6), 1565–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14690608

Kuehner, J. N., Pearson, E. L., & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews. Molecular Cell Biology*, *12*(5), 283–94. http://doi.org/10.1038/nrm3098

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., … Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(28), 10684–9. http://doi.org/10.1073/pnas.0600326103

Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)*, *339*(6122), 950–3. http://doi.org/10.1126/science.1229386

Land, H., Parada, L. F., & Weinberg, R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature*, *304*(5927), 596–602. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6308472

Langdon, W. Y., Harris, A. W., Cory, S., & Adams, J. M. (1986). The c-myc oncogene perturbs B lymphocyte development in E-mu-myc transgenic mice. *Cell*, *47*(1), 11–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3093082

Lavenu, A., Pournin, S., Babinet, C., & Morello, D. (1994). The cis-acting elements known to regulate c-myc expression ex vivo are not sufficient for correct transcription in vivo. *Oncogene*, *9*(2), 527–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8290263

Lazo, P. A., Klein-Szanto, A. J., & Tsichlis, P. N. (1990). T-cell lymphoma lines derived from rat thymomas induced by Moloney murine leukemia virus: phenotypic diversity and its implications. *Journal of Virology*, *64*(8), 3948–59.

Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=249690&tool=pmcentrez&rendertype=abstract

Lazo, P. A., Lee, J. S., & Tsichlis, P. N. (1990). Long-distance activation of the Myc protooncogene by provirus insertion in Mlvi-1 or Mlvi-4 in rat T-cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(1), 170–3. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=53222&tool=pmcentrez&rendertype=abstract

Lee, J. T., & Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*, *152*(6), 1308–23. http://doi.org/10.1016/j.cell.2013.02.016

Lee, J.-K., & Kim, K.-C. (2013). DZNep, inhibitor of S-adenosylhomocysteine hydrolase, down-regulates expression of SETDB1 H3K9me3 HMTase in human lung cancer cells. *Biochemical and Biophysical Research Communications*, *438*(4), 647–52. http://doi.org/10.1016/j.bbrc.2013.07.128

Lee, K., Hsiung, C. C.-S., Huang, P., Raj, A., & Blobel, G. A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. *Genes & Development*, *29*(19), 1992–7. http://doi.org/10.1101/gad.255265.114

Lees-Murdock, D. J., & Walsh, C. P. DNA methylation reprogramming in the germ line. *Epigenetics*, *3*(1), 5–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18259118

Lemaitre, J. M., Buckle, R. S., & Méchali, M. (1996). c-Myc in the control of cell proliferation and embryonic development. *Advances in Cancer Research*, *70*, 95–144. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8902055

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. *Cell*, *145*(4), 502–11. http://doi.org/10.1016/j.cell.2011.04.021

Li, J., Liu, Y., Rhee, H. S., Ghosh, S. K. B., Bai, L., Pugh, B. F., & Gilmour, D. S. (2013). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Molecular Cell*, *50*(5), 711–22. http://doi.org/10.1016/j.molcel.2013.05.016

Li, J.-Y., Patterson, M., Mikkola, H. K. A., Lowry, W. E., & Kurdistani, S. K. (2012). Dynamic Distribution of Linker Histone H1.5 in Cellular Differentiation. *PLoS Genetics*, *8*(8), e1002879. http://doi.org/10.1371/journal.pgen.1002879

Li, X.-Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. A., & Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biology*, *12*(4), R34. http://doi.org/10.1186/gb-2011-12-4-r34

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–93. http://doi.org/10.1126/science.1181369

Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., … Young, R. A. (2012a). Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell*, *151*(1), 56–67. http://doi.org/10.1016/j.cell.2012.08.026

Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., … Young, R. A. (2012b). Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell*, *151*(1), 56–67. http://doi.org/10.1016/j.cell.2012.08.026

Lin, S.-P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., … Ferguson-Smith, A. C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nature Genetics*, *35*(1), 97–102. http://doi.org/10.1038/ng1233

Liu, J., Akoulitchev, S., Weber, A., Ge, H., Chuikov, S., Libutti, D., … Levens, D. (2001). Defective Interplay of Activators and Repressors with TFIIH in Xeroderma Pigmentosum. *Cell*, *104*(3), 353–363. http://doi.org/10.1016/S0092-8674(01)00223-9

Liu, J., Kouzine, F., Nie, Z., Chung, H.-J., Elisha-Feil, Z., Weber, A., … Levens, D. (2006). The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression. *The EMBO Journal*, *25*(10), 2119–30. http://doi.org/10.1038/sj.emboj.7601101

Liu, J., & Levens, D. (2006). Making myc. *Current Topics in Microbiology and Immunology*, *302*, 1–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16620023

Liu, M., Maurano, M. T., Wang, H., Qi, H., Song, C.-Z., Navas, P. A., … Stamatoyannopoulos, G. (2015). Genomic discovery of potent chromatin insulators for human gene therapy. *Nature Biotechnology*, *33*(2), 198–203. http://doi.org/10.1038/nbt.3062

Loc, P. V, & Strätling, W. H. (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *The EMBO Journal*, *7*(3), 655–64. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=454370&tool=pmcentrez&rendertype=abstract

Lokody, I. (2014). Enhancers: BRD4 and JMJD6 regulate antipause enhancers. *Nature Reviews Genetics*, *15*(2), 67–67. http://doi.org/10.1038/nrg3671

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., … Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, *153*(2), 320–34. http://doi.org/10.1016/j.cell.2013.03.036

Lu, X., Wontakal, S. N., Kavi, H., Kim, B. J., Guzzardo, P. M., Emelyanov, A. V., … Skoultchi, A. I. (2013). Drosophila H1 Regulates the Genetic Activity of Heterochromatin by Recruitment of Su(var)3-9. *Science*, *340*(6128), 78–81. http://doi.org/10.1126/science.1234654

Ludérus, M. E., de Graaf, A., Mattia, E., den Blaauwen, J. L., Grande, M. A., de Jong, L., & van Driel, R. (1992). Binding of matrix attachment regions to

lamin B1. *Cell*, *70*(6), 949–59. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1525831

Lunyak, V. V, Prefontaine, G. G., Núñez, E., Cramer, T., Ju, B.-G., Ohgi, K. A., … Rosenfeld, M. G. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science (New York, N.Y.)*, *317*(5835), 248–51. http://doi.org/10.1126/science.1140871

Luo, G., Ivics, Z., Izsvák, Z., & Bradley, A. (1998). Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10769–73. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27970&tool=pmcentrez&rendertype=abstract

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., … Mundlos, S. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, *161*(5), 1012–1025. http://doi.org/10.1016/j.cell.2015.04.004

Malhas, A., Lee, C. F., Sanders, R., Saunders, N. J., & Vaux, D. J. (2007). Defects in lamin B1 expression or processing affect interphase chromosome position and gene expression. *The Journal of Cell Biology*, *176*(5), 593–603. http://doi.org/10.1083/jcb.200607054

Malik, S., & Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews. Genetics*, *11*(11), 761–72. http://doi.org/10.1038/nrg2901

Mancini-Dinardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S., & Tilghman, S. M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes & Development*, *20*(10), 1268–82. http://doi.org/10.1101/gad.1416906

Marinić, M., Aktas, T., Ruf, S., & Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Developmental Cell*, *24*(5), 530–42. http://doi.org/10.1016/j.devcel.2013.01.025

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., … Eyre, S. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nature Communications*, *6*, 10069. http://doi.org/10.1038/ncomms10069

Martinez, A.-M., & Cavalli, G. (2006). The role of polycomb group proteins in cell cycle regulation during development. *Cell Cycle (Georgetown, Tex.)*, *5*(11), 1189–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16721063

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional Regulatory Elements in the Human Genome. Retrieved from http://www.annualreviews.org/doi/full/10.1146/annurev.genom.7.080505.115623?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed&

Matangkasombut, O., Buratowski, R. M., Swilling, N. W., & Buratowski, S.

(2000). Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes & Development*, *14*(8), 951–62. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=316539&tool=pmcentrez&rendertype=abstract

Mathad, R. I., Hatzakis, E., Dai, J., & Yang, D. (2011). c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Research*, *39*(20), 9023–33. http://doi.org/10.1093/nar/gkr612

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15 Spec No*, R17–29. http://doi.org/10.1093/hmg/ddl046

McCord, R. P., Nazario-Toole, A., Zhang, H., Chines, P. S., Zhan, Y., Erdos, M. R., … Cao, K. (2013). Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Research*, *23*(2), 260–9. http://doi.org/10.1101/gr.138032.112

McGrath, J., & Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, *37*(1), 179–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6722870

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., … Mikkelsen, T. S. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*, *30*(3), 271–7. http://doi.org/10.1038/nbt.2137

Merika, M., Williams, A. J., Chen, G., Collins, T., & Thanos, D. (1998). Recruitment of CBP/p300 by the IFNβ Enhanceosome Is Required for Synergistic Activation of Transcription. *Molecular Cell*, *1*(2), 277–287. http://doi.org/10.1016/S1097-2765(00)80028-3

Milon, B. C., Cheng, H., Tselebrovsky, M. V, Lavrov, S. A., Nenasheva, V. V, Mikhaleva, E. A., … Nurminsky, D. I. (2012). Role of histone deacetylases in gene regulation at nuclear lamina. *PloS One*, *7*(11), e49692. http://doi.org/10.1371/journal.pone.0049692

Missra, A., & Gilmour, D. S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences*, *107*(25), 11301–11306. http://doi.org/10.1073/pnas.1000681107

Mitchell, N. C., Tchoubrieva, E. B., Chahal, A., Woods, S., Lee, A., Lin, J. I., … Quinn, L. M. (2015). S6 Kinase is essential for MYC-dependent rDNA transcription in Drosophila. *Cellular Signalling*, *27*(10), 2045–53. http://doi.org/10.1016/j.cellsig.2015.07.018

Moberg, K. H., Logan, T. J., Tyndall, W. A., & Hall, D. J. (1992). Three distinct elements within the murine c-myc promoter are required for transcription. *Oncogene*, *7*(3), 411–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1549358

Mogno, I., Kwasnieski, J. C., & Cohen, B. A. (2013). Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Research*, *23*(11), 1908–15. http://doi.org/10.1101/gr.157891.113

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell*, *147*(5), 1132–1145. http://doi.org/10.1016/j.cell.2011.10.023

Moon, A. M., & Ley, T. J. (1990). Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(19), 7693–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=54814&tool=pmcentrez&rendertype=abstract

Mosch, K., Franz, H., Soeroes, S., Singh, P. B., & Fischle, W. (2011). HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. *PloS One*, *6*(1), e15894. http://doi.org/10.1371/journal.pone.0015894

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., & Shaw, W. C. (2009). Cleft lip and palate. *Lancet*, *374*(9703), 1773–85. http://doi.org/10.1016/S0140-6736(09)60695-4

Muravyova, E. (2001). Loss of Insulator Activity by Paired Su(Hw) Chromatin Insulators. *Science*, *291*(5503), 495–498. http://doi.org/10.1126/science.291.5503.495

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., … Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. http://doi.org/10.1038/nature12593

Natarajan, M., Schiralli Lester, G. M., Lee, C., Missra, A., Wasserman, G. A., Steffen, M., … Henderson, A. J. (2013). Negative elongation factor (NELF) coordinates RNA polymerase II pausing, premature termination, and chromatin remodeling to regulate HIV transcription. *The Journal of Biological Chemistry*, *288*(36), 25995–6003. http://doi.org/10.1074/jbc.M113.496489

Nieminen, A. I., Partanen, J. I., & Klefstrom, J. (2007). C-Myc blazing a trail of death: Coupling of the mitochondrial and death receptor apoptosis pathways by c-Myc. *Cell Cycle*. http://doi.org/10.4161/cc.6.20.4917

Nolis, I. K., McKay, D. J., Mantouvalou, E., Lomvardas, S., Merika, M., & Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. *Proceedings of the National Academy of Sciences*, *106*(48), 20222–20227. http://doi.org/10.1073/pnas.0902454106

Noma, K., Cam, H. P., Maraia, R. J., & Grewal, S. I. S. (2006). A role for TFIIIC transcription factor complex in genome organization. *Cell*, *125*(5), 859–72. http://doi.org/10.1016/j.cell.2006.04.028

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., … Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-

inactivation centre. *Nature*, *485*(7398), 381–5. http://doi.org/10.1038/nature11049

Ohtsuki, S., Levine, M., & Cai, H. N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. *Genes & Development*, *12*(4), 547–56. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=316525&tool=pmcentrez&rendertype=abstract

Ott, C. J., Blackledge, N. P., Kerschner, J. L., Leir, S.-H., Crawford, G. E., Cotton, C. U., & Harris, A. (2009). Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 19934–9. http://doi.org/10.1073/pnas.0900946106

Palumbo, A. P., Boccadoro, M., Battaglio, S., Corradini, P., Tsichlis, P. N., Huebner, K., … Croce, C. M. (1990). Human homologue of Moloney leukemia virus integration-4 locus (MLVI-4), located 20 kilobases 3' of the myc gene, is rearranged in multiple myelomas. *Cancer Research*, *50*(20), 6478–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2208106

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., … Merkenschlager, M. (2008a). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, *132*(3), 422–33. http://doi.org/10.1016/j.cell.2008.01.011

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., … Merkenschlager, M. (2008b). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, *132*(3), 422–33. http://doi.org/10.1016/j.cell.2008.01.011

Park, S. W., Li, G., Lin, Y.-P., Barrero, M. J., Ge, K., Roeder, R. G., & Wei, L.-N. (2005). Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. *Molecular Cell*, *19*(5), 643–53. http://doi.org/10.1016/j.molcel.2005.08.008

Parnell, T. J., & Geyer, P. K. (2000). Differences in insulator properties revealed by enhancer blocking assays on episomes. *The EMBO Journal*, *19*(21), 5864–74. http://doi.org/10.1093/emboj/19.21.5864

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., & Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and Cellular Biology*, *27*(10), 3769–79. http://doi.org/10.1128/MCB.01432-06

Pavri, R., Lewis, B., Kim, T.-K., Dilworth, F. J., Erdjument-Bromage, H., Tempst, P., … Reinberg, D. (2005). PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. *Molecular Cell*, *18*(1), 83–96. http://doi.org/10.1016/j.molcel.2005.02.034

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W. M., Solovei, I., Brugman, W., … van Steensel, B. (2010). Molecular maps of the

reorganization of genome-nuclear lamina interactions during differentiation. *Molecular Cell*, *38*(4), 603–13. http://doi.org/10.1016/j.molcel.2010.03.016

Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V, … Kelsey, G. (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7), 3830–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22380&tool=pmcentrez&rendertype=abstract

Peterson, C. W., & Ayer, D. E. (2011). An extended Myc network contributes to glucose homeostasis in cancer and diabetes. *Frontiers in Bioscience : A Journal and Virtual Library*, *17*, 2206–2223. http://doi.org/10.2741/3848

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C. P., Schaffner, W., & Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Research*, *33*(12), 3743–50. http://doi.org/10.1093/nar/gki689

Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., … Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, *153*(6), 1281–95. http://doi.org/10.1016/j.cell.2013.04.053

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M., & van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature Genetics*, *38*(9), 1005–14. http://doi.org/10.1038/ng1852

Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J., & Kelsey, G. (2004). The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. *Nature Genetics*, *36*(8), 818–26. http://doi.org/10.1038/ng1397

Plasschaert, R. N., & Bartolomei, M. S. (2014). Genomic imprinting in development, growth, behavior and stem cells. *Development (Cambridge, England)*, *141*(9), 1805–13. http://doi.org/10.1242/dev.101428

Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., & Sandelin, A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biology*, *7*(8), R78. http://doi.org/10.1186/gb-2006-7-8-R78

Popescu, N. C., DiPaolo, J. A., & Amsbaugh, S. C. (1987). Integration sites of human papillomavirus 18 DNA sequences on HeLa cell chromosomes. *Cytogenetic and Genome Research*, *44*(1), 58–62. http://doi.org/10.1159/000132342

Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Reviews. Molecular Cell Biology*, *16*(3), 190–202. http://doi.org/10.1038/nrm3943

Prickett, A. R., & Oakey, R. J. (2012). A survey of tissue-specific genomic imprinting in mammals. *Molecular Genetics and Genomics : MGG*, *287*(8), 621–

30. http://doi.org/10.1007/s00438-012-0708-6
Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., … Kamakaka, R. T. (2012). Human tRNA genes function as chromatin insulators. *The EMBO Journal*, *31*(2), 330–50. http://doi.org/10.1038/emboj.2011.406
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279–83. http://doi.org/10.1038/nature09692
Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., … Young, R. A. (2010). c-Myc regulates transcriptional pause release. *Cell*, *141*(3), 432–45. http://doi.org/10.1016/j.cell.2010.03.030
Rando, O. J., & Chang, H. Y. (2009). Genome-wide views of chromatin structure. *Annual Review of Biochemistry*, *78*, 245–71. http://doi.org/10.1146/annurev.biochem.78.071107.134639
Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665–1680. http://doi.org/10.1016/j.cell.2014.11.021
Rauen, K. A. (2013). The RASopathies. *Annual Review of Genomics and Human Genetics*, *14*, 355–69. http://doi.org/10.1146/annurev-genom-091212-153523
Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., … Reik, W. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development (Cambridge, England)*, *136*(4), 525–30. http://doi.org/10.1242/dev.031328
Ribeiro de Almeida, C., Stadhouders, R., Thongjuea, S., Soler, E., & Hendriks, R. W. (2012). DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation. *Blood*, *119*(26), 6209–18. http://doi.org/10.1182/blood-2012-03-402586
Richard, P., & Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes & Development*, *23*(11), 1247–69. http://doi.org/10.1101/gad.1792809
Rollins, R. A., Morcillo, P., & Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics*, *152*(2), 577–93. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460629&tool=pmcentrez&rendertype=abstract
Román, A. C., González-Rico, F. J., & Fernández-Salguero, P. M. (2011). B1-SINE retrotransposons: Establishing genomic insulatory networks. *Mobile Genetic Elements*, *1*(1), 66–70. http://doi.org/10.4161/mge.1.1.15455
Rosenwald, I. B. (1996). Upregulated expression of the genes encoding translation initiation factors eIF-4E and eIF-2alpha in transformed cells. *Cancer Letters*, *102*(1-2), 113–123.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., … Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(24), 8309–14. http://doi.org/10.1073/pnas.0801273105

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics*, *43*(4), 379–86. http://doi.org/10.1038/ng.790

Sahlén, P., Abdullayev, I., Ramsköld, D., Matskova, L., Rilakovic, N., Lötstedt, B., … Sandberg, R. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biology*, *16*(1), 156. http://doi.org/10.1186/s13059-015-0727-9

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108. http://doi.org/10.1038/nprot.2008.73

Schuettengruber, B., Martinez, A.-M., Iovino, N., & Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nature Reviews. Molecular Cell Biology*, *12*(12), 799–814. http://doi.org/10.1038/nrm3230

Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S., & Jeltsch, A. (2015). Activity and specificity of the human SUV39H2 protein lysine methyltransferase. *Biochimica et Biophysica Acta*, *1849*(1), 55–63. http://doi.org/10.1016/j.bbagrm.2014.11.005

Schulz, R., Woodfine, K., Menheniott, T. R., Bourc'his, D., Bestor, T., & Oakey, R. J. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics*, *3*(2), 89–96. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2492384&tool=pmcentrez&rendertype=abstract

Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nature Reviews. Genetics*, *14*(12), 853–64. http://doi.org/10.1038/nrg3603

Sebastiani, P., Timofeev, N., Dworkis, D. A., Perls, T. T., & Steinberg, M. H. (2009). Genome-wide association studies and the genetic dissection of complex traits. *American Journal of Hematology*, *84*(8), 504–15. http://doi.org/10.1002/ajh.21440

Seitan, V. C., Faure, A. J., Zhan, Y., McCord, R. P., Lajoie, B. R., Ing-Simmons, E., … Merkenschlager, M. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Research*, *23*(12), 2066–77. http://doi.org/10.1101/gr.161620.113

Sheiness, D., Fanshier, L., & Bishop, J. M. (1978). Identification of nucleotide sequences which may encode the oncogenic capacity of avian retrovirus MC29. *Journal of Virology*, *28*(2), 600–10. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=354308&tool=pmcentrez&rendertype=abstract

Shinkai, Y., & Tachibana, M. (2011). H3K9 methyltransferase G9a and the related molecule GLP. *Genes & Development*, *25*(8), 781–8. http://doi.org/10.1101/gad.2027411

Simon, J. A., & Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature Reviews. Molecular Cell Biology*, *10*(10), 697–708. http://doi.org/10.1038/nrm2763

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., … de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, *38*(11), 1348–54. http://doi.org/10.1038/ng1896

Sims, R. J., & Reinberg, D. (2009). Processing the H3K36me3 signature. *Nature Genetics*, *41*(3), 270–1. http://doi.org/10.1038/ng0309-270

Singh, A. M., & Dalton, S. (2009). The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. *Cell Stem Cell*. http://doi.org/10.1016/j.stem.2009.07.003

Singh, P., Lee, D.-H., & Szabó, P. E. (2012). More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. *Frontiers in Genetics*, *3*, 214. http://doi.org/10.3389/fgene.2012.00214

Sirotkin, A. M., Edelmann, W., Cheng, G., Klein-Szanto, A., Kucherlapati, R., & Skoultchi, A. I. (1995). Mice develop normally without the H1(0) linker histone. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(14), 6434–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41532&tool=pmcentrez&rendertype=abstract

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–3. http://doi.org/10.1038/415810a

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., … Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal*, *32*(24), 3119–3129. http://doi.org/10.1038/emboj.2013.237

Solovei, I., Wang, A. S., Thanisch, K., Schmidt, C. S., Krebs, S., Zwerger, M., … Joffe, B. (2013). LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell*, *152*(3), 584–98. http://doi.org/10.1016/j.cell.2013.01.009

Souleimani, A., & Asselin, C. (1993). Regulation of c-myc expression by sodium butyrate in the colon carcinoma cell line Caco-2. *FEBS Letters*, *326*(1-3), 45–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8325387

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews. Genetics*, *13*(9), 613–26. http://doi.org/10.1038/nrg3207

Spitz, F., Herkenne, C., Morris, M. A., & Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes.

Nature Genetics, *37*(8), 889–93. http://doi.org/10.1038/ng1597

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., … de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development*, *20*(17), 2349–54. http://doi.org/10.1101/gad.399506

Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M., & Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell*, *20*(2), 451–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7388947

Steffen, P. A., & Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nature Reviews. Molecular Cell Biology*, *15*(5), 340–56. http://doi.org/10.1038/nrm3789

Sun, H. B., Shen, J., & Yokota, H. (2000). Size-dependent positioning of human chromosomes in interphase nuclei. *Biophysical Journal*, *79*(1), 184–90. http://doi.org/10.1016/S0006-3495(00)76282-5

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., … Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, *24*(3), 390–400. http://doi.org/10.1101/gr.163519.113

Taniura, H., Glass, C., & Gerace, L. (1995). A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones. *The Journal of Cell Biology*, *131*(1), 33–44. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2120604&tool=pmcentrez&rendertype=abstract

Thorvaldsen, J. L., Duran, K. L., & Bartolomei, M. S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes & Development*, *12*(23), 3693–702. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=317260&tool=pmcentrez&rendertype=abstract

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., … Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development (Cambridge, England)*, *136*(18), 3131–41. http://doi.org/10.1242/dev.037127

Tomlinson, I. P. M., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., … Houlston, R. S. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genetics*, *40*(5), 623–30. http://doi.org/10.1038/ng.111

Towbin, B. D., Gonzalez-Sandoval, A., & Gasser, S. M. (2013). Mechanisms of heterochromatin subnuclear localization. *Trends in Biochemical Sciences*, *38*(7), 356–63. http://doi.org/10.1016/j.tibs.2013.04.004

Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G. R., & Bishop, J. M. (2001). c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature*, *414*(6865), 768–73.

http://doi.org/10.1038/414768a
Tsai, Z. T.-Y., Shiu, S.-H., & Tsai, H.-K. (2015). Contribution of Sequence Motif, Chromatin State, and DNA Structure Features to Predictive Models of Transcription Factor Binding in Yeast. *PLoS Computational Biology*, *11*(8), e1004418. http://doi.org/10.1371/journal.pcbi.1004418
Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. a, Marioni, J. C., & Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*, 1–24. http://doi.org/10.1038/ng.2971
Valenzuela, L., Dhillon, N., & Kamakaka, R. T. (2009). Transcription independent insulation at TFIIIC-dependent insulators. *Genetics*, *183*(1), 131–48. http://doi.org/10.1534/genetics.109.106203
Valenzuela, L., & Kamakaka, R. T. (2006). Chromatin insulators. *Annual Review of Genetics*, *40*, 107–38. http://doi.org/10.1146/annurev.genet.39.073003.113546
Valin, A., Ouyang, J., & Gill, G. (2013). Transcription factor Sp3 represses expression of p21CIP[1] via inhibition of productive elongation by RNA polymerase II. *Molecular and Cellular Biology*, *33*(8), 1582–93. http://doi.org/10.1128/MCB.00323-12
Van Bortle, K., & Corces, V. G. (2012). tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription*, *3*(6), 277–84. http://doi.org/10.4161/trns.21579
Van Bortle, K., Nichols, M. H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z. S., & Corces, V. G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biology*, *15*(5), R82. http://doi.org/10.1186/gb-2014-15-5-r82
van Riggelen, J., Yetil, A., & Felsher, D. W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature Reviews. Cancer*, *10*(4), 301–309. http://doi.org/10.1038/nrc2819
Vaquero, A., Scher, M., Erdjument-Bromage, H., Tempst, P., Serrano, L., & Reinberg, D. (2007). SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation. *Nature*, *450*(7168), 440–4. http://doi.org/10.1038/nature06268
Veloso, A., Kirkconnell, K. S., Magnuson, B., Biewen, B., Paulsen, M. T., Wilson, T. E., & Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*, *24*(6), 896–905. http://doi.org/10.1101/gr.171405.113
Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology*, *16*(3), 178–89. http://doi.org/10.1038/nrm3941
Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, *291*(5507), 1304–51. http://doi.org/10.1126/science.1058040
Verschure, P. J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A. E.,

Belmont, A. S., & van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Molecular and Cellular Biology*, *25*(11), 4552–64. http://doi.org/10.1128/MCB.25.11.4552-4564.2005

Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D., & Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(15), 4672–7. http://doi.org/10.1073/pnas.1504783112

Visel, A. (2009). ChIP-seq Accurately Predicts Tissue-Specific Activity of Enhancers. *Lawrence Berkeley National Laboratory*. Retrieved from http://escholarship.org/uc/item/51v7d58p

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research*, *35*(Database issue), D88–92. http://doi.org/10.1093/nar/gkl822

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. http://doi.org/10.1038/nature08451

Vogelmann, J., Valeri, A., Guillou, E., Cuvier, O., & Nollmann, M. Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. *Nucleus (Austin, Tex.)*, *2*(5), 358–69. http://doi.org/10.4161/nucl.2.5.17860

Voss, T. C., & Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews. Genetics*, *15*(2), 69–81. http://doi.org/10.1038/nrg3623

Vujatovic, O., Zaragoza, K., Vaquero, A., Reina, O., Bernués, J., & Azorín, F. (2012). Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic Acids Research*, *40*(12), 5402–14. http://doi.org/10.1093/nar/gks224

Wada, H., & Netz, R. R. (2009). Plectoneme creation reduces the rotational friction of a polymer. *EPL (Europhysics Letters)*, *87*(3), 38001. http://doi.org/10.1209/0295-5075/87/38001

Wagschal, A., Sutherland, H. G., Woodfine, K., Henckel, A., Chebli, K., Schulz, R., … Feil, R. (2008). G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Molecular and Cellular Biology*, *28*(3), 1104–13. http://doi.org/10.1128/MCB.01111-07

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., & Cremer, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *The Journal of Cell Biology*, *160*(5), 685–97. http://doi.org/10.1083/jcb.200211103

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., … Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, *22*(9), 1680–1688.

http://doi.org/10.1101/gr.136101.111
Wasserman, N. F., Aneas, I., & Nobrega, M. a. (2010a). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer, 1191–1197. http://doi.org/10.1101/gr.105361.110
Wasserman, N. F., Aneas, I., & Nobrega, M. A. (2010b). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Research*, *20*(9), 1191–7. http://doi.org/10.1101/gr.105361.110
Weber, F., de Villiers, J., & Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell*, *36*(4), 983–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6323032
Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., … Peters, J.-M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796–801. http://doi.org/10.1038/nature06634
West, A. G., Gaszner, M., & Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes & Development*, *16*(3), 271–88. http://doi.org/10.1101/gad.954702
Wierstra, I., & Alves, J. (2008). The c-myc promoter: still MysterY and challenge. *Advances in Cancer Research*, *99*, 113–333. http://doi.org/10.1016/S0065-230X(07)99004-1
Wilkins, J. F. (2014). Genomic imprinting of Grb10: coadaptation or conflict? *PLoS Biology*, *12*(2), e1001800. http://doi.org/10.1371/journal.pbio.1001800
Wittig, B., Wölfl, S., Dorbic, T., Vahrson, W., & Rich, A. (1992). Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *The EMBO Journal*, *11*(12), 4653–63. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=557041&tool=pmcentrez&rendertype=abstract
Wood, A. M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B. C., … Corces, V. G. (2011). Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Molecular Cell*, *44*(1), 29–38. http://doi.org/10.1016/j.molcel.2011.07.035
Xiao, T., Wallace, J., & Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Molecular and Cellular Biology*, *31*(11), 2174–83. http://doi.org/10.1128/MCB.05093-11
Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M., & Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proceedings of the*

National Academy of Sciences of the United States of America, *104*(17), 7145–50. http://doi.org/10.1073/pnas.0701811104

Xu, J., Watts, J. A., Pope, S. D., Gadue, P., Kamps, M., Plath, K., … Smale, S. T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes & Development*, *23*(24), 2824–38. http://doi.org/10.1101/gad.1861209

Yamaguchi, Y., Shibata, H., & Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochimica et Biophysica Acta*, *1829*(1), 98–104. http://doi.org/10.1016/j.bbagrm.2012.11.007

Yamashita, R., Suzuki, Y., Sugano, S., & Nakai, K. (2005). Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene*, *350*(2), 129–36. http://doi.org/10.1016/j.gene.2005.01.012

Yang, J., Ramos, E., & Corces, V. G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. *Genome Research*, *22*(11), 2199–207. http://doi.org/10.1101/gr.142125.112

Yie, J., Merika, M., Munshi, N., Chen, G., & Thanos, D. (1999). The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome. *The EMBO Journal*, *18*(11), 3074–89. http://doi.org/10.1093/emboj/18.11.3074

Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E., & Korsmeyer, S. J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10632–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27946&tool=pmcentrez&rendertype=abstract

Yukawa, M., Akiyama, T., Franke, V., Mise, N., Isagawa, T., Suzuki, Y., … Aoki, F. (2014). Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. *PloS One*, *9*(3), e92689. http://doi.org/10.1371/journal.pone.0092689

Yusufzai, T. M., Tagami, H., Nakatani, Y., & Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Molecular Cell*, *13*(2), 291–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14759373

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2014). Enhancer—core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, *518*(7540), 556–9. http://doi.org/10.1038/nature13994

Zanke, B. W., Greenwood, C. M. T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M., … Dunlop, M. G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature Genetics*, *39*(8), 989–94. http://doi.org/10.1038/ng2089

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing

competence for gene expression. *Genes & Development*, *25*(21), 2227–41. http://doi.org/10.1101/gad.176826.111

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., … Young, R. A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Genetics*, *39*(12), 1512–6. http://doi.org/10.1038/ng.2007.26

Zentner, G. E., & Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*, *20*(3), 259–66. http://doi.org/10.1038/nsmb.2470

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, *21*(8), 1273–83. http://doi.org/10.1101/gr.122382.111

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., … Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, *38*(11), 1341–7. http://doi.org/10.1038/ng1891

Zhou, V. W., Goren, A., & Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews. Genetics*, *12*(1), 7–18. http://doi.org/10.1038/nrg2905

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., … Rosenfeld, M. G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Molecular Cell*, *29*(1), 69–80. http://doi.org/10.1016/j.molcel.2007.11.002

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., & Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell*, *138*(6), 1122–36. http://doi.org/10.1016/j.cell.2009.07.031

Zlatanova, J., & Caiafa, P. (2009). CTCF and its protein partners: divide and rule? *Journal of Cell Science*, *122*(Pt 9), 1275–84. http://doi.org/10.1242/jcs.039990

Zlotorynski, E. (2015). Chromosome biology: CTCF-binding site orientation shapes the genome. *Nature Reviews. Molecular Cell Biology*, *16*(10), 578–579. http://doi.org/10.1038/nrm4057

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., … Wendt, K. S. (2013). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences*, *111*(3), 996–1001. http://doi.org/10.1073/pnas.1317788111

Zuleger, N., Boyle, S., Kelly, D. A., de las Heras, J. I., Lazou, V., Korfali, N., … Schirmer, E. C. (2013). Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biology*, *14*(2), R14. http://doi.org/10.1186/gb-2013-14-2-r14

8 Appendix

8.1 List of Figures

8.2 Abbreviations and Acronym

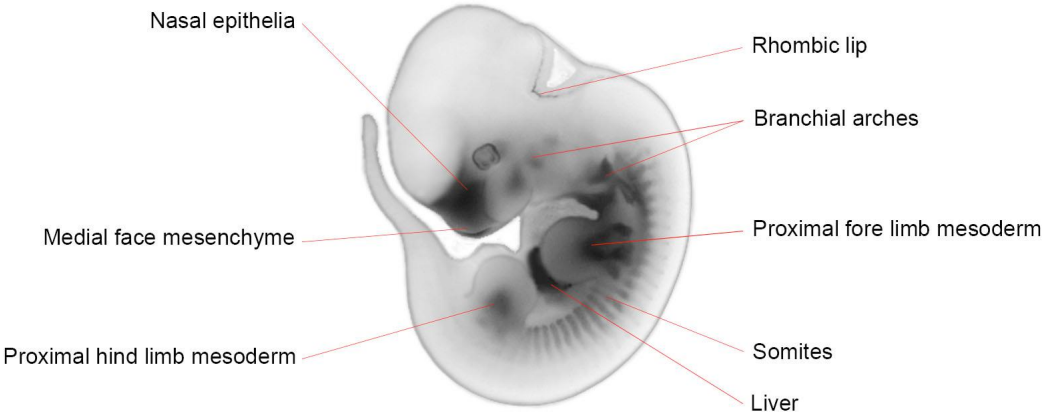
Acronym	Full name
$(NH_4)_2SO_4$	Ammonium sulfate
3'UTR	3' Untranslated region
3C	Chromatin conformation capture
4C	Circularized chromatin conformation capture
5C	Chromatin conformation capture carbon copy
A1bg	Alpha-1-B glycoprotein
ANOVA	Analysis of variance
AP1/FOS-JUN	V-Jun Avian Sarcoma Virus 17 Oncogene Homolog
ATP	Adenosine triphosphate
BA	Branchial arches
BEAF	Boundary element associated factor
BET	Bromodomain-extraterminal
BMP	Bone morphogenetic protein
bp	Base pairs
Brd2/3/4	Bromodomain containing 2/3/4
Brdt	Bromodomain testis-specific
BRE	TFIIB recognition element
c-Myc	*v-Myc* Avian Myelocytomatosis Viral Oncogene Homolog
CaCl2	Calcium chloride
Capture-C	Hi-C with sequence capture
ChIP-seq	Chromatin immunopreciptation-sequencing
CLP	Cleft lip and/or cleft palate
CPF-CF	Cleavage and polyadenylation factor-cleavage factor
Cre	Cre recombinase
CREB	cAMP response element-binding protein
Ct	Treshold cycle
CT	Chromosome territories
CTCF	CCCTC-binding factor
CTD	Carboxy terminal domain
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
DPE	Downstream Promoter Element
E2F(s)	Retinoblastoma-Binding Protein
E430025E21Rik	Spastic paraplegia 8 (autosomal dominant)
EDTA	Ethylenediaminetetraacetic acid
EGF	Epithelial growth factor

EGTA	Tetra(acetoxymethyl Ester)
EMBL	European molecular biology laboratory
ER	Estrogen receptor
Fam49b	Family with sequence similarity 49, member b
Fam84b	Family with sequence similarity 84, member B
FB	Fore brain
FBP	FUSE binding protein
FBS	Fetal serum bovine
FIR	FUSE interacting repressor
FISH	Fluorescence in situ hybridization
FM	Facial mesenchyme
Ftz	Fushi tarazu
FUSE	Far upstream sequence element
GRO-cap	Genomic run-on-5'cap enriched
GRO-seq	Genomic run-on-sequencing
GROMIT	Genome regulatory organization mapping with integrated transposons
Gsdmc	Gasdermin C
GTF	General transcription factors
Gusb	Glucuronidase beta
GWAS	Genome wide association studies
H2AK119ub	H2A lysine 119 ubiquitination
H_2O	Water
H3K27ac	H3 lysine 27 acetylation
H3K27me3	H3 lysine 27 tri-methylation
H3K4me1	H3 lysine 4 mono-methylation
H3K4me3	H3 lysine 4 tri-methylation
H3K9me3	H3 lysine 9 tri-methylation
Hbb	Hemoglobin subunit beta
Hi-cap	Hi-C with sequence capture of promoter regions
HMGA1	High mobility group proteins A1
hnRNP	Heterogeneous ribonucleoprotein particle
HP1	Heterochromatin protein 1
Hprt	Hypoxanthine phosphoribosyltransferase 1
HS4	Chicken hypersensitivity site 4
HSB16	Hyperactive SB16 transposase
ICR	Imprinting control regions
Igf2	Insulin like growth factor 2
IgH	Immunoglobulin heavy chain
INV1	Inversion1
INV2	Inversion2

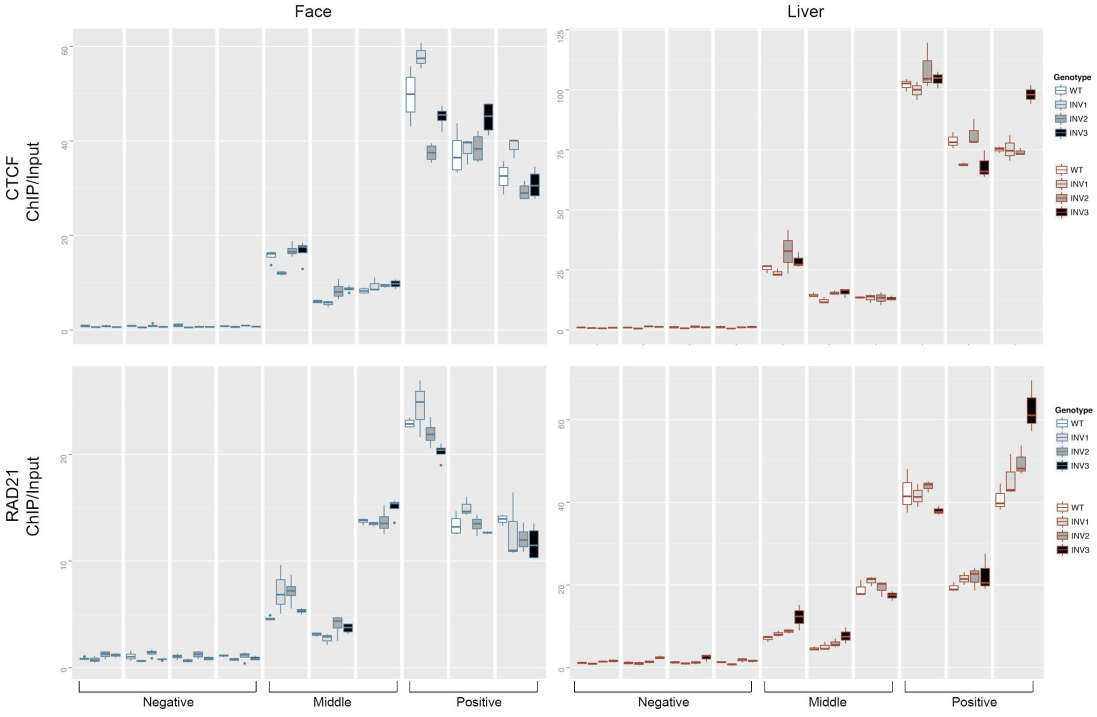
INV3	Inversion3
Irf6	Interferon regulatory factor 6
$K_3Fe_3(CN)_6$	Potassium ferrocyanide
$K_4Fe_2(CN)_6$	Potassium ferrocyanide
KCl	Potassium chloride
KH_2PO_4	Monopotassium phosphate
LacZ	β galactosidase
LAP	Lamin-associated membrane proteins
LCR	Locus control region
LEF-TCF	lymphoid enhancer-binding factor - transcription factor 7 (T-cell specific, HMG-Box)
LFM	Lateral face mesenchyme
MAR	Matrix attachment regions
Med12	Mediator complex subunit 12
mESC	Mouse embryonic stem cells
MFM	Medial facial mesenchyme
$MgCl_2$	Magnesium chloride
MoMuLV	Moloney murine leukemia virus
MTE	Motif Ten Element
Na_2HPO_4	Sodium phosphate
NaCl	Sodium chloride
NaDeoxycholate	Sodium deoxycholate
NE	Nasal epithelia
NFκB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
NGS	Next-generation-sequencing
Nsmce2	Non-SMC element 2, MMS21 homolog
ON	Over night
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PFLM	Proximal fore limb mesoderm
PIC	Pre-initiation complex
PK	Proteinase K
PLM	Proximal limb mesoderm
Prm1	Protamine gene
Pvt1	Pvt1 oncogene (non-protein coding)
qPCR	Quantitative PCR
RD	Regulatory domain
Rfx(s)	Regulatory Factor X
RIPA	Radioimmunoprecipitation assay buffer

RL	Rhombic lip
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RT	Room temperature
RT-qPCR	Reverse transcript qPCR
SB	Sleeping beauty transposon
SCC1	Sister-chromatid cohesion protein 1
SCC3/SA	Sister-chromatid cohesion protein 3/Stromalin
Sdha	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
SDS	Sodium dodecyl sulphate
Segment-C	Region of the telomeric end of *c-Myc locus* covering the ~130Kbp centromeric to the GROMIT insertion 179039 (17a)
Segment-T	Region of the telomeric end of *c-Myc locus* extending from the GROMIT insertion 179039 (17a) to the gene *Gsdmcl-ps*
SINE	Short interspersed element
SMAD(s)	Sma-And Mad-related protein(s)
SMC1	Cohesin subunit1
SMC3	Cohesin subunit3
SNP(s)	Single nucleotide polymorphism(s)
Sp1-3	Sp1-3 Transcription Factor
Src	Sarcoma
ssDNA	Single strand DNA
Stat(s)	Signal transducer and activator of transcription
su(Hw)	Protein suppressor of hairy wing
SWI/SNF	Switch/Sucrose non-fermentable
TAD	Topologically associated domans
Taq	Thermus aquaticus
Tbp	TATA box binding protein
TF	Transcription factor
TGF-β	Transforming growth factor β
Trib1	Tribbles pseudokinase 1
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
USF1	Upstream transcription factor 1
Vax1	Ventral anterior homeobox 1
VEZF1	Vascular endothelial zinc finger 1
WNT	Wingless-type MMTV (mouse mammary tumour virus)
Wt	Wild type

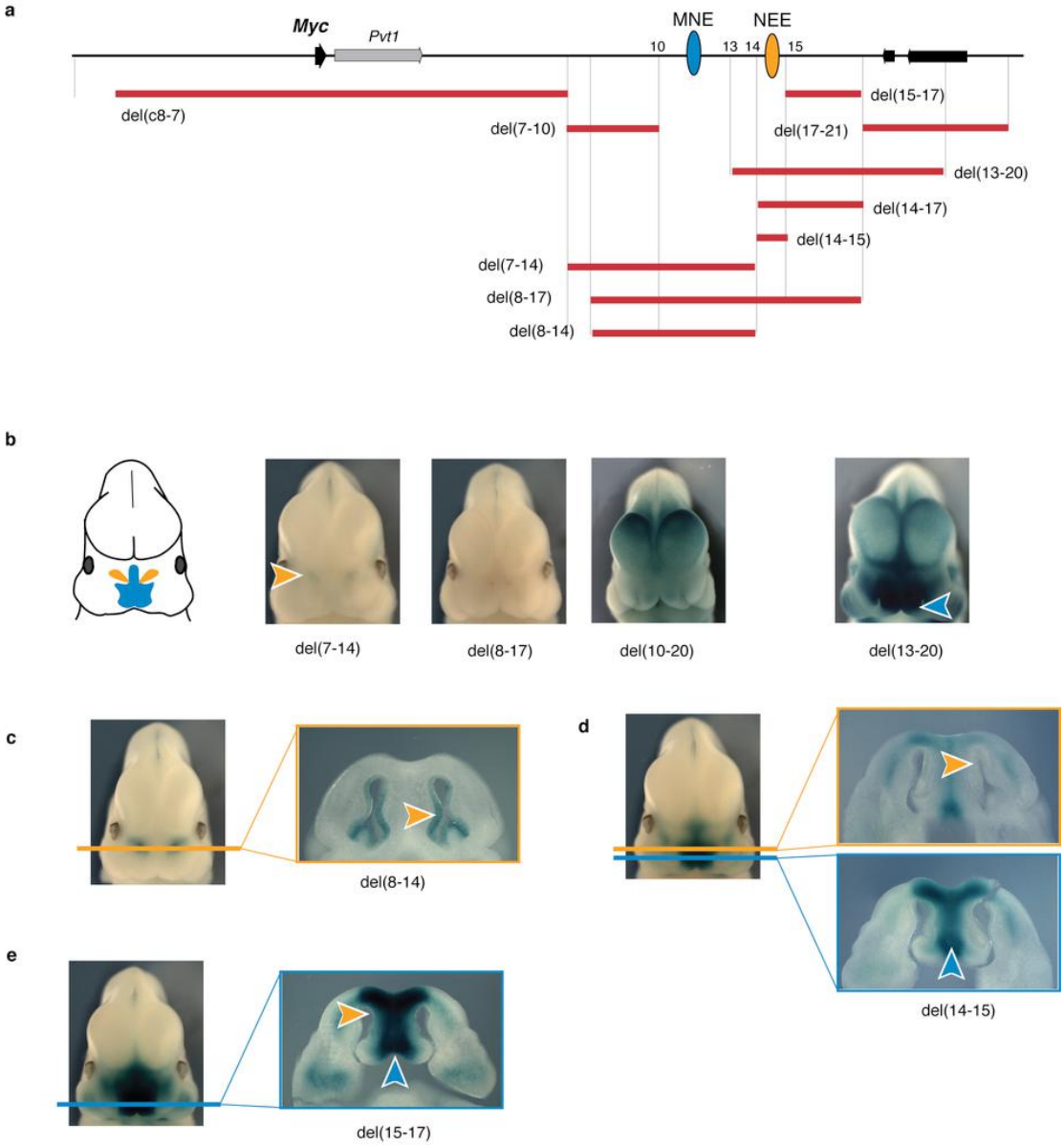
8.3 Supplementary figures



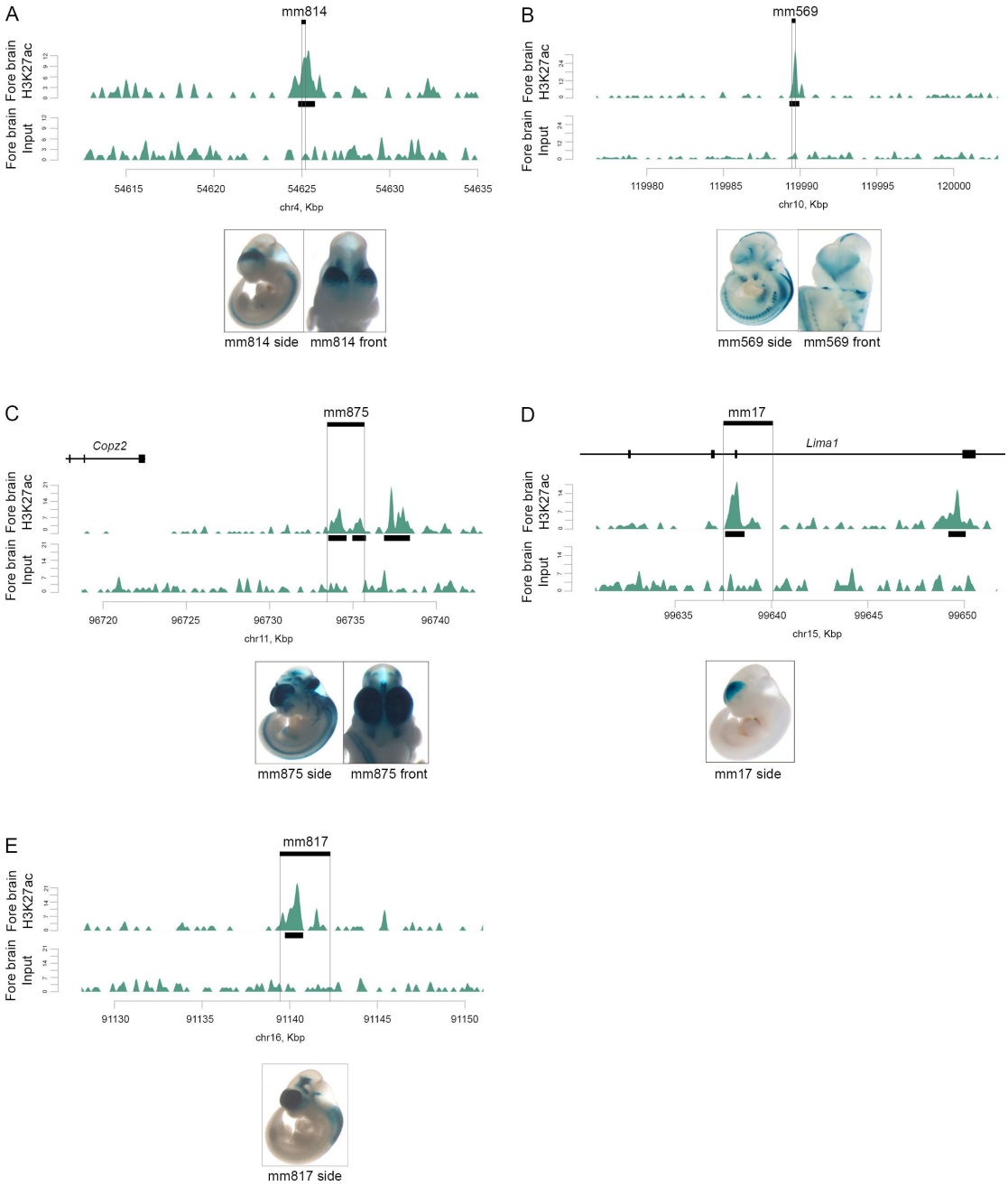
Supplementary figure 1 | *c-Myc locus* regulatory inputs. Schematic representation of all the regulatory inputs captured by GROMIT insertions integrated inside *c-Myc locus*.



Supplementary figure 2 | Architectural proteins binding on control regions, in inversion lines. CTCF and RAD21 occupancy at selected control regions in Wt, INV1$^{+/+}$, INV2$^{+/+}$, INV3$^{+/+}$ E11.5 facial mesenchyme and liver (sample size $n \geq 3$). The y-axis represents the fold enrichment of ChIP samples to the input chromatin control (background = 1). The control regions were chosen outside *c-Myc locus* (assumed to be unaffected by the chromosomal inversions) and were selected for having different degrees of enrichment compared to the input. They are labeled as negative, middle and positive controls.



Supplementary figure 3 | Identification of facial enhancers in *c-Myc locus.* Picture obtained from (Uslu et al., 2014) (supplementary figure 2). A) Locations of the deletion lines generated to identify the facial regulatory elements. The approximate position of the medial face mesenchyme (MFM) and the nasal epithelia (NE) enhancers is depicted by a blue and a yellow ellipse respectively. MNE, medionasal enhancer; NEE, nasal epithelial enhancer. B–E) Frontal view of the head of E11.5 mouse embryos carrying different chromosomal deletions, subjected to LacZ staining procedure. The regions corresponding to the MFM and NE are shown by a blue and a yellow arrow respectively. C–E) Vibrotome sections (150µm) of the face of E11.5 mouse embryos (from different deletion lines) subjected to LacZ staining procedure. C) Panel showing the loss of LacZ staining in the MFM and the preservation of the LacZ staining in the NE, upon Del(8–14). D) Pictures showing the LacZ staining present in the facial mesenchyme (both medial and lateral) but absent in the NE. E) Pictures showing the strong LacZ expression in the MFM and the weak (although still present) one on the NE.



Supplementary figure 4 | Sequences driving reporter expression in the fore brain. A–E) Picture showing the LacZ staining pattern of mouse embryos at stage E11.5, subjected to reporter assays (Visel, Minovitsky, Dubchak, & Pennacchio, 2007). The genomic locations, and the H3K27ac density (ChIP-seq on fore brain tissue of E11.5 Wt embryos) are shown on the top of each panel. Peak called regions and input chromatin controls are shown underneath the H3K27ac tracks.

8.4 Acknowledgements

I thank Dr. François Spitz for giving me the opportunity to work in his laboratory and the TAC committee members: Dr. Alexander Aulehla, Dr. Wolfgang Huber, Prof. Dr. Herbert Steinbeisser and Prof. Dr. Joachim Wittbrodt, for their availability and for providing useful feedbacks during the TAC sessions.

I thank all the people in Spitz laboratory and in the institute in general who spent some of their valuable time to help me, showing great friendship and cooperative attitude.

I thank Veli Uslu for generating the material that constituted the basis of this thesis and for providing me the GROMIT lines 196231, 192857 and 194578 necessary to create the line INV1 and INV3. I thank him also for introducing me the peculiar laboratory techniques and his work on *c-Myc locus*, being always present for troubleshooting and for his help in countless occasions, amongst which the large scale tissue collection dedicated to the enhancer marks ChIP-seq. Finally, I thank Veli Uslu for authorizing me to use his unpublished data regarding the *c-Myc locus* regulatory landscape and the imprinting.

I thank Sandra Ruf for providing me with the GROMIT line 179039 (used for creating INV2) and for maintaining the stocks of Hprt:::Cre and Prm1:::HSB16 animals, always ready to be bred, indispensable to generate all the inversion lines and to remove either of the two SB transposons from them.

I thank Katja Langenfeld and Leslie Pan for their help with the large-scale tissue collection aimed to perform the H2AZ and CTCF ChIP-seq, spending their own time, sometimes even on weekends, in order to help completing these experiments.

I thank the EMBL facilities, in particular for the LAR, Silke Brohn, Silke Feller, Andrea Genthner, Simone Plaztek and Michaela Wesch, for all their work and for managing time consuming and complicated requests, often very difficult to arrange with the resources available at the facility. I wish to acknowledge also Christian Tischer from the microscopy facility for his constant assistance with the imaging. Finally, I thank the genomics facility, in particular Bianka Baying and Paul Collier for their maniacal care in every procedure and for being always present (often dedicating a lot of extra time) to provide helps and tips on top of their standard assistance duties.

The work I described here in this thesis, despite little, would not have been possible without the contribution of these people and without their good care.

9 Annex

9.1 Publications

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. a, Marioni, J. C., & Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*, 1–24. http://doi.org/10.1038/ng.2971.

Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis

Veli Vural Uslu[1], Massimo Petretich[1], Sandra Ruf[1], Katja Langenfeld[1], Nuno A Fonseca[2], John C Marioni[2] & François Spitz[1]

Cleft lip with or without cleft palate (CL/P) is one of the most common congenital malformations observed in humans, with 1 occurrence in every 500–1,000 births[1,2]. A 640-kb noncoding interval at 8q24 has been associated with increased risk of non-syndromic CL/P in humans[3–5], but the genes and pathways involved in this genetic susceptibility have remained elusive. Using a large series of rearrangements engineered over the syntenic mouse region, we show that this interval contains very remote *cis*-acting enhancers that control *Myc* expression in the developing face. Deletion of this interval leads to mild alteration of facial morphology in mice and, sporadically, to CL/P. At the molecular level, we identify misexpression of several downstream genes, highlighting combined impact on the craniofacial developmental network and the general metabolic capacity of cells contributing to the future upper lip. This dual molecular etiology may account for the prominent influence of variants in the 8q24 region on human facial dysmorphologies.

CL/P is present in a large number of malformation syndromes with mendelian inheritance, but most cases correspond to isolated non-syndromic forms[1]. The influence of environmental risk factors[6] further complicates definition of the genetic components of CL/P. Genome-wide association studies (GWAS) have identified common variants contributing to increased risk of CL/P[3–5], defining several susceptibility loci. Of these, a 640-kb locus at 8q24 showed the strongest effect size in different populations, particularly in those with European ancestry[1,3–5]. This large interval corresponds to a gene desert (**Fig. 1**). Such situations are a common outcome of GWAS and constitute a challenge in moving from statistical associations to functional mechanisms[7], particularly for phenotypes that cannot be easily reproduced in cell lines.

To obtain insights into the role of the 8q24 noncoding interval during craniofacial development and to identify the gene(s) and pathway(s) it can influence, we sought to dissect its function *in vivo*. Similarity of craniofacial development in mice and humans[8] and conserved synteny of the 8q24 region (**Supplementary Fig. 1**) suggested that the mouse constitutes a suitable model. To identify regulatory activities present in this interval, we exploited a mouse line carrying, at the distal end of the locus, an insertion of a *LacZ* reporter gene whose activity reflects endogenous surrounding regulatory activities[9]. As this regulatory sensor lay in a *Sleeping Beauty* transposon, we produced, by repeated *in vivo* transposition, 38 new insertions in the locus (**Fig. 1b** and **Supplementary Table 1**). Strikingly, insertions within or adjacent to the region orthologous to the CL/P-associated interval showed specific LacZ expression in the developing fronto- and medionasal process of embryonic day (E) 11.5 mouse embryos, when this process fuses with the maxillary process to form the future upper lip[10] (**Fig. 1** and **Supplementary Table 1**).

To more precisely localize the region(s) responsible for this regulatory activity, we used the *lox*P site present in the transposon to engineer a series of rearrangements spanning the entire interval[9] (**Fig. 1b** and **Supplementary Table 2**). We first monitored the effects of the different deletions on expression of the reconstituted *LacZ* sensor (**Fig. 1f** and **Supplementary Fig. 2**). Confirming our earlier observations[9], del(8–17) showed no LacZ expression in the developing face. All deletions encompassing the 8–14 interval also led to loss of LacZ expression in the medionasal process, whereas deletions retaining the 10–13 interval had no or minor effects on this expression domain. We also analyzed mice carrying different duplications that 'moved' genomic intervals away from the CL/P-associated interval, linking them to a different regulatory domain (**Supplementary Fig. 3**). We observed LacZ expression in the medionasal process in dup(10–20) and dup(8–21) but not in dup(13–20) embryos (**Supplementary Fig. 3**), confirming the key role of the 10–13 interval. Altogether, the 14 overlapping rearrangements showed that the 280-kb region between 10a and 13a contains specific regulatory information relevant to CL/P, defining a medionasal enhancer (MNE) region. To identify potential regulatory modules constituting the MNE region, we carried out chromatin immunoprecipitation (ChIP) on face samples from E11.5 embryos for enhancer-associated histone modifications[11,12] (**Fig. 1g** and **Supplementary Table 3**). The 10–13 interval contained several regions enriched for acetylation of histone H3 at lysine 27 (H3K27ac) and monomethylation of histone H3 at lysine 4 (H3K4me1) (**Fig. 1g**, bottom). These regions are often evolutionarily conserved and were not detected in other tissues or cell types[13], suggesting that they have a specific function in the face. Supporting this hypothesis, it was recently shown that the human sequence

[1]Developmental Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [2]European Bioinformatics Institute–European Molecular Biology Laboratory (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. Correspondence should be addressed to F.S. (spitz@embl.de).

Received 13 December 2013; accepted 8 April 2014; published online 25 May 2014; doi:10.1038/ng.2971

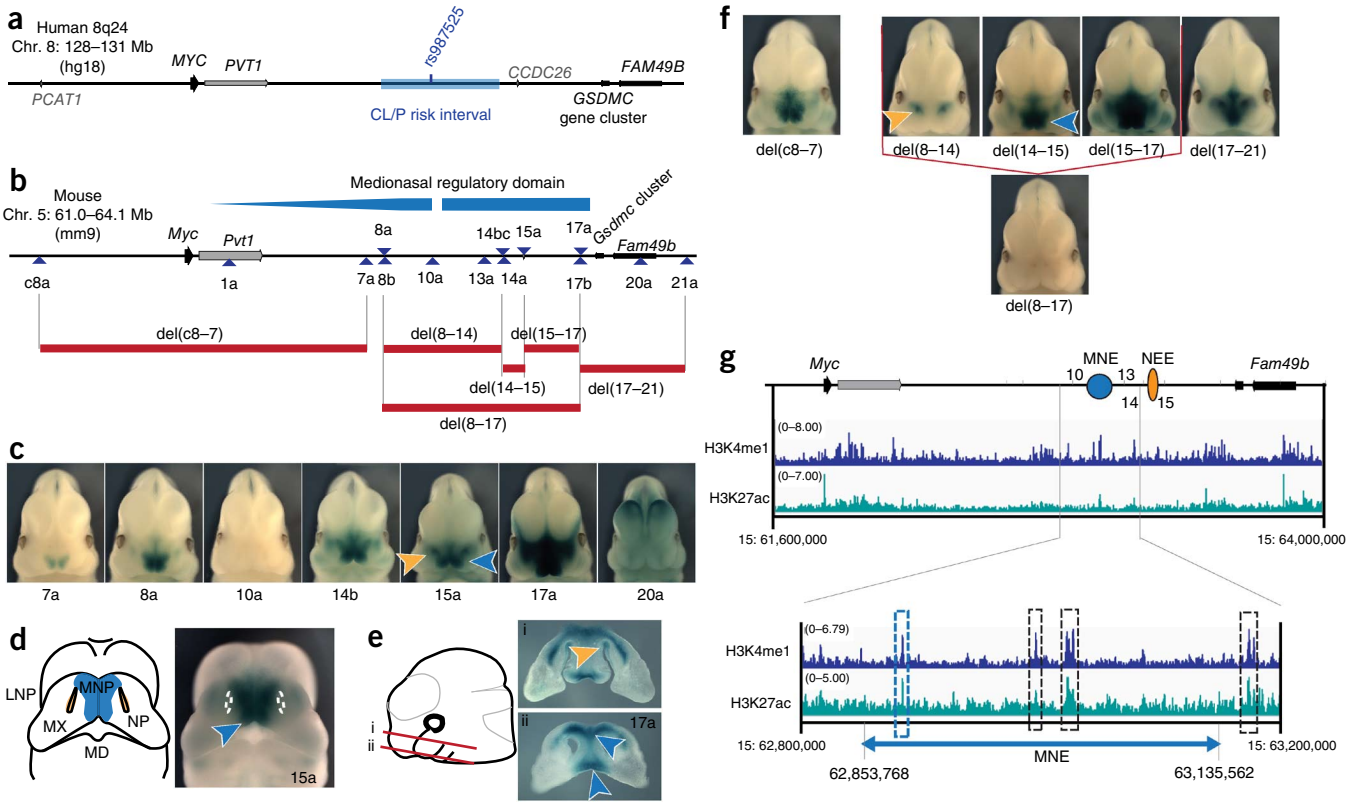


Figure 1 Functional characterization of the 8q24 CL/P regulatory landscape. (**a**) The human 8q24 interval associated with CL/P risk. Genes are shown as plain arrows (black, protein-coding genes; gray, annotated noncoding transcripts). The CL/P interval[3] and the most significantly associated SNP (rs987525) are indicated in blue. (**b**) Syntenic organization of the mouse locus, depicting transposon insertions (blue triangles) and deletions (red bars) used in this study. An expanded list of insertions and alleles is given in **Supplementary Table 1**. The expression patterns of adjacent insertions (shown in **c**) define a broad 'medionasal regulatory domain' indicated by a blue bar whose width represents relative LacZ expression levels. (**c**) LacZ staining of E11.5 embryos with various insertions, with insertion number indicated below. Arrowheads indicate expression in the medionasal process (blue) and in the nasal epithelium (orange). Insertions located in the *Gsdmc* gene cluster did not result in any expression, whereas the more telomeric ones around *Fam49b* resulted in a different expression pattern. (**d**,**e**) Magnified view (**d**) and sections (**e**) of E11.5 embryos with strong LacZ expression in the medionasal process (MNP; blue) and nasal epithelium (NP; orange). MX, maxillary process; MD, mandibulary process; LNP, lateral nasal process. Most insertions in the 7a–17a interval also showed LacZ expression in somites and in the limb mesenchyme of E11.5 embryos but not in heart or liver[9]. (**f**) LacZ staining in the faces of E11.5 embryos carrying the different deletions. The medionasal process and nasal epithelium expression domains can be separated by different deletions, highlighting two distinct regulatory regions, the MNE and NEE regions, respectively. See also the description of additional lines and sections in **Supplementary Figures 2** and **3**. Precise genomic positions are given in **Supplementary Table 2**. (**g**) Enrichment for enhancer-associated marks (blue, H3K4me1; turquoise, H3K27ac) profiled by ChIP sequencing (ChIP-seq) of E11.5 facial tissues highlighted several candidate face-specific regulatory regions (dashed boxes; positions and evolutionary conservation are given in **Supplementary Table 3**). One representative profile of two biological replicates is shown for each track. The blue boxed peak corresponds to Vista element hs1877, which showed enhancer activity in a number of regions, including the facial mesenchyme[14].

corresponding to one of these face-specific H3K27ac and H3K4me1 peaks showed, in E11.5 transgenic mouse embryos, enhancer activity in the face, among other domains[14].

Most of the genes surrounding this large region are expressed in the developing face (**Supplementary Fig. 4**), and their promoters show active chromatin marks (**Supplementary Table 3**). To identify MNE region–regulated gene(s), we measured by quantitative PCR (qPCR) the expression of these genes in dissected tissues from wild-type E11.5 embryos and those with deletion of the MNE region (**Fig. 2a**). In the face of del(8–17) embryos, *Myc* expression was reduced to 15% of its normal expression level. None of the other surrounding genes tested showed significant ($P < 0.05$) changes in expression. Notably, *Myc* levels in embryonic heart and liver were unaffected (**Fig. 2b**), showing that the 8–17 interval controls *Myc* expression only in tissues where it has enhancer activity. Whole-mount *in situ* hybridization confirmed strong reduction in *Myc* transcript levels in the face but not in the liver of del(8–17) homozygous embryos (**Fig. 2c**). The smaller deletion of the 8–14 interval, which retains the MNE region, led to a similar reduction in *Myc* expression in the developing face, whereas deletions that did not include the MNE region (**Fig. 2d**) or simple transposon insertions (**Supplementary Fig. 5**) had no or minor effects. From these experiments, we conclude that the MNE region controls *Myc* expression specifically in the future upper lip. If the MNE region regulates *Myc* expression in *trans*, both *Myc* alleles should be equally affected. Instead, in compound embryos carrying del(8–17) on one chromosome and a *GFP*-tagged *Myc* gene[15] on the other, *GFP* expression was not diminished (but was instead slightly higher), whereas the *Myc* allele on the same chromosome as del(8–17) was downregulated to 30% of wild-type levels, after normalization to *GFP* levels (**Fig. 2e**). The very different response of the two alleles demonstrates that the MNE region acts on *Myc* in *cis* across more than 1 Mb of DNA. The intermediate effects of heterozygous deletions of the MNE region on *Myc* expression also supported direct regulation in *cis*.

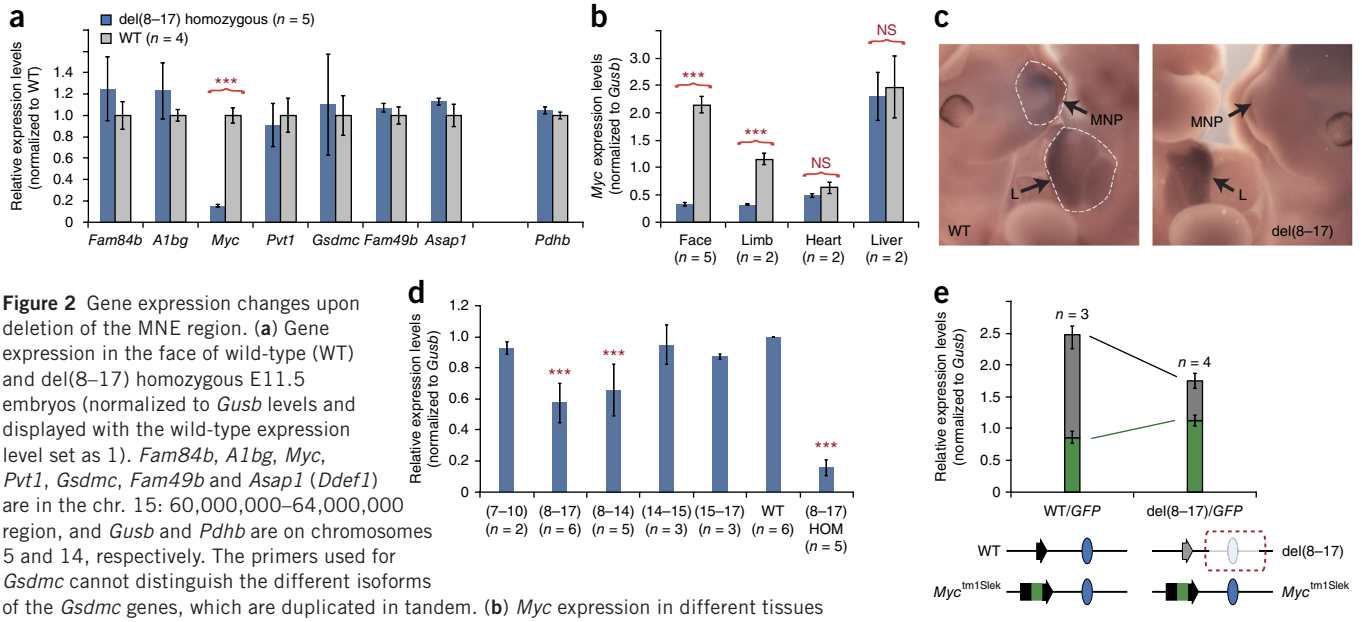


Figure 2 Gene expression changes upon deletion of the MNE region. (**a**) Gene expression in the face of wild-type (WT) and del(8–17) homozygous E11.5 embryos (normalized to *Gusb* levels and displayed with the wild-type expression level set as 1). *Fam84b*, *A1bg*, *Myc*, *Pvt1*, *Gsdmc*, *Fam49b* and *Asap1* (*Ddef1*) are in the chr. 15: 60,000,000–64,000,000 region, and *Gusb* and *Pdhb* are on chromosomes 5 and 14, respectively. The primers used for *Gsdmc* cannot distinguish the different isoforms of the *Gsdmc* genes, which are duplicated in tandem. (**b**) *Myc* expression in different tissues from E11.5 embryos. Values are expressed after normalization to *Gusb* levels (set as 1 in each tissue). (**c**) Loss of *Myc* expression detected by whole-mount *in situ* hybridization in the medionasal process (MNP) but not in the liver (L) of del(8–17) homozygous embryos relative to wild-type E11.5 embryos. Staining in control embryos is highlighted by dashed circles. (**d**) *Myc* expression in the face of E11.5 embryos heterozygous for different deletions. Values are expressed after normalization to *Gusb* levels (set as 1 in each tissue). (**e**) Allele-specific analysis of *Myc* expression in the faces of E11.5 embryos showed that the MNE region acts in *cis*. The reference allele $Myc^{tm1Slek}$ carries an insertion of *GFP* in the second exon of *Myc*[15], and its expression is detected with *GFP*-specific primers. Expression levels of the other alleles (wild type or del(8–17)) were determined by subtracting *GFP* expression levels from the overall *Myc* expression levels (the *Myc*-specific primers used amplified both alleles). *Myc* expression levels are significantly lower in del(8–17) than in wild-type mice (Student's *t* test, $P < 5.6 \times 10^{-5}$), whereas *GFP* expression is instead higher. In all charts, error bars represent s.d. ***$P < 0.005$, Student's *t* test; NS, not statistically significant ($P > 0.05$).

We then investigated the physiological consequences of the deletion of the MNE region on facial development. Mice homozygous for del(8–17) were born in normal mendelian proportions, without gross abnormalities or the several global skull malformations observed upon *Myc* deletion in the neural crest lineage[16]. Homozygous del(8–17) mice usually showed impaired postnatal growth, and some died before weaning, suggesting that the 8–17 interval contains additional elements[9] that may regulate *Myc* expression in other organs and tissues. Interestingly, compared to their littermates, del(8–17) homozygous mice had a significantly smaller snout ($P < 0.005$) (**Fig. 3** and **Supplementary Fig. 6a**), with an abnormal suture of the nasal and frontal bones (**Fig. 3b**). Other craniofacial regions were unaffected (**Fig. 3c**), and we did not observe palatal clefts in any of the pups. Changes in face morphology could already be detected in embryos, with a reduction in the width of the medionasal process in del(8–17) embryos compared to control E11.5 embryos (**Fig. 3d**). We did not detect any change in apoptosis in the developing medionasal prominence at E11.5 (data not shown). However, we found a small but significant reduction in the number of mitotic cells ($P < 0.05$) (**Fig. 3e** and **Supplementary Fig. 6b**), suggesting abnormal cell proliferation. Notably, we recovered a small number of E14–E15 embryos heterozygous and homozygous for del(8–17) with cleft lip and/or abnormal facial development (**Fig. 3f**): out of 121 embryos produced from heterozygous × heterozygous crosses, 4 del(8–17) embryos showed cleft lip and 4 other del(8–17) embryos showed cleft palate. The normal rate of spontaneous CL/P on the C57BL/6J background, on which our alleles are maintained, is extremely low[17], and none of the 35 wild-type embryos from these crosses had abnormal craniofacial morphology (Fisher's exact probability test $P = 0.0589$). Thus, these sporadic cases further support the involvement of this region in susceptibility to CL/P.

To identify genetic pathways that might contribute to the increased incidence of CL/P, we compared the mRNA transcriptomes of the medionasal regions in del(8–17) and control embryos at E11.5. We found 100 genes that showed significant (false discovery rate (FDR) < 0.05), albeit mild, differences in expression (**Fig. 4a**, **Supplementary Figs. 7** and **8**, and **Supplementary Table 4**). We tested and confirmed differences in expression for a subset of these genes using qPCR in del(8–17) and del(8–14) homozygous embryos. We also showed that the expression of these genes was not altered in del(14–17) embryos, in which the deleted interval does not include the MNE region (**Fig. 4b**). Consistent with our qPCR analysis, *Myc* expression was reduced to 17% of wild-type levels, and we found no significant changes for the genes located in the surrounding 5 Mb of sequence (**Supplementary Table 5**). Gene Ontology (GO) analysis showed that the majority of the genes misexpressed in del(8–17) embryos were involved in ribosome assembly and translational control (**Supplementary Tables 6** and **7**). For these genes, the fold change in expression was small (20–25%), but all showed lower expression levels in del(8–17) embryos and enrichment of the associated terms was maintained even when considering only highly expressed genes. Interestingly, MYC has already been shown to be a direct regulator of ribosome biogenesis in cancer and cellular models[18]. Our results stress the notion that physiological, non-oncogenic expression levels of *Myc* modulate ribosomal and translation-associated protein expression. We also found significant yet mild misexpression of a number of transcription factors and signaling molecules known to be important for facial development (**Supplementary Table 8**). Intriguingly, the altered frontonasal sutures in del(8–17) mice resembled the phenotype induced by facial-specific knockout of *Tfap2a*[19]. Whereas *Tfap2a* expression seemed unaltered in del(8–17) embryos, *Nr2f1*, whose product, together with TFAP2A, binds to a large subset of neural crest cell (NCC) enhancers[20], showed

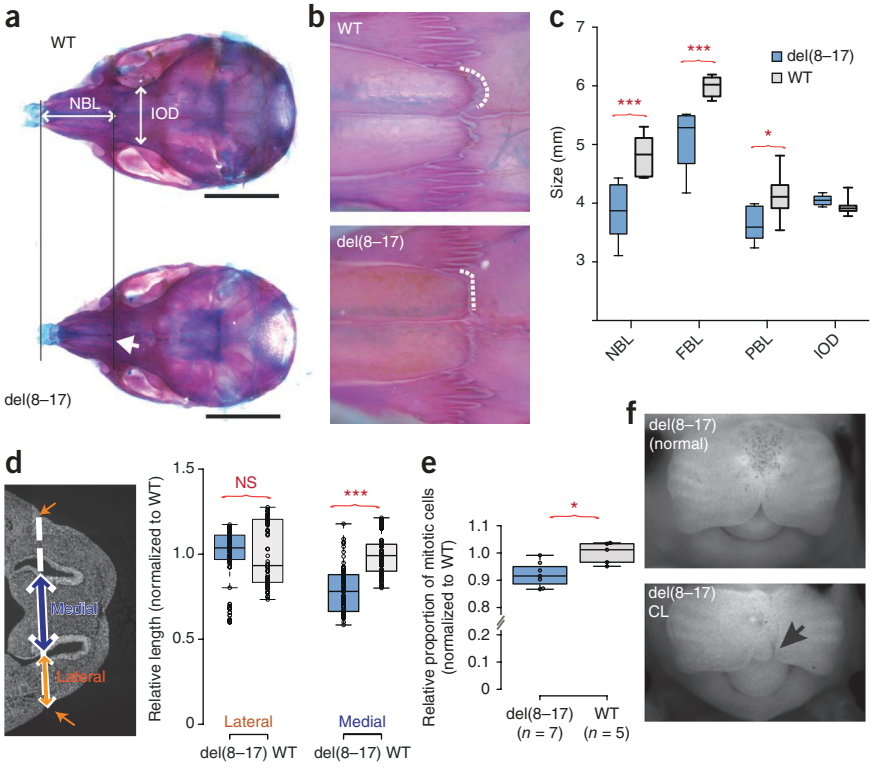


Figure 3 Facial dysmorphologies upon deletion of the CL/P-associated 8q24 region. (**a**) Comparison of the skulls of del(8–17) homozygous mice to those of their wild-type littermates. Dorsal views of representative skulls from 3-week-old wild-type and mutant (del(8–17)) mice, stained with Alcian blue (cartilage) and Alizarin red (bone). IOD, interorbital distance; NBL, nasal bone length. Scale bars, 5 mm. (**b**) Enlarged views of the frontonasal regions showing the abnormal suture in del(8–17) mice (highlighted by dashed lines). (**c**) Comparison of different bone lengths and skull measures (FBL, frontal bone length; PBL, parietal bone length) in 3-week-old (del(8–17), $n = 5$; wild type, $n = 7$) mice. Del(8–17) mice showed reduced nasal and frontal bone lengths (Student's t test, $P = 0.0033$ and 0.0028, respectively). Box plots show medians and first and third quartiles. Whiskers indicate minimum and maximum values. (**d**) Comparison of the widths of the lateral and medial parts of the developing face in wild-type ($n = 3$) and del(8–17) ($n = 4$) E11.5 embryos. The landmarks (arrows) used to compare the lateral and medial widths of the developing faces of E11.5 embryos are shown on the left. NS = not significant. (**e**) Quantification of proliferation in the faces of wild-type ($n = 5$) and (del(8–17) homozygous ($n = 7$) E11.5 embryos. Values represent the proportion of mitotic cells, determined by staining for phosphorylated histone H3, averaged from 10–30 serial sections per embryo. In **d** and **e**, box plots show medians and first and third quartiles. Whiskers indicate 1.5 times the interquartile ranges of the first and third quartiles. $^*P = 0.011$ (Student's t test). (**f**) Most E14.5 del(8–17) embryos have normal face morphology (top), but a minority have CL/P (arrow) with other craniofacial malformations (bottom). In all panels, statistical significance was determined with Student's t tests. $^{***}P < 0.005$, $^*P < 0.05$.

lower expression levels. TFAP2A also interacts both genetically and directly with MYC[21,22]. These interactions may contribute to propagating the consequences of *Myc* downregulation in the facial NCC gene regulatory network (GRN).

The distinct facial morphology of mice carrying deletion of the MNE region and the sporadic CL/P cases observed demonstrate the key functional role of this region in craniofacial morphogenesis. These findings support the implication of genetic variation at 8q24 in CL/P[3–5] and in normal variation in facial shape[23,24] in humans. They refine the proposed critical interval to a smaller region, distal to the marker most significantly associated with CL/P, rs987525 (ref. 3; **Supplementary Fig. 9**).

At the molecular level, we show that this region is a key tissue-specific *cis*-regulatory enhancer for *Myc*. Its extremely remote location further emphasizes the role of distant regulatory elements in controlling gene activity[14,25], not only for tissue-specific developmental regulators but also for genes with more general functions[26]. The MNE

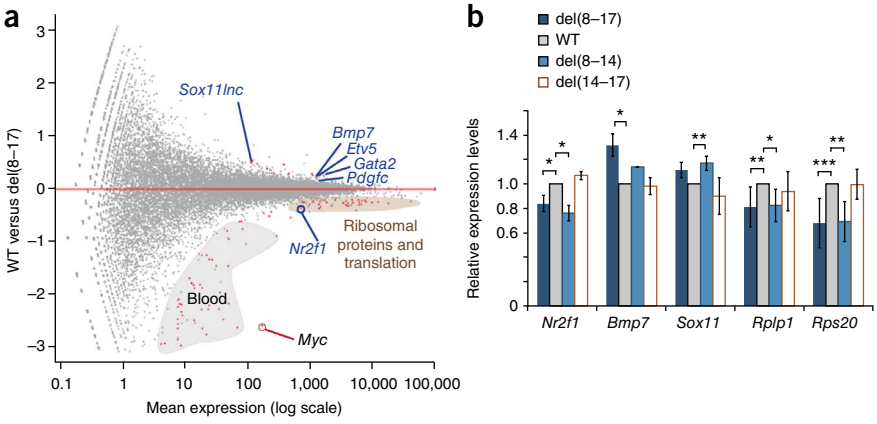


Figure 4 Alterations in gene expression upon deletion of the CL/P-associated 8q24 region. (**a**) Changes in expression measured by RNA sequencing in the medial faces of del(8–17) homozygous embryos (four replicate libraries from four different embryos) compared to wild-type controls (four replicate libraries from littermates). Genes with significant (FDR < 0.05) changes in expression are shown in red. See **Supplementary Table 4** for a list of the misregulated genes and **Supplementary Figure 7** for heat-map representations of the data. RNA sequencing analysis also showed downregulation of several blood-specific genes; their presence arose from the small blood vessels in the dissected tissues and suggested an additional role of the 8–17 interval in hematopoiesis, in which *Myc* has an important role[49]. Notably, the genomic region involved appears to be distinct from the MNE region, as the downregulation of *Apoe* and *Csf1* was observed in del(14–17) but not in del(8–14) embryos (**Supplementary Fig. 8**). (**b**) Validation by qPCR of expression changes for some of the genes identified by RNA sequencing. The levels of *Rplp1* and *Rps20* are significantly ($^{***}P < 0.005$, $^{**}P < 0.01$, Student's t test) decreased to similar extents in the medial faces of del(8–17) ($n = 4$ to 7) and del(8–14) ($n = 3$) homozygote E11.5 embryos but not in del(14–17) ($n = 3$) homozygote E11.5 embryos. *Nr2f1* levels also appeared lower ($^*P < 0.05$, Student's t test). Error bars, s.d.

region comprises several adjacent blocks carrying enhancer-associated epigenetic marks, which may act collectively to control *Myc* expression, as described for other loci[27–29]. In such situations, deletions of individual modules may show only limited or subtle effects[14,27] and may lead to an underestimation of their collective role. Indeed, it was recently proposed that GWAS may identify haplotypes with multiple polymorphisms affecting adjacent enhancers instead of a single 'master' element[30]. For this reason, larger rearrangements, such as the ones used here for 8q24 or earlier to characterize the 9p21 coronary artery disease (CAD)-associated interval[31], may be a useful strategy in following up GWAS findings, particularly for intervals with multiple candidate enhancer regions[32].

The different types of genes found to be misexpressed upon deletion of the MNE region give insights into the nature of the risk associated with variants in 8q24 (**Supplementary Fig. 10**). MYC is known for its role in NCCs[16,33], and our analyses suggest that the altered expression levels of downstream genes involved in facial development observed with deletion of the MNE region may account for the slight but penetrant concurrent facial dysmorphism. In this context, orofacial clefts may constitute the extreme end of a spectrum of normal phenotypic variation[23,34]. In addition to this hypothesis, we suggest that the global misexpression of ribosomal subunits and translation factors observed in del(8–17) mice may contribute to CL/P incidence through a distinct process. Heterozygous mutations of ribosomal genes, including *Rps7* and *Rps20*, whose expression was downregulated in del(8–17) mice, have phenotypic consequences in mice and humans[35–37]. Of note, humans with Diamond-Blackfan anemia, which is caused by mutations affecting diverse ribosomal proteins, are also prone to CL/P malformations[38]. Abundance of ribosomal proteins therefore constitutes a limiting factor: even mild, reduced expression of ribosomal and translation regulatory genes in del(8–17) mice may affect cellular metabolic capacity, leading to a general but spatially restricted sensitivity. This phenomenon may account for the influence of diverse metabolic and environmental conditions (for example, tobacco or alcohol use, diabetes or folate deficiency) that have been proposed to contribute to CL/P etiology[6] by transforming broad-spectrum but minor growth perturbations into growth defects affecting a specific embryonic process. The combination of these different effects might explain why this region has such a high population attributable risk[3] and a prominent influence on facial (dys)morphologies.

Interestingly, GWAS for other traits—prominently, susceptibility to diverse types of cancer—have identified associations with other regions within 8q24 (refs. 39–42), which were proposed to coincide with variants in tissue-specific enhancers for *Myc* or other genes[43–47]. As illustrated recently[48], validating or determining the role of these variants *in vivo* will be essential. The series of overlapping deletions and rearrangements described here opens up avenues to further dissect this important regulatory landscape.

URLs. TRACER, http://tracerdatabase.embl.de/; HTSeq-count, http://www-huber.embl.de/users/anders/HTSeq/doc/index.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) under accessions GSM1279691, GSM1279692 and GSM1279693. RNA sequencing data have been deposited in GEO under accessions GSM1279694–GSM1279701.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank B. Sleckman (Washington University, St. Louis) and A. Trumpp (DKFZ, Heidelberg) for providing the *Myc*^{tm1Slek} strain. We thank members of the EMBL Laboratory Animal Resources Facility for animal welfare and husbandry; the EMBL Genomics Core Facility for advice and support in processing ChIP and RNA sequencing experiments; and Genome Biology Computational Support for help with the analyses. We thank members of the Spitz laboratory and colleagues at EMBL for sharing reagents and helpful comments. V.V.U. and M.P. were supported by PhD fellowships from the Jeff Schell Darwin Trust and the EMBL International PhD program, respectively. This work was supported by EMBL.

AUTHOR CONTRIBUTIONS

F.S. designed the experiments. V.V.U., M.P., S.R. and K.L. performed the experiments. N.A.F. and J.C.M. performed RNA-seq data, bioinformatics and statistical analyses. V.V.U., M.P. and F.S. analyzed the data. F.S. wrote the manuscript with V.V.U., M.P. and J.C.M.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Dixon, M.J., Marazita, M.L., Beaty, T.H. & Murray, J.C. Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* **12**, 167–178 (2011).
2. Mangold, E., Ludwig, K.U. & Nöthen, M.M. Breakthroughs in the genetics of orofacial clefting. *Trends Mol. Med.* **17**, 725–733 (2011).
3. Birnbaum, S. *et al.* Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.* **41**, 473–477 (2009).
4. Beaty, T.H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*. *Nat. Genet.* **42**, 525–529 (2010).
5. Ludwig, K.U. *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat. Genet.* **44**, 968–971 (2012).
6. Murray, J.C. Gene/environment causes of cleft lip and/or palate. *Clin. Genet.* **61**, 248–256 (2002).
7. Freedman, M.L. *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.* **43**, 513–518 (2011).
8. Gritli-Linde, A. The etiopathogenesis of cleft lip and cleft palate: usefulness and caveats of mouse models. *Curr. Top. Dev. Biol.* **84**, 37–138 (2008).
9. Ruf, S. *et al.* Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379–386 (2011).
10. Jiang, R., Bush, J.O. & Lidral, A.C. Development of the upper lip: morphogenetic and molecular mechanisms. *Dev. Dyn.* **235**, 1152–1166 (2006).
11. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
12. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
13. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
14. Attanasio, C. *et al.* Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013).
15. Huang, C.-Y., Bredemeyer, A.L., Walker, L.M., Bassing, C.H. & Sleckman, B.P. Dynamic regulation of *c-Myc* proto-oncogene expression during lymphocyte development revealed by a *GFP–c-Myc* knock-in mouse. *Eur. J. Immunol.* **38**, 342–349 (2008).
16. Wei, K. *et al.* Neural crest cell deficiency of *c-myc* causes skull and hearing defects. *Genesis* **45**, 382–390 (2007).
17. Karolyi, J., Erickson, R.P. & Liu, S. Genetics of susceptibility to 6-aminonicotinamide–induced cleft palate in the mouse: studies in congenic and recombinant inbred strains. *Teratology* **37**, 283–287 (1988).
18. van Riggelen, J., Yetil, A. & Felsher, D.W. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* **10**, 301–309 (2010).
19. Nelson, D.K. & Williams, T. Frontonasal process–specific disruption of AP-2α results in postnatal midfacial hypoplasia, vascular anomalies, and nasal cavity defects. *Dev. Biol.* **267**, 72–92 (2004).
20. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).
21. de Crozé, N., Maczkowiak, F. & Monsoro-Burq, A.H. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proc. Natl. Acad. Sci. USA* **108**, 155–160 (2011).
22. Gaubatz, S. *et al.* Transcriptional activation by Myc is under negative control by the transcription factor AP-2. *EMBO J.* **14**, 1508–1519 (1995).

23. Boehringer, S. *et al.* Genetic determination of human facial morphology: links between cleft-lips and normal variation. *Eur. J. Hum. Genet.* **19**, 1192–1197 (2011).
24. Liu, F. *et al.* A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.* **8**, e1002932 (2012).
25. Visel, A., Rubin, E.M. & Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
26. Hallikas, O. *et al.* Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47–59 (2006).
27. Montavon, T. *et al.* A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
28. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
29. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
30. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).
31. Visel, A. *et al.* Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* **464**, 409–412 (2010).
32. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* **470**, 264–268 (2011).
33. Bellmeyer, A., Krase, J., Lindgren, J. & LaBonne, C. The protooncogene *c-Myc* is an essential regulator of neural crest formation in *Xenopus*. *Dev. Cell* **4**, 827–839 (2003).
34. Parsons, T.E. *et al.* Phenotypic variability and craniofacial dysmorphology: increased shape variance in a mouse model for cleft lip. *J. Anat.* **212**, 135–143 (2008).
35. Terzian, T. & Box, N. Genetics of ribosomal proteins: "curiouser and curiouser.". *PLoS Genet.* **9**, e1003300 (2013).
36. Watkins-Chow, D.E. *et al.* Mutation of the Diamond-Blackfan anemia gene *Rps7* in mouse results in morphological and neuroanatomical phenotypes. *PLoS Genet.* **9**, e1003094 (2013).
37. McGowan, K.A. *et al.* Ribosomal mutations cause p53-mediated dark skin and pleiotropic effects. *Nat. Genet.* **40**, 963–970 (2008).
38. Ito, E., Konno, Y., Toki, T. & Terui, K. Molecular pathogenesis in Diamond-Blackfan anemia. *Int. J. Hematol.* **92**, 413–418 (2010).
39. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
40. Tomlinson, I.P.M. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008).
41. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* **41**, 899–904 (2009).
42. Goode, E.L. *et al.* A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.* **42**, 874–879 (2010).
43. Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41**, 885–890 (2009).
44. Pomerantz, M.M. *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with *MYC* in colorectal cancer. *Nat. Genet.* **41**, 882–884 (2009).
45. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with *MYC*. *Proc. Natl. Acad. Sci. USA* **107**, 9742–9746 (2010).
46. Jia, L. *et al.* Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet.* **5**, e1000597 (2009).
47. Sotelo, J. *et al.* Long-range enhancers on 8q24 regulate *c-Myc*. *Proc. Natl. Acad. Sci. USA* **107**, 3001–3005 (2010).
48. Sur, I.K. *et al.* Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
49. Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).

ONLINE METHODS

Mouse alleles. Transposon insertion sites were generated and mapped as described previously[9]. Additional information on each insertion can be found on the TRACER website[50]. To distinguish the different *Myc* alleles, we used *Myc*^{tm1Slek} (ref. 15). The reference allele *Myc*^{tm1Slek} carries an insertion of the *GFP* gene in the second exon of *Myc*, and the encoded fusion protein appeared functional[15]. Yet, we noticed that the *GFP*-tagged allele was expressed at lower levels than the wild-type allele, suggesting that it might be a hypomorph. *Myc* is known to negatively regulate its own transcription through its promoter region[51], and this feedback can account for overall *Myc* levels in compound mice, on top of the specific effect of del(8–17) in *cis*.

Mouse alleles were genotyped by PCR of DNA purified from tail biopsies (for mice after birth) or yolk sacs (for embryos). Sequences for the primers for each insertion are provided in **Supplementary Table 9**. Deletions and duplications were generated by TAMERE[52] using the *Hprt*$^{tm1(cre)Mnn}$ line[53,54]. Rearrangements were further verified by long-range PCR (Expand Long-Range dNTPack, Roche), using primers flanking the reconstituted transposon. Mouse lines were maintained by backcrossing on a C57BL/6J background. Both male and female embryos were used for the experiments. Sample sizes were not predetermined.

Mouse experiments were conducted in accordance with the principles and guidelines in place at EMBL, as defined and overseen by its Institutional Animal Care and Use Committee, in accordance with European Convention 18/3/1986 and Directives 86/609/EEC and 2010/63/EU.

LacZ staining and *in situ* hybridization. LacZ staining and mRNA whole-mount *in situ* hybridization were performed as previously described[9]. Except where indicated, LacZ staining was performed on embryos heterozygous for the corresponding insertion, obtained by mating a transgenic male with a wild-type female.

Quantitative RT-PCR. Total RNA was purified from the dissected faces, forelimbs, hearts and livers of E11.5 embryos (45–48 somites) (RNeasy Mini kit, Qiagen) and eluted in 30–50 µl of nuclease-free water. For each RNA sample, we first measured concentration and purity with a Nanodrop ND-1000 spectrophotometer, and we then checked quality with a Bioanalyzer (Agilent Technologies) (**Supplementary Fig. 11a**). Possible DNA contamination was checked for by PCR using the 179039L and 179039R primers. Samples that did not meet quality control standards were discarded. First-strand cDNA was synthesized from 250–750 ng of total RNA using the Protoscript M-MuLV kit (NEB) with random hexamers. Quantitative RT-PCR was carried out in 0.1-ml and 0.2-ml wells of 96-well plates. We included 1 µl of cDNA, 1 µM primer mix and 10 µl of SYBR Green PCR Master Mix (Applied Biosystems) in a 20-µl final volume. Primers are listed in **Supplementary Table 10**. PCR was carried out on StepOnePlus Real-Time PCR systems (Life Technologies) and AB7500 Real-Time PCR systems (Applied Biosystems). Each primer pair was tested for efficiency and specificity following recommended procedures[55] (**Supplementary Fig. 11b**). We included *Gusb* and *Pdhb* as references. As *Gusb* expression levels (C_t values) were more in the range of those for the genes of interest, we used this gene for normalization, although the same significant differences were observed using normalization to *Pdhb* levels. Gene expression data were processed using Microsoft Excel and StepOne Real-Time PCR Software v2.0. For each group, two to five biological replicates were used (specific numbers are indicated in the figures), and, for each sample, quantification was performed on two to three technical replicates.

Skeletal preparation and analysis. Skulls from postnatal day (P) 19–21 and P40 mice were prepared following a standard procedure[56] and stained with 0.3% Alcian blue or 0.1% Alizarin red. For morphological measurements, ten landmarks from the top view of the skull and six landmarks from the bottom view were used[57]. Measurements were compared between age-matched groups. No mice were excluded from the analyses. Comparisons between the del(8–17) homozygous mice and wild-type controls were carried out with Student's *t* tests, and the statistical variance was calculated by *f* test. Exclusion of any of the samples did not have a critical impact on the statistical significance of the difference between genotypes.

Cell proliferation assays. Antibody to phosphorylated histone H3 (Ser10; 06-570, Millipore, rabbit polyclonal) was used at a 1:200 dilution in 10% FCS, 0.2% PBS with Tween-20 (PBST) blocking solution on 5-µm paraffin sections prepared from E11.5 embryos. Antigen retrieval was carried out in citrate buffer, pH 6.0 (Dako), and an Alexa Fluor 594–conjugated anti-rabbit secondary antibody was used at a 1:500 dilution with 1× 4′,6-diamidino-2-phenylindole (DAPI) in the blocking solution. Images were acquired at 10× and 20× magnification with a Zeiss CellObserver HS Automated wide-field microscope. Four to 10 images were stitched together to obtain a full E11.5 embryonic face section. Image analysis was carried out with Fiji software[58], using an automated macro to minimize subjective errors in counting nuclei positive for phosphorylation of histone H3. In brief, for each section, phosphorylated histone H3 signal was captured with the DsRed filter, and the signal obtained with the GFP-filtered image was subtracted out to remove background autofluorescence. The resulting signal image (S) was subjected to background rolling, Gaussian blur and Renyi Entropy dark thresholding. To define parameters for the automated counting of particles positive for phosphorylated histone H3 in 10× and 20× wide-field images, mitosis-phase particles were counted manually on 63× magnification images of the two sections that covered the whole face, obtained by stitching images recorded with a Spinning-Disk Microscope (PE Ultraview VoX) and using Volocity software (PerkinElmer). Particle counting was then performed and overlaid on the DAPI image: particle signals that were too elongated or did not overlap with DAPI signal were eliminated manually to obtain the final number of particles positive for phosphorylated histone H3 (pSc). The average number of pSc particles for a wild-type sample was 291. Cell numbers were estimated by measuring DAPI area (aD) on each section, as current Fiji plug-ins for DAPI signal counting were computationally inefficient for large samples. We verified that DAPI area and cell count performed with the Cell Counter plug-in (I. Levenfus, TU Dresden) showed a linear correlation of >99.5% in the range of 20 to 8,000 cells. Data were collected for sections from seven del(8–17) homozygous embryos and five wild-type littermate controls. The proliferation rate was defined as pSc/aD for each section. On average, proliferation rate was higher in the 130 sections from the 5 wild-type embryos than in the 165 sections from the 7 del(8–17) homozygous embryos (Student's *t* test, $P < 1 \times 10^{-5}$) (**Supplementary Fig. 6b**). The average proliferation rate was also calculated for each embryo (as the average of the values measured for all the sections from one embryo). Again, the seven del(8–17) homozygous embryos showed lower values than the five controls (Student's *t* test, $P < 0.05$) (**Fig. 3e**).

ChIP-seq experiments. For a given ChIP experiment, 25 facial mesenchyme samples were collected from E11.5 embryos, pooled, minced and fixed for 10 min in 1% formaldehyde in PBS. Formaldehyde was quenched through incubation with 250 mM glycine for 1 min on ice, and cells were lysed in buffer A (10 mM HEPES, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100 and 1:100 freshly added PMSF) for 10 min at 4 °C on a rotating wheel. Cells were then incubated with buffer B (10 mM HEPES, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100 and 1:100 freshly added PMSF) for 10 min at 4 °C on a rotating wheel. Cells were resuspended in 300 µl of sonication buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 0.1% SDS and 1:100 freshly added PMSF) and sonicated with a Bioruptor for 24 min, with a duty cycle of 0.5 on the high-power setting. An aliquot of 2% of chromatin was taken as the input control. Chromatin was then equilibrated to 300 µl of RIPA buffer and immunoprecipitated with 1.2 µg of antibody to H3K27ac (Abcam, ab4729) or 0.75 µg of antibody to H3K4me1 (Abcam, ab8580) with rotation overnight at 4 °C. Protein A Dynabeads (10001D, Life Technologies) were added (20 µl per tube), incubated for 5 h in rotation and washed four times in RIPA buffer and 1 time in Tris-EDTA. Immunoprecipitated complexes were eluted in 100 µl of elution buffer, and cross-links were reversed at 65 °C overnight. Afterward, 20 µg of proteinase K was added to each tube, and samples were incubated for 2 h in 55 °C. Finally, DNA was purified with QIAquick columns.

Libraries were prepared with NEB Next ChIP-seq Sample Prep Master Mix Set I (E6240S, NEB) and NEB Next Multiplex Oligos for Illumina (E7335S, NEB) and sequenced with a depth of 100 million reads for the input samples and 30 million reads for the ChIP samples.

RNA sequencing. The facial tissues of four del(8–17) homozygous and four wild-type E11.5 littermate embryos were dissected and treated independently as biological replicates. *Drosophila melanogaster* S2 cells (72,000) were spiked into each tube to detect any global perturbation in mRNA levels. Total RNA was isolated with an RNeasy Mini kit (Qiagen). RNA quality and quantity were measured with a 2100 Bioanalyzer (Agilent Technologies). RNA samples were prepared according to the TruSeq RNA sample preparation guide (Illumina). We performed 50-bp, single-end sequencing on an Illumina HiSeq instrument. The eight samples were barcoded and run on the same flow cell. RNA sequencing data were processed with the iRAP (version 0.2.0; N.A.F., R. Petryszak, J.C.M. and A. Brazma, unpublished data) analysis pipeline (with default settings). More specifically, reads were mapped to the mouse reference genome (GRCm38.72) using TopHat[59] (version 1.4.1) with the default number of mismatches. Read counts for each gene were estimated by running HTSeq-count (version 0.5.3.p9) with intersection non-empty mode. Data normalization and identification of genes with significantly different expression levels were performed using DESeq (version 1.10.1)[60]. *P* values were adjusted for multiple testing using the Benjamini-Hochberg procedure, with a corrected *P* value of 0.05 being used to identify differentially expressed genes. The ratio of reads mapped to the mouse genome and to the *Drosophila* genome was similar for all samples, suggesting the absence, in this context and for this range of *Myc* expression levels, of a global amplifier effect that would change the expression levels of most genes[61,62].

50. Chen, C.-K. *et al.* TRACER: a resource to study the regulatory architecture of the mouse genome. *BMC Genomics* **14**, 215 (2013).
51. Facchini, L.M., Chen, S., Marhin, W.W., Lear, J.N. & Penn, L.Z. The Myc negative autoregulation mechanism requires Myc-Max association and involves the *c-myc* P2 minimal promoter. *Mol. Cell. Biol.* **17**, 100–114 (1997).
52. Hérault, Y., Rassoulzadegan, M., Cuzin, F. & Duboule, D. Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat. Genet.* **20**, 381–384 (1998).
53. Tang, S.-H.E., Silva, F.J., Tsark, W.M.K. & Mann, J.R. A Cre/*lox*P-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis* **32**, 199–202 (2002).
54. Wu, S., Ying, G., Wu, Q. & Capecchi, M.R. Toward simpler and faster genome-wide mutagenesis in mice. *Nat. Genet.* **9**, 922–930 (2007).
55. Bustin, S.A. *et al.* The need for transparency and good practices in the qPCR literature. *Nat. Methods* **10**, 1063–1067 (2013).
56. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R.R. *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2003).
57. Kawakami, M. & Yamamura, K.-I. Cranial bone morphometric study among mouse strains. *BMC Evol. Biol.* **8**, 73 (2008).
58. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
59. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
60. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
61. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151**, 68–79 (2012).
62. Lin, C.Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).

doi:10.1038/ng.2971