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
Tuğçe Aktaş, B.Sc.
Born in Ankara, Türkiye

Date of oral examination: 26.10.2011

MusD transposable elements and their impact on endogenous gene regulation

Referees:

Dr. Eileen Furlong
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg
Tel.: +49(0)6221 387- 8416

Prof. Dr. Joachim Wittbrodt
Dept. of Molecular Developmental Biology & Physiology
Centre for Organismal Studies
Im Neuenheimer Feld 230, 69120 Heidelberg
Tel: +49(0)6221 54-6499

Erişir menzil-i maksuduna âheste giden
*Tiz-reftâr olanın pâyine dâmen dolaşır**

Hatemi İbrahim Bey

*The slow proceeding reaches to the intended target,
The skirt becomes entangled to the feet of the rushing.

Table of Contents

1. Summary

1.1 Summary in English

Transposable elements are traditionally perceived as “junk DNA”, but they are a major evolutionary force in shaping genes and genomes. In this study, I investigated the role of MusD endogenous retroviruses as controlling elements in the loci they are inserted in and a novel mechanism through which the host genome could restrict their ability to change gene expression patterns.

De novo insertions of two MusD elements into the *Fgf8* locus were found to be the cause of *Dactylaplasia* mutations. When MusDs are situated between the regulatory elements driving *Fgf8* in the limb apical ectodermal ridge (AER) and its promoter, they act as enhancer blockers and down-regulate *Fgf8* expression in this domain. Concomitantly, they also hijack these enhancers to drive their own expression in the limbs. We propose that MusD’s enhancer-blocking activity concurrent with its transcription has a broader impact on the locus by re-routing some *Fgf8* enhancers to new target genes, giving rise to ectopic expression of these genes in lieu of *Fgf8*. This model could also account for the phenotypically related but genomically distinct *SHFM3* condition, where large duplications in the orthologous human locus cause a limb malformation similar to *Dactylaplasia*. They are moving a subset of *Fgf8*-enhancers away from this gene, a situation that possibly makes them accessible to other genes. Using mouse chromosomal engineering, we showed indeed that such structural changes in the locus are leading to ectopic expression of genes in *Fgf8* expression domains. We tested two candidate genes from the locus to assess if their ectopic expression could phenocopy the disease. While we did not manage to reproduce the ectrodactyly phenotype, ectopic-expression of *Lbx1* in the AER led to preaxial polydactylies, a feature also observed in several *SHFM3* patients (specific for this form of ectrodactly). Thus, it is possible that this already complex disorder involves the combined action of multiple ectopically expressed genes from the locus.

Therefore, gene expression programs in this *SHFM3/Dactylaplasia* locus seem to be altered in related ways for genomically distinct human and mouse mutations. To further understand how MusD could exert this effect, I examined the enhancer blocking activity of MusDs using *ex-vivo* assays. These experiments identified several regions within MusD, which have insulator properties as strong as

the prototypic HS4 region from the chicken beta-globin locus. This is providing strong evidence that MusDs could interfere with the expression of endogenous genes, by working as a mobile insulator element. Indeed, I showed that of a MusD found between the co-regulated *Olig2* and *Olig1* genes led to changes in their overall and relative expression levels. These changes and the effects of MusD in *Dactylaplasia* mice were intriguingly only observed if these elements were unmethylated. When their 5' LTRs were epigenetically repressed, gene expression levels were similar to wild type (i.e. in the absence of MusD) and no phenotype was observed.

These observations suggested that MusD expression and effects were depending on epigenetic control over this element. Importantly, the epigenetic status of MusD appeared to be strictly dependent on the presence of an additional locus, *Mdac*, which is polymorphic amongst mouse strains. The resistant strain had completely methylated MusD 5'LTRs in contrast to almost complete lack of methylation in the permissive strain, hence the origin of the *Mdac* allele determined the cytosine methylation levels of MusD 5'LTR. This effect is limited to a few of MusD elements, as many of them are in heterochromatin regions. Nevertheless, Mdac seems to be a general controlling factor of MusD, since the strain-specific, differential methylation is consistent in all tested tissues independent of its expression. MusD elements have almost identical LTRs with ETnII elements. However, *Mdac* did not affect the 5'LTR methylation status of ETnIIs, indicating that Mdac is acting specifically on MusD elements.

I genetically mapped the *Mdac* locus to a small interval of 1.3-1.7 Mb. Interestingly, this region is structurally variable between resistant and permissive strains, with the permissive strains carrying deletion of a cluster of KRAB-ZFP genes and pseudogenes present in the resistant strains. KRAB-ZFPs are good candidates for *Mdac,* as their zinc finger domains provide a modular sequence specific binding and their KRAB domain recruits repressing chromatin modifiers. Supporting this identification, we found that a BAC that partially covers the mapped region and contains one KRAB-ZFP from the resistant strain could lead to MusD repression when added to ES cells from a permissive strain. Furthermore, we showed that the deletion of KAP1 in resistant strains led to up-regulation of MusD, along side with other elements. Our findings argue that KRAB-ZFP play major role in counteracting ERVs and that a specific KRAB-ZFP from the *Mdac* region is targeting repressive modification to MusD elements.

1.2 Deutsche Zusammenfassung

Transposons werden traditionell als “junk DNA” aufgefasst. Dennoch stellen sie eine bedeutende evolutionäre Kraft da, die Gene und Genome gestaltet. In dieser Arbeit habe ich die regulativen Auswirkungen von Insertionen des MusD endogenen Retrovirus in einen Lokus sowie einen neuen Mechanismus untersucht, der Wirtsgenomen erlaubt diese Auswirkungen zu unterdruecken.

Zwei unabhängige *de novo* MusD-Insertionen im *Fgf8* Lokus sind ursächlich für *Dactylaplasia* in Mäusen. Diese MusDs befinden sich zwischen den regulatorischen Elementen, die die *Fgf8* Expression in der apikalen ektodermalen Randleiste (AER) steuern, und dessen Promoter. Hierbei blockieren sie die Enhancer-Promoter Interaktion, welches letztendlich zu einer reduzierten *Fgf8* Expression in dieser Domäne führt. Parallel dazu nutzten die MusDs diese Enhancer um ihre eigene Expression in den Extremitäten zu steuern. Wir stellen die Hypothese auf, dass diese Vorgänge umfassendere Auswirkungen auf den Lokus haben können und beispielsweise einige *Fgf8* Enhancer auf neue Zielgene wirken, welche daraufhin ektop in den Extremitäten exprimimiert werden. Dieses Modell könnte auch das Auftreten der phänotypisch ähnlichen, doch genomisch unschiedlichen Missbildung SHFM3 erklären, der große Duplikationen im orthologen, humanen Lokus zu Grunde liegen. Diese trennen eine Reihe von *Fgf8* Enhancern von ihrem Zielgen und machen sie dadurch verfügbar für andere Gene. Tatsächlich konnten wir mittels „mouse chromosomal engineering“ zeigen, dass solche strukturellen Veränderungen des Lokus zur ektopen Expression von Genen in den *Fgf8* Domänen führen können. Daraufhin haben wir zwei Kandidatengene in diesen Bereichen ektop exprimiert, um ihr Potential den Phänotyp auszulösen zu ermitteln. Obwohl wir nicht in der Lage waren den Ektrodaktyly-Phänotyp zu reproduzieren, führte die Überexpression von *Lbx1* in der AER zu preaxialer Polydactyly. Dieser Phänotyp wurde ebenfalls in mehreren SHFM3 Patienten beobachtet, wobei er spezifisch für diese Form der Ektrodaktyly ist. Daher ist es möglich, dass diese Entwicklungsstörung den kombinierten Effekt von einigen ektop exprimierten Genen beinhaltet.

Die Genexpression des *SHFM3/Dactylaplasia* Lokus scheint also in ähnlicher Weise durch unterschiedliche Mutationen in Mensch und Maus verändert zu sein. Um besser verstehen zu können, wie ein MusD Element diesen Effekt auslösen kann, habe ich das Potential von MusD, Enhancer zu blockieren, mittels *ex vivo* Studien untersucht. Durch diese Experimente wurden mehrere Abschnitte innerhalb von MusD bestimmt, die ähnlich starke Insulatoreigenschaften wie bei der

prototypischen HS4 Region des Huhn-Betaglobin Lokus aufwiesen. Dies gibt einen deutlichen Hinweis darauf, dass MusD Elemente die Expression von endogenen Genen stören können. Tatsächlich konnte ich zeigen, dass ein endogenes MusD Element zwischen den ko-regulierten Genen Olig2 und Olig1 zu einer Veränderung der absoluten und relativen Expression dieser Gene führt. Interessanterweise waren die Veränderungen in *Dactylaplasia* Mäusen nur zu beobachten, wenn die MusDs unmethyliert vorlagen. Waren hingegen ihre 5' LTRs epigenetisch reprimiert, waren auch die Genexpressionsstärken dem Wildtyp ähnlich und kein Phänotyp trat auf.

Diese Beobachtungen wiesen darauf hin, dass die Expression von MusD, wie auch die weiteren genannten Effekte, von der epigenetischen Kontrolle über das Element abhängen. Dabei ist wichtig zu erwähnen, dass dieser epigentische Zustand von einem weiteren Lokus, *Mdac*, abhängt, der in Mauslinien polymorph vorliegt. Die resistente und phänotypisch normal Mauslinie zeigt eine vollständige Methylierung der MusD 5' LTRs, welche hingegen bei der permissiven Linie fast völlig demethyliert vorliegen. Diese Wirkung des *Mdac* Allels ist auf wenige MusD Elemente begrenzt, da die meisten in Heterochromatin liegen. Gleichwohl scheint Mdac ein universeller Kontrollfaktor für MusD zu sein, da die Linien spezifische, differentielle Methylierung durchgängig in allen getesteten, exprimierenden wie auch nicht exprimierenden, Geweben auftrat. MusDs besitzen nahezu identische LTRs wie die verwandten EtnII Elemente. Allerdings beeinflusst Mdac nicht den Methylierungsstatus dieser Elemente und scheint daher MusD spezifisch zu sein.

Ich habe den *Mdac* Lokus auf ein kleines Intervall von 1,3 – 1,7 Mb genetisch kartiert. Interessanterweise ist diese Region zwischen resistenten und permissiven Mauslinien strukturell variabel, wobei die permissiven Linien eine Deletion eines KRAB-ZFP Gen-Clusters und einiger Pseudogene aufweisen. Diese KRAB-ZFPs sind gute Kandidatengene für *Mdac*, da ihre Zinkfingerdomänen DNA Sequenz-spezifisch binden und ihre KRAB-Domänen Chromatin modifizierende Proteine rekrutieren. Folgerichtig weisen ES Zellen von einer permissiven Mauslinie eine Hemmung von MusD auf, wenn sie mit einem BAC aus der kartierten Region mit einem KRAB-ZFP der resistenten Linie injiziert wurden. Des Weiteren konnten wir zeigen, dass die Deletion von KAP1 in resistenten Mauslinien zu einer gesteigerten Expression von MusD und weiteren retroviralen Elementen führt. Zusammengenommen zeigen unsere Resultate, dass KRAB-ZFP Proteine eine wichtige Rolle bei dem Schutz gegen ERVs spielt und dass ein spezifisches KRAB-ZFP aus der *Mdac* Region MusD Elemente durch DNA-Modifikation reprimiert.

2. Introduction

Developmental processes require the harmonized activation of multiple genes in a strictly controlled manner. This coordination is achieved at the level of transcription through the action of *cis*-regulatory elements such as enhancers. In vertebrates these elements are frequently not in close proximity to the promoters of the genes they control. Thus, modulated interactions between the remote elements and their targets contribute to the precise regulation of gene transcription. The specificity of such interactions is often critical, particularly for developmental genes, as shown in multiple mutations that interfere with *cis*-regulatory function (reviewed in (Kleinjan and Lettice 2008)). Genome sequencing projects have revealed that a large part of the genome is composed, not only of these elements with clear functional roles, but also of repeated sequences, to a very large extent (more than 40% in mammals) originating from more or less recent transposable elements. While these mobile elements are considered to be "junk" or even harmful remnants of their proviral ancestors, we now have a different understanding of TEs in shaping genomes as initiated by the studies of Barbara McClintock and followed by other studies showing TEs as "controlling elements". As I will discuss here, their role and notably in gene regulation, is far from being negligible.

Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and Insulators

There are several crucial steps at which gene expression is controlled from the production of an mRNA to its maturation, export outside the nucleus and translation into proteins. Among these steps, transcription plays an essential role in defining the spatial, temporal and quantitative distribution of gene products. The production of mRNA transcripts is the result of the action of several genomic regulatory elements with distinct functions, and that are distributed around the gene they act on. The main regions involved in transcription and transcriptional regulation could be distinguished as: the core promoter, proximal promoter elements, enhancers, insulators and silencers (as shown in Figure 1).

1. The promoter: core promoter and promoter proximal region: Classically, the core promoter is described as the region around the transcription start site of a gene, which contains several DNA elements that facilitate the binding of regulatory proteins (J. Courey 2008). These binding events are associated with the recruitment of the transcription initiation complex and thus the

initiation of the transcription. Canonical core promoter elements include the TATA box, the INR, DPE and BRE (Smale and Kadonaga 2003), but the number of core promoter elements is increasing with the advancements in sequencing technologies (such as the pair-end sequencing strategy in (Ni et al. 2010)) that provide precise and comprehensive maps of transcription initiation sites within the genome. Lately, core promoters are classified into two groups as; those that have a single transcription start site and those that have a cluster of transcription sites over a broad region (reviewed in (Sandelin et al. 2007)). The first group of promoters, which are also known as "focused" promoters, constitute the majority of core promoters in eukaryotes. Interestingly, in vertebrates only one-third of the core promoters belong to this class and these promoters are associated with regulated, tissue-specific genes (Juven-Gershon et al. 2008). In contrast, the latter group, also referred to as "dispersed" promoters, are the minority in eukaryotes, but comprises the majority of genes in higher eukaryotes and drive the expression of most of the house-keeping genes which are more or less uniformly expressed throughout development (Carninci et al. 2006). Other promoter elements, which are located mostly 100-500bp upstream of the core promoter constitute proximal promoter elements (or upstream promoter elements). Examples of such elements include the GC box to which the transcription factor Sp1 binds and the CAAT box which is recognized by NF1 (Kadonaga and Tjian 1986), (Jones et al. 1987). Binding of these transcription factors to promoters that contain these elements boosts transcription by enhancing the recruitment of RNA polymerase II to nearby core promoter (Jones, Kadonaga et al. 1987). Abundant transcription factors such as Sp1 can work co-operatively with tissue specific transcription and/or other nuclear factors that also bind to the promoter proximal region (Safe and Kim 2004). In addition to these functions, some elements were proposed to function as tethering elements for active distant regulatory elements (enhancers) and allowing them to interact with the core promoter (Calhoun et al. 2002).

2. Enhancers: These distal regulatory elements can be found up- or downstream, and at considerable distances. They can be within the intron of the gene that they control or in those of the surrounding genes (Levine and Tjian 2003). Recent examples have also shown that coding exons could have enhancer functions (Tümpel et al. 2008). A gene regulatory region is often

composed of several physically distant enhancers which act in a modular manner: each module can carry out some aspect of the gene expression pattern in a specific cell-type or at a given stage during development (Blackwood and Kadonaga 1998). However, this modular aspect of regulation is also marked by some redundancy with the presence of the so-called called "shadow" enhancers. These elements are proposed to ensure precise gene expression patterns during embryogenesis (Hong et al. 2008). Such remote enhancers may have important functions in controlling gene expression as shown by their contribution to phenotypic robustness in response to environmental changes (Frankel et al. 2010). Establishing functional interactions between an enhancer and its target promoter is crucial for gene expression. Enhancers, which are typically kilobases away from the genes they influence but which could be found at much considerable distances up to several hundreds of kilobases (Lettice et al. 2003) are physically brought close to the relevant promoters for these interactions by looping of the chromatin structure. It has been shown by DNA-FISH (Osborne et al. 2004), (Jhunjhunwala et al. 2008) or by chromatin conformation capture (3C) method (Dekker et al. 2002) and later by other follow-up methods (4C, Hi-C ((Simonis et al. 2006), (Lieberman-Aiden et al. 2009))).

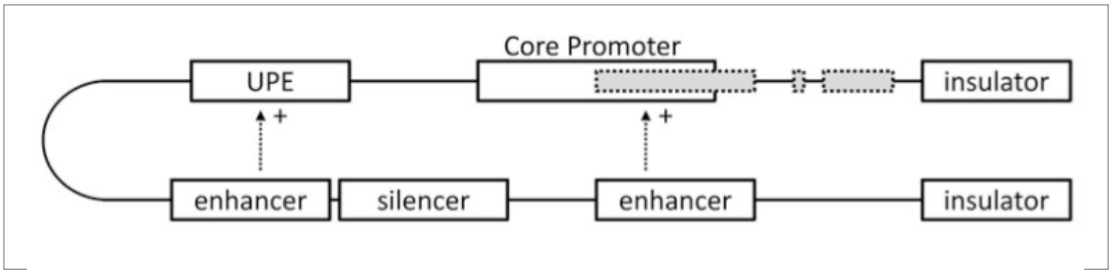


Figure 1. Mammalian transcriptional unit (taken from (Riethoven 2010))
This complex arrangement of the mammalian transcriptional unit includes a core promoter overlaps with the first exon and upstream promoter elements (UPE) further upstream that are brought close to enhancer elements by DNA looping

3. Specificity of enhancer interactions: Given the large number of genes and enhancers present within a genomic locus, the specificity of interactions should be ensured. Two major mechanisms have been suggested to act together to achieve the enhancer-promoter assembly in a specific manner. First, factors that bind both enhancers and promoters effectively could bring them together and lead to the transcription of a gene of interest. Secondly,

there must be other elements that block unintended interactions between an enhancer and the promoter of a gene that should not be expressed at the given place and time (reviewed in (Blackwood and Kadonaga 1998)). Interestingly, many enhancers are able interact with heterologous promoters when they are taken out of context with a few exceptions. One such exception is the distinct preference of interaction of an enhancer with the core promoters that contain a particular element relative to the ones which do not contain it (Ohtsuki et al. 1998), (Butler and Kadonaga 2001), (Juven-Gershon et al. 2008). Furthermore, specific enhancer-promoter interactions depend on the homotypic relations of common proteins bound to both the enhancer and promoter-proximal DNA, such as Sp1 mediated formation of DNA loops (Mastrangelo et al. 1991) or GATA binding within the distal locus-control region (LCR) as well as promoter-proximal regions of globin genes (Gong and Dean 1993).

Blocking the enhancer-promoter interactions or shutting down the promoter of a surrounding gene can also prevent inappropriate interactions. The first manner could be achieved by insulator elements with enhancer-blocking activity. The enhancer-blocking insulators interfere with the enhancer-promoter interactions and prevent gene activation, when situated between the enhancer and the promoter. Some insulators have as well a chromatin barrier activity. These elements lie in the border of heterochromatin and euchromatin domains and prevent the spreading of heterochromatin (reviewed in (Gaszner and Felsenfeld 2006)). Both activities contribute to make such an element a Locus-Control Region, functioning in a manner which is independent of the surrounding sequence, as described in the beta globin locus (reviewed in (Mahajan et al. 2007)). The activity of the insulator sequence in the globin locus is dependent on the binding of the Zinc-finger protein, CCCTC-binding factor (CTCF) (Bell et al. 1999). Subsequently, most of the insulators identified so far bind to CTCF. Interestingly, CTCF is ubiquitously expressed and ChIP-seq experiments showed that it binds to its target sequences largely in a cell type invariant manner (Kim et al. 2007). However, some other insulator sequences have more complex features and notably has been shown to act in a polar manner. A chromosomal inversion within the HoxD cluster involving such a polar element resulted in a reciprocal re-assignment of specific regulation for two tissues (Kmita et al. 2000).

Insulators are not the only defined elements that are capable of preventing the promoter activation. Silencers are DNA sequences generally located in the 5' upstream region of the promoter of a gene, which recruit transcription factors and interfere with the initiation of transcription in a position independent fashion. The position dependent silencers (also called as negative regulatory elements, NREs) passively prevent the binding of transcription factors to promoters and are not only specific to the upstream region (reviewed in (Ogbourne and Antalis 1998)). NRSF/REST was one of the first transcriptional regulators found to negatively regulate gene expression during vertebrate development. It was initially identified as a silencer binding protein that is controlling neuronal differentiation (Lunyak et al. 2002). Transcription factor binding mostly plays a transient role in gene expression regulation therefore they should be continuously supplied. The polycomb group (Pc-G) proteins seem to have overcome this requirements by initiating a complex assembly that occupies target genes stably and keeps them heritably silenced via histone modifications (Schwartz and Pirrotta 2007).

Altogether, these DNA elements that constitute most of the genome are essential for the regulation of a relatively smaller but obviously very important portion: protein coding genes. Furthermore, the physiological and behavioral complexity that we observe in living organisms is correlated with the wide range of gene expression patterns, which an organism might acquire by the expansion of *cis*-regulatory regions, underscoring the importance of these elements (Levine and Tjian 2003).

The repeated genome

The repeated portion of eukaryotic genomes mostly consists of various classes of **T**ransposable **E**lements (TEs). About half of the human genome is composed of such elements (Lander et al. 2001). Although the origin of TEs is obscure, they most likely evolved from a group of proviruses that entered the germ lines or germ cell progenitors of their respective animal hosts.

Eukaryotic transposable elements are usually classified into two main groups depending on their mechanism of transposition and both classes of elements have autonomous and non-autonomous members. The first class consists of elements, called retroelements, which transpose through the re-insertion of an RNA intermediate transcribed from the transposon. These elements are further classified based on their sequence composition (**L**ong **T**erminal **R**epeat containing (LTR) or not

containing (non-LTR)). The LTR is the control center for gene expression as it contains enhancer, promoter, transcription initiation (capping), transcription terminator and poly-adenylation signal. Complete copies of non-LTR transposons, such as mammalian **L**ong **I**nterspersed **N**uclear **E**lements (LINE), encode proteins needed for reverse transcription, but defective transposons lacking proteins necessary for transposition are often abundant. A good example of such element is provided by the mammalian **S**hort **I**nterspersed **N**uclear **E**lements (SINE) which do not encode proteins needed for reverse transcription. These non-autonomous elements need the action of LINEs-proteins in *trans* to complete their reproduction cycle (Dewannieux and Heidmann 2005). *Drosophila copia* and *gypsy* elements and three superfamilies of **e**ndogenous **r**etro**v**iruses (ERV-1,2,3) can be given as examples of LTR repeats. The second class of elements transpose themselves into new locations directly through their DNA copies. These elements are further split into three groups depending on their replication strategy: 1) cut and paste, 2) rolling circle and 3) self-synthesizing. Cut-and-paste transposons are mobilized by a transposase (which determines the superfamily they belong to) that binds to the ends of the transposon sequence and introduces nicks to DNA for excision. Usually, the transposase is encoded by the transposons, but defective elements can be transposed by other elements' transposases produced by other members of the same family. Helitrons transpose via replicative rolling-circle transposition and do not generate a target site duplication. Finally, polintons propagate through protein primed self-synthesis and are the most complex (with their potential to code ten proteins) of all the known eukaryotic transposable elements up to date. (summarized in Table 1, information gathered from (Charlesworth et al. 1994), (Jurka et al. 2007), (Kapitonov and Jurka 2008)) (more specifically for human and mouse transposons see Table 2)

TEs are often considered to be "selfish" elements that use part of the host system to maintain their self replication ability and be transmitted through generations without a significant contribution to the host phenotype. Hickey called such a DNA sequence a "sexually-transmitted nuclear parasite" (Hickey 1982). These parasites are not innocuous: on the contrary, there is a reverse correlation between an increase in their copy number and the fitness of the host (Orgel and Crick 1980), (Doolittle and Sapienza 1980). The high abundance of TEs in genomes of obligate sexual species is not representing a symbiotic situation, but a steady state of a permanent conflict between aggressive transposons and their hosts, as suggested by Bestor (Bestor 1999). To oppose the potentially harmful effects of active TEs, the host genome has

evolved multiple “defense” mechanisms that suppress their activity, notably through epigenetic silencing. Importantly, this on-going competition between the host silencing mechanisms and persistent TEs appeared to have open the way to many genetic innovations with transposable elements being used as molecular sources of novel genes and functional genomic elements, and with defense mechanisms being re-used to control endogenous gene expression.

Class		Superfamily
Retrotransposons	non-LTR	*CRE*
		NeSL
		R4
		R2
		L1
		RTE
		Jockey
		CR1
		Rex1
		I
		RandI
		Tx1
		SINE1
		SINE2
		SINE3
		Penelope
	LTR	*Copia*
		Gypsy
		BEL
		ERV1
		ERV2
		ERV3
		DIRS
DNA Transposons	“cut and paste”	*Chapaev*
		En/Spm
		hAT
		Harbinger
		ISL2EU
		Kolobok
		Tc1/Mariner
		Merlin
		Mirage
		MuDR
		Novosib
		P
		PiggyBac
		Rehavkus
		Transib
	Helitrons	
	Polintons	

Table 1. Classification of Eukaryotic Transposable Elements based on Repbase (adapted from Kapitonov, V.V. and J. Jurka, *A universal classification of eukaryotic transposable elements implemented in Repbase.* Nat Rev Genet, 2008. **9**(5): p. 411-2; author reply 414.)

Human Genome ~3200Mb		#of Copies (x1000)	% of Genome	Activity
LINEs		**868**	**20.42**	
LINE1		516	16.89	Active
LINE2		315	3.22	
LINE3		37	0.31	
SINEs		**1558**	**13.29**	
Alu		1090	10.6	Active using L1 RT
MIR		393	2.2	
MIR3		75	0.34	
SVA		2.76	0.15	Active using L1 RT
LTR retro-transposons		**443**	**8.29**	
ERV class I		112	2.89	
ERV (K) class II		8	0.31	
ERV (L) class III		83	1.44	
MaLR		240	3.65	
DNA Transposons		**294**	**2.84**	
hAT	Charlie	182	1.39	
	Zaphod	13	0.16	
Tc-1	Tigger	57	1.02	
	Tc2	4	0.03	
	Mariner	14	0.1	
PiggyBac-like		2	0.02	
Unclassified		22	0.12	

Mouse Genome ~2800Mb		#of Copies (x1000)	% of Genome	Activity
LINEs		**660**	**19.21**	
LINE1		599	18.78	Active
LINE2		53	0.38	
LINE3		8	0.05	
SINEs		**1498**	**8.22**	
B1 (Alu)		564	2.66	Active using L1 RT
MIR/MIR3		115	0,57	
B2		348	2.39	Active using L1 RT
B4/RSINE		391	2.36	
ID		79	0.25	
LTR retro-transposons		**631**	**9.87**	
ERV class I		34	0.68	Active
ERV class II		127	3.14	Active
ERV class III		37	0.58	Active
MaLR (III)		388	4.82	Active
DNA transposons		**112**	**0.88**	
hAT	Charlie	82	0.62	
	Other hATs	8	0.06	
Tc-1	Tiggger	24	0.17	
	Mariner	1	0.01	

Table 2. Classification of Human and Mouse Transposable Elements (taken from Mandal, P. and H. Kazazianjr, *SnapShot: Vertebrate Transposons.* Cell, 2008. **135**(1): p. 192-192.e1.)
Mice have greater transposon activity than humans because of a higher content of lineage-specific repeats. Endogenous retroviruses (ERVs) are extinct in humans, whereas all three classes of ERVs have active members in the mouse.

2.1 Silencing Mechanisms against TE activity

The defense mechanisms that could prevent potentially harmful effects and invasion of the genome by TEs are quite diverse and target different steps of the transposon cycle. They could act at the level of:

(1) Transcription of the TEs

(2) Post-transcriptional processing of TE RNAs

(3) Integration of new TE copies.

2.1.1 Transcriptional control of TE activities

1. The role of DNA Methylation: (Figure 2.1) In mammals, DNA methylation, which targets cytosines within CpG dinucleotides, has a key role in transcriptional silencing. DNA methylation was even proposed to have evolved primarily as a defense system against TE activity, since most of the methylated DNA is found within

transposons (Yoder et al. 1997). When associated with transcriptional machineries (5'LTR/5'UTR) of the transposons, 5'methyl cytosine (5mC) causes transcriptional repression of TEs by blocking the accessibility of transcription factors to their binding sites containing a CG or by recruiting transcription repressor complexes such as the **P**oly**c**omb **G**roup (PcG) proteins (reviewed in (Beisel and Paro 2011)). Moreover, methylation of a cytosine eventually may lead to deamination generating a uracil, which will be replaced by a thymine base in subsequent DNA replications. These base conversions introduce changes in binding motifs of regulatory factors and immobilize TEs due to irreversible inactivation of their promoters (Rollins et al. 2006).

In mammals DNA methylation is governed by four enzymatically functional DNA(Cytosine-5-)methyltransferases (Dnmts), Dnmt1, Dnmt3a, Dnmt3b and Dnmt2 (Jones and Liang 2009), which are all capable of catalyzing the methyl group to C-5 of cytosine but have different biological activities (Liu et al. 2003). Methylation of retroelements has to be re-instituted after the global erasure of methylation that occurs in cleavage embryos and later again in primordial germ cells. These marks are re-established co-operatively by Dnmt3A, Dnmt3B, and the related Dnmt3L in early embryos. A detailed analysis of the mutants revealed a certain degree of specificity of *de novo* methyltransferases for given types of retroelements. For example, Dnmt3A methylates SINE-B1 repeats, Dnmt3B methylates satellite repeats and Dnmt3A with the co-factor Dnmt3L are required for **L**ong **In**terspersed **E**lement-1 (LINE1-L1) and **I**ntracisternal **A-p**article (IAP) repeat methylation in the male germline (Kato et al. 2007). Dnmt1 is unique amongst these methylases in the sense that it has a strong preference for hemi-methylated DNA and thus is responsible for the maintenance of DNA methylation after DNA replication. A loss of Dnmt1 leads to progressive loss of cytosine methylation in dividing cells. Consequently, mouse embryos lacking Dnmt1 lose methylation on different types of TEs and die before mid-gestation stage (~8.5 d.p.c.) (Maksakova et al. 2008). The simultaneous inactivation of the Dnmt3A and Dnmt3B genes results in the undermethylation of endogenous MLV and IAP elements, but the phenotype is not as severe as the one observed with the Dnmt1 knockout (that is a loss of methylation down to 25% of wild type levels) (Okano et al. 1999).

In addition to the DNA-methyltransferase themselves, many other proteins that assist the methylation reaction and have therefore a role in TE transcriptional repression, as exemplified by Lymphoid-specific helicase, Lsh, a member of SNF2 chromatin-remodelling ATPase. *Lsh* knockout mice suffer from DNA demethylation of

retroelements including IAP, L1 and SINE B1, and subsequently an up-regulation of repeat elements (Dennis et al. 2001), (Huang et al. 2004).

Besides the 5mC, a novel cytosine modification, 5'hydroxymethyl cytosine (5hmC), has been recently discovered in mammals. It has been shown that 5hmC base interferes with DNMT1-mediated methylation (*in-vitro* tests in (Valinluck and Sowers 2007)), suggesting that this base is thought to be responsible for passive DNA demethylation by excluding DNMT1 binding/interaction. In trypanosomes a similar DNA modification is established by the JBP1/2 enzymes. In mammals paralogous proteins to JBP1/2 have been identified and named as **T**en **E**leven **T**ranslocation (TET) proteins (Tahiliani et al. 2009). All three proteins can convert 5mC to 5hmC in **E**mbryonic **S**tem (ES) cells (Ito et al. 2010). To date, the role of TET proteins regarding TE silencing has only been functionally tested for TET1, by a knock-down of its expression in ES cells. This analysis showed that TET1 is not contributing to IAP or microsatellite repeats' 5mC content (Williams et al. 2011); however LINE1 promoters accumulate 5mC while losing 5hmC (Ficz et al. 2011). These observations correlated with the methyl cytosine status of IAPs and LINEs in wild type ES cells. IAPs are more than 90% methylated whereas LINEs are only ~40% methylated (Tsumura et al. 2006)) and therefore could be further methylated in the absence of TET. Taken as a whole, 5mC has in many ways proven to be functioning against TE activity and in the light of newly found cytosine modifications it looks like there will be more to add on the mechanisms that regulate TE activity by changing DNA accessibility.

2. Chromatin remodeling: (Figure 2.2) Covalent modifications on certain amino acids of histone tails function as one of the main determinants of chromatin function and impinge on gene expression. Some histone tail modifications are negatively correlated with transcription and are called as "repressive marks". In general, methylation of histone tails (except histone3 lysine4 (H3K4) mono- and di-methylation) is associated with transcriptional repression (Zhou et al. 2011). DNA methylation and histone tail methylation are linked together *in vivo* (notably through the interaction of histone methyltransferase G9a with Dnmt3A/B (Dong et al. 2008)) and these mechanisms are probably acting in parallel. Repressive marks such as histone3 lysine9 (H3K9) tri-methylation and histone4 lysine20 (H4K20) mono-methylation were found on inactive TE promoters in ES cells (Martens et al. 2005),(Mikkelsen et al. 2007). Additionally the simultaneous inactivation of **P**olycomb **R**epressive **C**omplexes PRC1 and PRC2, which are needed to establish the

H3K27me3 marks, led to the up-regulation of TEs in ES cells, suggesting a functional correlation between TE silencing and the presence of H3K27me3 over them (Leeb et al. 2010). Mutation of another histone modifier, the H3K9 methyltransferase gene *Suv39* resulted also in the up-regulation of TE transcription (Martens, O'Sullivan et al. 2005). Similarly, the knockout of the H3K9 di and trimethylation mediating protein ESET (SETDB1) resulted in the up-regulation of **M**o**us**e type-**D** virus (MusD) and IAP transcription (Matsui et al. 2010). Other proteins associated with chromatin remodeling or with the recruitment of complexes establishing different chromatin modifications have also been implicated in TE silencing. The conditional knock-out of KRAB-associated protein 1 (KAP1) in mouse ES cells was shown to lead to the up-regulation of IAP and MusD elements (Rowe et al. 2010). Importantly KAP1 acts as a scaffold for ESET, HP1 family heterochromatin proteins and the NuRD histone deacetylase complex (Urrutia 2003). Interestingly KAP1 is brought to specific sites through interaction with the KRAB Zinc Finger proteins (KRAB-ZFPs) to which their multiple zinc fingers confer sequence specific DNA binding properties. For example, binding of ZFP809 to the **P**rimer **B**inding **S**ite (PBS) of Murine leukemia virus (MLV) recruits KAP1 complex and is needed for silencing of these transposons (Wolf and Goff 2009). As KRAB-ZFPs belong to a very large and rapidly evolving family (with 304 members in human and 219 in mouse, amongst them 20% are contained in lineage-specific expansions (Emerson and Thomas 2009), one can speculate that some members could target specific chromatin states in a sequence specific manner.

2.1.2 Post-transcriptional control

1. Degrading TE transcripts through RNA interference: (Figure 2.3) RNA interference (RNAi) also plays a major role in silencing transposable elements. The small interfering RNAs (siRNAs) are 21-23 nt long non-coding RNAs processed from double stranded RNAs (dsRNAs) by the RNAi machinery. This siRNA pathway is involved in transposon silencing in plants, fungi and *Drosophila* (reviewed in (Girard and Hannon 2008)). Repeated sequences contribute abundantly to dsRNAs production due to anti-sense or read-through transcription. The antisense transcript could be produced either by promoters of adjacent genes or by cryptic antisense transcriptional activities of the transposon itself (Girard and Hannon 2008). Notably transposon terminal inverted repeats are able to generate direct reverse read-through transcripts, therefore corresponding dsRNA, such as the Tc1 transposon in

C.elegans (Sijen and Plasterk 2003). The Dicer proteins that are core components of RNAi machinery as they process double-stranded transcripts into siRNAs, play therefore a central role in detection and silencing of transposons. Indeed, ES cells with an inactivation of *Dicer* gene by knockout displayed elevated levels of IAP and L1 and furthermore *Dicer*-depleted embryos showed also a transcriptional increase of IAP and **M**urine **E**ndogenous **R**etrovirus-**L** (MuERV-L) (Svoboda et al. 2004). Moreover, disruption of *Dicer* in growing oocytes leads to the up-regulation of Mouse Transposon (MT) and SINEs. The deficiency of Tc1 transposable element silencing in *C.elegans mut-7* mutants (obtained from a screen of RNAi resistance) provide another evidence that RNAi action is a general system for TE silencing in many organisms (Ketting et al. 1999). Interestingly, many of the genes up-regulated in oocytes without *Dicer* have 3'UTR containing sequences derived from TEs. This is suggesting that the regulatory role of Dicer on these genes is mediated via targeting the TE-derived motifs that have been gained in these mRNAs (Murchison et al. 2007).

Another class of small RNAs has been linked to RNA-interference-based TE repression. However, these small RNAs are not derived from dsRNA precursors but from single-stranded RNA transcribed from genomic clusters and the processed through a complex containing PIWI proteins. These PIWI-interacting RNAs (piRNAs) are slightly longer than siRNAs (24-30 nt) (Aravin et al. 2007). Importantly, PIWI proteins are strictly restricted to the germline. In *Drosophila* PIWI-piRNA complex interacts with HP1a and is localized at genomic sites enriched for methylated H3K9, indicating a close relation to triggering heterochromatin formation for TE silencing (Brower-Toland et al. 2007). Mice have three PIWI proteins and among them, MILI and MIWI2 are involved in the host defense against TEs (Siomi et al. 2011). Mutants of either protein fail to down-regulate L1 and IAP transcripts and lead to a loss of DNA methylation of L1 elements resulting in male sterility due to spermatogenic arrest (Aravin et al. 2007). Absence of PIWI protein has less dramatic consequences for the female since endo-siRNAs can also target TEs in female germline (Watanabe et al. 2008).

Altogether, these findings indicate that both piRNA and siRNA pathways suppress transposons expression and activity in mice.

2. Nucleic acid editing: (Figure 2.4) TEs with RNA intermediates that manage to escape transcriptional silencing mechanisms may be targeted by RNA-editing enzymes. These enzymes could introduce mutations in TE RNA that may render the

newly integrated copy of the transposon inactive. This RNA-editing process has an important role in innate immune response against infectious RNA viruses (Muramatsu et al. 2000), but also plays a role in controlling endogenous retroelements. The proteins of the ADAR family convert adenosines into inosines and targets dsRNAs that are formed from inverted Alu and L1 repeats. The APOBEC proteins belong to another group of RNA-editing enzymes, which catalyzes deamination of cytosines to uracils. A member of this family, APOBEC3G was shown to inhibit retrotransposition of IAP and MusD elements, and to induce G-to-A hypermutations in their DNA copies (Esnault et al. 2005). Another member of same the family, APOBEC1, has the same activity on IAP and MusD elements, but was also found to prevent retrotransposition of L1 (Ikeda et al. 2011).

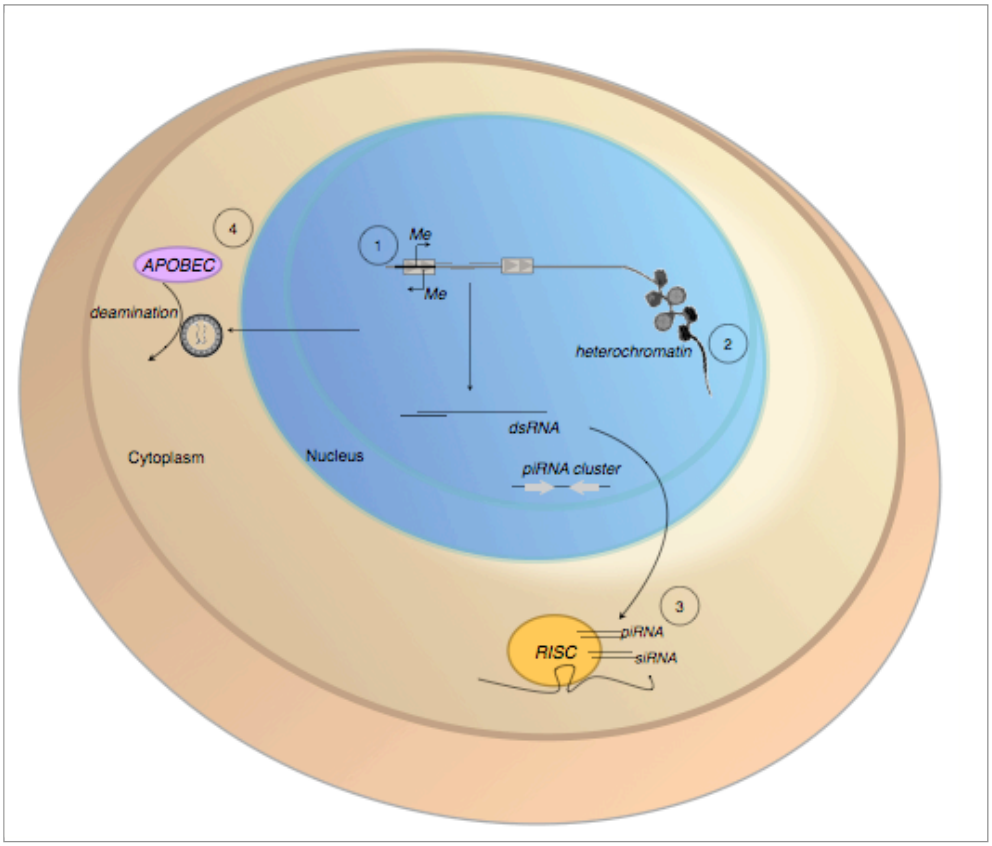


Figure 2. Silencing mechanisms against TE activity
(adapted from Goodier, J.L. and H.H. Kazazian, *Retrotransposons revisited: the restraint and rehabilitation of parasites.* Cell, 2008. 135(1): p. 23-35.)

2.1.3 Blocking the integration

Finally, the last step of the TE cycle, integration, can also be regulated by host proteins, notably by the host DNA repair machinery. The ERCC1/XPF complex takes part in nucleotide excision repair pathway, through its endonuclease activity. The knock-down of XPF in human cells led to an increase in L1 retrotransposition, suggesting that this complex might be cleaving the target-site primed reverse transcription intermediate, which would effectively block retrotransposition. In an opposing direction, another DNA repair enzyme, ATM (a member of double-strand break repair) was shown to facilitate L1 integration (reviewed in (Zamudio and Bourc'his 2010)). These observations suggest that different DNA repair pathways can modulate the process of transposon integration either negatively or positively.

2.2 Deleterious Effects of Transposable Elements that influence the host gene expression

Besides increasing the cost of DNA replication through their multiplication, active transposons could jump into new loci, and there, have specific deleterious effects by interfering with the activity of neighboring endogenous genes. Supporting this idea, TE insertions result in 0.1% of *de novo* human mutations. In mice, more than 10% of spontaneous mutations are caused by ERV insertions owing to the high activity of these elements (Maksakova et al. 2006). Since mobile elements are interspersed between the host's genes, they could influence gene expression and function in many ways. Interestingly, in some cases, these influences primarily are not caused by the insertion of transposon sequences themselves, but by the consequences of action of TE silencing mechanisms in these new locations, which could spread and then affect the neighboring bystander genes as well.

2.2.1 Impact on gene transcription

Many transposons have strong constitutive promoters. When inserted nearby a gene, if they are in an active state, they may function for this gene as an alternative transcription start site or a cryptic promoter, leading to its ectopic expression (Figure 3a). For example, in the A^{iapy}/A^{vy} locus the IAP retrotransposon produces an outward-reading transcript that extends into the *Agouti* coat-color gene. *Agouti* gene level is therefore determined in part by the transcriptional level of the retrotransposon, which lead to mice with variegated coat colors (because of the stochastic epigenetic silencing of IAP) as well as to obesity and diabetes (Michaud et al. 1994), (Morgan et

al. 1999). Alternatively, TEs may change the activity of endogenous genes by disrupting their *cis*-regulatory elements (Figure 3b). They could also modify an endogenous gene promoter sequence by introducing new transcription factor-binding sites and drive the expression of the gene in a different context; such a case was reported for an IAP insertion upstream of Interleukin-3 (IL-3) gene that led to constitutive expression of IL-3 in a leukemia cell line (Ymer et al. 1985)(Figure 3c).

When inserted in an intron or downstream of a gene, a mobile element may interfere with the original transcript of the gene by producing an antisense transcript (Figure 3d). As described before (in section 2.1.2) TEs can take part in dsRNA formation. In such situations, these dsRNAs could incorporate sequences from endogenous nearby genes and therefore impinge on their expression levels. Conley *et al.* proposed such a regulatory role for TEs as they observed an enrichment of TE initiated transcripts in cis-natural antisense transcripts (cis-NATs are RNAs that are transcribed from the antisense strand of a gene locus) (Conley et al. 2008))

Chromatin domains have rather imprecise boundaries, as exemplified by the phenomenon of position-effect variegation, representing the stochastic silencing of transgenes inserted close to heterochromatic domains (Sun et al. 2004). Similarly, repressive chromatin marks can spread from silenced TEs (chromatin spreading in yeast and plants is reviewed in (Medstrand et al. 2005)). In humans, Alu sequences are major targets for H3K9 methylation thus Alu elements can locally alter the chromatin state around their insertion points (Figure 3e) (Martens, O'Sullivan et al. 2005). This spreading of histone methylation from human Alus has been implicated in silencing some tumor-suppressor genes (Goodier and Kazazianjr 2008), suggesting that in general TEs could have a role in repressing proximally located genes by altering the local chromatin structure.

2.2.2 Post-transcriptional interference

A transposon inserted at the 3'UTR of a gene can introduce an alternative poly-adenylation site (Figure 3f). Since 3'UTR contributes to RNA stability, export and translation (Mazumder et al. 2003), an alternative 3'UTR provided by a TE can crucially modify the expression of a targeted gene. For example, the polyadenylation signal donated by the gibbon ape leukemia virus insertion into *IL-2* gene results in the constitutive production of IL-2 in MLA144 leukemia cell line (reviewed in (Keshet et al. 1991)). TE insertion at the 3'UTR could also bring a binding sequence for a miRNA, leading to post-transcriptional silencing of the affected gene in cells

expressing the corresponding miRNA (Figure 3g), as proposed by Feschotte and colleagues (Feschotte 2008).

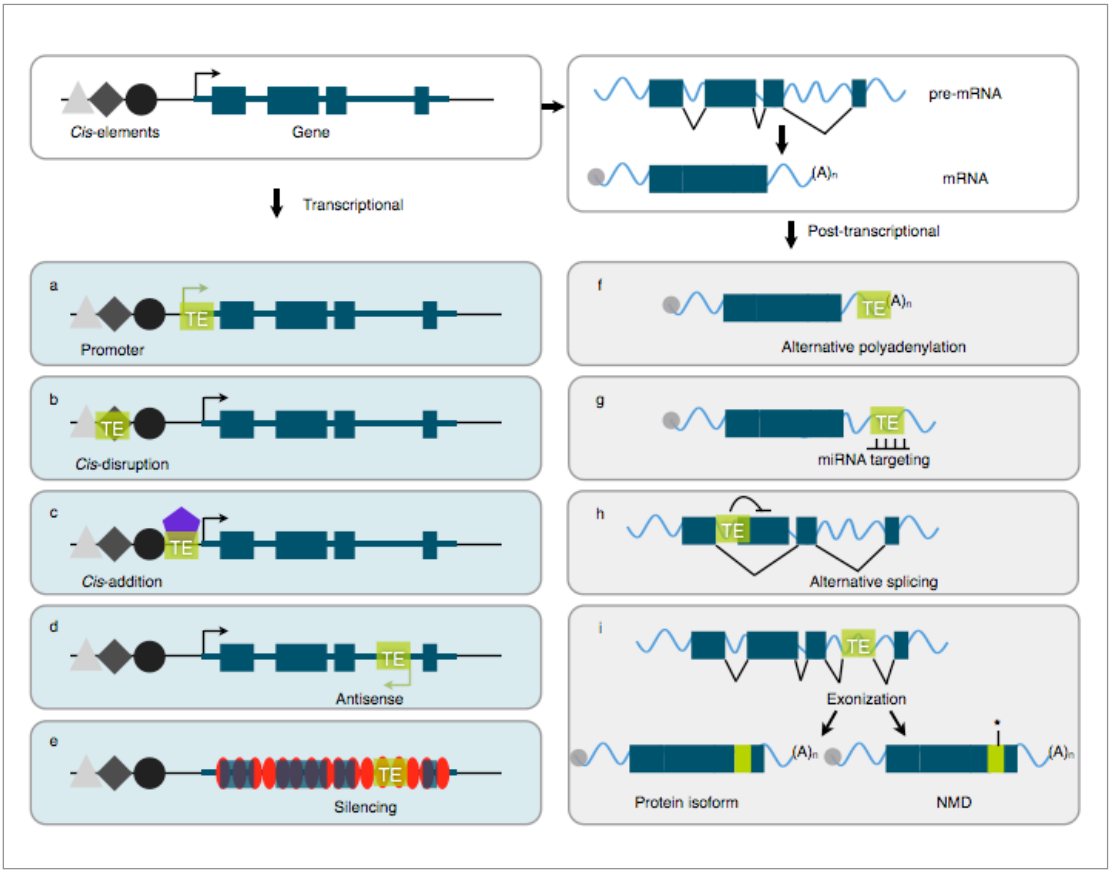


Figure 3. Influence of TEs on gene expression
(adapted from Feschotte, C., *Transposable elements and the evolution of regulatory networks.* Nat Rev Genet, 2008. **9**(5): p. 397-405.)

TE insertions, especially in introns, can interfere with the normal splicing process and lead to the production of abnormal (and often non-functional) transcripts. This accounts for most of the mutations caused by *de novo* insertions of active TEs in mice. For instance, the mutation of the *Gli3* gene in *Polydactyly Nagoya (Pdn,* a disease that is characterized by a mild polydactyly on the anterior side of the hind limbs*)* is caused by an **E**arly **T**ranspos**o**n (ETn) insertion into 3rd intron, leading to alternatively spliced transcripts that produce non-functional protein variants (Thien and Rüther 1999) (Figure 3h).

Finally, when TE sequences are incorporated into coding sequences, they may modify protein structure or introduce premature stop codons and trigger nonsense mediated decay (NMD) (Figure 3i). This was for example demonstrated in

the case of the exonization of an Alu generating splice variants of the *Survivin* gene which are then subject to NMD in the cytoplasm (Mola et al. 2007).

2.2.3 Alteration of the Genome Structure

The activity of TEs can also lead to important changes in the genome structure which are not limited to the insertion of their own sequences. The 3' processing machinery can sometimes skip the poly-adenylation signal of an L1 element, and use a downstream signal instead. In such cases, the non-L1 3'end flanking sequence, usually from the host genome, is transported together with the copied L1 to the new genomic location. This process is called transduction and is rather common in L1 retrotransposition (Goodier et al. 2000): young L1 elements could generate transduction of host sequences in 15-20% of the transposition events. Based on this, Pickeral *et al*. predicted that as high as 1% of the human genome could be derived from transduction events (Pickeral et al. 2000).

Retrotransposition sometimes generates target site deletions, which can be quite large as seen in the deletion event of an entire HLA-A gene caused by an SVA element insertion, which resulted in leukemia (Takasu et al. 2007). In addition non-homologous recombination events between different copies of transposons could cause deletions, duplications or rearrangements of gene sequence (Goodier and Kazazianjr 2008). **N**on-**a**llelic **h**omologous **r**ecombination (NAHR) is the most common mechanism underlying the disease associated genome rearrangements and it is often triggered by the misalignment between highly homologous sequence elements usually arising from segmental duplication and termed **L**ow **c**opy **r**epeats (LCRs). The identification of Alu sequences at the junctions of genes/pseudogenes within LCR regions led to the proposition that the generation of LCRs could be associated with Alu elements (Shaw and Lupski 2004). Alu elements (or longer TEs) could provide short regions of homology and when positioned in frequent intervals could favor unequal crossing-over caused by these elements. This could result in duplications and deletions of the intervening regions or even more complex re-arrangements. In addition to Alu repeats, LTR repeats can also participate in recombination reactions that generate retention of solitary LTRs and an excised episomal fragment (reviewed in (Prak and Kazazian 2000)). The observation that both human and mouse segmental duplication boundaries are enriched in various classes of retrotransposons (Bailey et al. 2004) emphasizes the extent of genome alterations associated with transposons and suggests that transposons either

through their activities or presence as repeated sequences could promote such re-arrangements, thus shape the organization and structure of host genomes.

2.2.4 Transposon Free Regions in mammalian genomes

Interestingly, like genes, transposons show an overall unequal distribution of transposon sequences across the genome (Lander, Linton et al. 2001). About 20% of the genome is considered to be gene poor and these regions also known as gene deserts are inconsistently distributed over chromosomes (Venter et al. 2001). These gene deserts are however not a pure storage places for junk DNA, as they are enriched in non-coding elements showing very high evolutionary constraints and conservation (Bejerano et al. 2004). Many of these elements were shown to have tissue specific enhancer activities and probably contribute to the expression of developmental genes that are often found next to these gene deserts (Nobrega et al. 2003), (Pennacchio et al. 2006). Interestingly, these gene deserts associated with important developmental regulators such as transcription factors, show with a decreased density of SINE and an increased number of LINE sequences. The depletion of SINEs could be explained by purifying selection: the high-CG composition of the SINE sequences which could lead to spreading of DNA methylation, and impair the normal activities of regulatory elements localized within (Ovcharenko et al. 2005). Strikingly, some genomic regions are almost completely depleted of any transposon. These **T**ransposon-**f**ree **r**egions (TFRs) are enriched around regulatory genes (such as the longest determined TFR (81kb) in the HOXA cluster) and miRNA genes in mammals. The expression of regulatory genes usually has to be tightly controlled, and this is often achieved through a complex interplay of multiple *cis*-regulatory elements spread over large genomic intervals surrounding each gene. Insertion of TEs in this intricate context could modify the delicate regulatory interactions and therefore would be eliminated through purifying selection. The observation that murine retroviral insertions into TFRs are significantly associated with cancer further supports the suggestion of the negative impact of TE insertions into such regions (Simons et al. 2006).

However, it is interesting to note that, in contrast to mammalian genomes, the green anole lizard genome *Hox* genes clusters have accumulated TEs of non-LTR retrotransposon family (Di-Poï et al. 2009). While *Hox* patterns are usually conserved, changes in *Hox13* and *Hox10* expression features were observed during somitogenesis in species which have accumulated repeat elements within *Hox*

clusters, coinciding with altered axial morphologies (Di-Poï et al. 2010). However, one could not conclude from these observations whether these TE insertions have contributed to *Hox* expression changes or they were tolerated after changes in the *Hox* regulatory mechanisms that led to relaxed constraints on the structure of the region.

2.2.5 Somatic Retrotransposition

As mentioned earlier, the different mechanisms leading to transposon silencing are mostly established early during embryogenesis: as a consequence retrotransposition occurs mostly during gametogenesis. However, recent work has revealed substantial somatic activities of retro-elements leading to somatic retrotranspositions as well. First evidence of the existence of a somatic event was identified with a colorectal tumor case where an L1 element is inserted into tumor suppressor *APC* gene and disrupting its expression (Miki et al. 1992). Later on taking the advantage of cell culture based systems and transgenesis, it has been shown that engineered L1 elements can retrotranspose in multipotent neural progenitor cells (Muotri et al. 2005) and in human cells (Kubo et al. 2006). This retrotransposition is restricted to neural progenitor cells and the suppression of the retrotransposition is correlated with Sox2 (Muotri, Chu et al. 2005) and **m**ethyl-**C**pG-binding **p**rotein **2** (MeCP2) expression. As MeCP2 is mutated in Rett syndrome patients, these patients could be more sensitive to L1 retrotransposition, raising a possibility that increased L1 activity in neurons could contribute to their neurological symptoms (Muotri et al. 2010). New somatic insertions, especially at early developmental stages, may contribute to changes in the genetic and epigenetic status of mature neurons at later stages of life and this genomic plasticity created by somatic retrotransposition events could contribute to individual variation and to the progression of neurological diseases.

2.3 Co-option of Transposable Elements for the host chromosome form and function

As exemplified above, there is a large and accumulating body of evidence that shows that TEs could impact endogenous gene expression in many ways and that the resulting changes could be highly detrimental to the host. However, TEs evolve to survive and amplify, and in the process they may accumulate numerous

beneficial features that are in fact useful for the host too (Kidwell and Lisch 2000). Following up on McClintock's idea regarding TEs as controlling elements Britten and Davidson proposed that mobile elements could offer an opportunity to the host as a vector to spread rapidly similar, potential, gene regulatory regions through the genome and generate a 'battery' of co-regulated genes (Britten and Davidson 1969). Indeed, growing evidence show that TEs provide a useful source of raw material to their hosts that could assist them in evolving new functions.

2.3.1 Centromeres and telomeres

Centromeres and telomeres are essential for genome integrity as they are controlling chromosome segregation during cell division and preventing chromosome shortening after replication, respectively. Both centromeres and telomeres are made of constitutive heterochromatin and are partially composed of retrotransposons. In the plant genome, these regions are frequently associated with the TEs that form heterochromatic boundaries within local euchromatic environments (Lippman et al. 2004). Pericentromeric regions of the human chromosomes are mainly composed of LINE and SINE transposons. Reactivation of TEs stored in this portion of constitutive heterochromatin leads to impaired genomic stability during mammalian meiosis. This chromosomal instability was also demonstrated in *Lsh* mutants where tandem repeats DNA were demethylated (De La Fuente et al. 2006) and in *Suv39h* mutants which lack H3K9 methylation (Peters et al. 2001) in pericentromeric regions. Moreover, TEs might have been the origin of centromeric satellite repeats themselves. **Cen**tromere **p**rotein **B** (CENPB) is highly conserved in mammals and helps to regulate centromeric heterochoromatin by binding to a motif in satellite repeats. This highly conserved protein resembles the transposase of *Tc1/mariner* DNA transposons. More generally, many organisms have centromeric satellite tandem repeats with homology to TE sequences, suggesting that they might have originated from TEs (reviewed in (Slotkin and Martienssen 2007)).

Telomeres are composed of short tandem repeats, which makes the DNA replication at these regions inefficient and leads to their shortening. Telomerase is the enzyme that is needed for the maintenance of telomere length. This enzyme has a reverse transcriptase domain that is structurally similar to the reverse transcriptase of non-LTR retrotransposons (Slotkin and Martienssen 2007). Interestingly, *Drosophila* species lack a telomerase homologue; instead, non-LTR

retrotransposons named *HeT-A* and *TART* have taken over the telomerase function by transposing to the chromosome ends (reviewed in (Volff 2006)).

2.3.2 TEs as regulatory elements: enhancers and chromatin insulators

The influence of TEs on gene regulation and expression was first discovered through the analysis of mutations found to be caused by individual TE insertions. More recent genome studies defined that 25% of experimentally characterized human promoters contain TE-derived sequences (Jordan et al. 2003) and some TEs are over-represented within predicted (using either analysis of transcription factor binding site motifs or DNaseI hypersensitive sites) *cis*-regulatory modules (Gentles et al. 2007), (Mariño-Ramírez and Jordan 2006). Correspondingly, Bourque and his colleagues demonstrated that TEs contributed in expanding the repertoire of transcription factor binding sites in eukaryotic genomes (Bourque et al. 2008). In human ES cells, TEs constitute up to 25% of binding sites for key regulatory proteins (OCT4, NANOG and CTCF) for stemness (Kunarso et al. 2010). A large number of non-coding elements under selective pressure could also be identified as ancient conserved transposable elements, suggesting that they may have been exapted to contribute to gene regulation. Interestingly, these elements are mostly located in gene deserts and enriched near developmental genes (Lowe et al. 2007). Specific examples of such elements include a distal enhancer of the neuro-developmental gene *Isl1* which was discovered initially as a member of the ancient lobed-fin SINE family (Bejerano et al. 2006). Another study showed that extant copies of AmnSINE (a member of Deu-SINE superfamily) act as enhancers of *Fgf8* and *Satb2* genes for developing mammalian forebrain. (Sasaki et al. 2008). Besides these examples where the TE is apparently directly controlling gene expression, some TEs could have a rather modulatory role. An LTR retrotransposon, ERV-9, located upstream of human fetal γ- and adult β-globin genes modulates gene switching mechanism in long-range. This switch is achieved by RNA polymerase II occupancy on the LTR of this element in embryonic cells. When RNA polymerase II is no longer on its LTR, the transcription factor occupancies change and fetal to adult globin switch occurs (Pi et al. 2010). Another modulatory role of a TE is observed with the *Tal1* gene. One of its upstream elements is found in open chromatin configuration but does not function as an enhancer. It has a quantitative effect on gene expression and boosts the activity of a nearby tissue-specific enhancer. Importantly, the **M**ammalian **I**nterspersed

Repeat (MIR) that is comprised in this element is needed for this boosting activity (Smith et al. 2008).

Besides this contribution of TEs to specific and peculiar role for disparate genes, TEs contribute to an important part of the transcripts deposited maternally or expressed in early embryonic stages. Peaston et al. found that various classes of TEs are expressed in mouse oocytes and preimplantation embryos, with a domination of Class III ERV family elements. These transcripts are characterized as alternative chimeric gene products as their 5' regions carry TE-derived sequences. They contribute to maternal mRNA pool largely with many isoforms of the same gene. These isoforms could provide additional means of gene regulation as they may have variable post-translational modifications. Furthermore, different LTR retrotransposons have specific and developmentally regulated expression patterns in oocyte-to-embryo transition and could constitute alternative promoters and 5'exons of the host genes (Peaston et al. 2004).

Besides being enhancers and promoters, TEs may function as insulators. Chromatin insulators set up territories of gene expression (active or silent states) along the chromosome and several TE or TE-derived elements have been shown to have similar activities. The insertion of the *Drosophila* gypsy retrotransposon into regulatory regions causes tissue-specific mutations at a number of loci. This insulator effect is mediated by Su(Hw) protein, which binds to gypsy sequences, disrupting the enhancer-promoter communication (Corces and Geyer 1991) (Jack et al. 1991). A more recent study in mammals reported B2 SINE repeat as a boundary element in organogenesis. This activity is associated with bidirectional, non-coding transcription from this repeat element. This boundary corresponded to local changes of chromatin state during development, but also displayed enhancer-blocking activity in a cellular assay (Lunyak et al. 2007).

2.3.3 X inactivation and Imprinting

Transposons have also been proposed to contribute to regulatory mechanisms by controlling large chromosomal domains such as X Chromosome Inactivation (XCI). In mammals, females inherit two X chromosomes whereas males inherit one X and the degenerate Y chromosome. This dosage difference is compensated by the females shutting off one of their X chromosomes by heterochromatization. Inactivation is initiated at the **X**-chromosome **i**nactivation **c**enter (Xic) by expression of noncoding *Xist* RNA and then spreads to the rest of the

chromosome in *cis*. This action is regulated by *Tsix,* the antisense noncoding regulator partner of *Xist*. Interestingly during XCI, *Tsix* expression is controlled by a remnant of an ancient retrotransposon (*DXPas34*), which initially acts as an enhancer and once XCI is established, as a repressor of *Tsix* (Cohen et al. 2007). In addition to *DXPas34* it is hypothesized that LINE retrotransposons help as boosters for the efficient spreading of the silencing away from the Xic. The decrease in LINE sequence density due to translocations between autosomes and X chromosome results in the loss of efficient spreading from Xic (Lyon 2000).

Similar to their proposed role in mammalian X chromosome regulation, TEs are densely present within imprinted loci on autosomes. In mammals, interestingly, different types of TEs are silenced between females and males: the IAP and LINE1 retrotransposons are hypomethylated in the female germ line, and contribute highly to mRNAs in oocytes, which are deposited in the embryo (Peaston, Evsikov et al. 2004), whereas non-autonomous SINEs are hypomethylated in sperm (Rubin et al. 1994). These observations taken together imply that different classes of TEs might contribute to establishment of the differential epigenetic marks of paternally and maternally imprinted genes. Consistent with this, there is an excess of LINE1 elements that are associated with paternally expressed autosomal imprinted genes (Allen et al. 2003).

2.3.4 Protein coding host genes domesticated from transposable elements

TE genes can evolve as new genes with functions beneficial to the host. In human genome ~4% of the protein coding genes harbor TEs in their coding sequence (Nekrutenko and Li 2001). One of the best studied mechanisms is the domestication of a transposase protein derived from Transib DNA transposons to form the *Rag1* gene that is controlling somatic V(D)J recombination in the B and T cells (Agrawal et al. 1998). Another example of domestication is of a retroviral envelope protein called Syncytin. Syncytin is the envelope protein of a defective human ERV, HERV-W. It is expressed in placental syncytiotrophoblasts (multinucleated cells that originate from fetal trophoblasts) and contribute to placental morphogenesis by mediating the cytotrophoblast fusion, which might help extending the surface for maternal-fetal exchange (Mi et al. 2000). In mouse and other *Muridae,* different retroviral envelope genes related to Syncytin gene have been identified, suggesting a convergent domestication of these genes in the primate and murine lineages (Volff 2006).

2.4 Phenotypic variation and Evolutionary Implications

The extent of the effects produced by TEs added to the fact that they can change location rather rapidly (on an evolutionary time scale) suggests that they could contribute significantly to intra-species variation and evolution of species. Importantly, they can produce genetic and phenotypic variations between individuals by active mechanisms (mutations) as well as by serving as a source of variable epigenetic effects on endogenous genes, leading to further phenotypic variation without genetic changes.

This hypothesis has long been discussed, especially for a young L1, specific to the human genome and present with variable allelic frequencies among human populations (Sheen et al. 2000). A comprehensive map of genetic variation of all forms (such as SNPs, indels, and transposon insertions) is in the process of being established through the effort of the 1000 Genomes Consortium (Consortium 2010) and should help to get better views and understanding of the contribution of TEs. So far, analysis of this data set revealed that ~30% of genomic **S**tructural **V**ariations (SV) in human populations is due to mobile element insertions, with these SVs overlapping with genes region in one third of the cases (Mills et al. 2011).

Intriguingly, a study on the koala genome discovered a recently inserted endogenous retrovirus, KoRV (koala retrovirus) that is vertically transmitted and quite active. Some of the isolated koala populations are lacking KoRV and the copy numbers are variable between the individual animals where it is present. With its high level of activity, KoRV suggests that it is at a transition stage between an exogenous virus and endogenous element (Tarlinton et al. 2006). The dynamic character of KoRV provides a unique chance to study the invasion of a mammalian genome by retroviruses and to observe how it could change the genome of a species.

As previously introduced in section 2.2.4, transposons may stimulate all types of genome re-arrangement events that may eventually lead to speciation. Interestingly, there is an association of TEs with species-specific chromosomal re-arrangements. For instance, in fishes the burst of non-LTR retrotransposons seemed to be associated with speciation events (Volff et al. 2000), (Volff et al. 2001). More hypothetically, it was suggested that the massive transposition of TEs observed in *Drosophila* inter-species hybrids could contribute to genomic changes and speciation (reviewed in (Böhne et al. 2008)). More generally, extending on the view that chromosome re-arrangements may play significant roles in speciation (Ayala and Coluzzi 2005) the re-arrangements mediated by TEs in vertebrates might have been

an important factor for divergence between species (Böhne, Brunet et al. 2008). For instance, the primate-specific burst of Alu elements resulted in increased Alu-mediated recombination events that lead to large homologous segmental duplications that are common in the human genome (Bailey et al. 2003). In addition, species-specific transposons might have contributed to evolution of humans and chimpanzees since the divergence from their last common ancestor (Mills et al. 2006).

An important property of mobile elements is that they could generate phenotypic variation by creating epialleles, i.e. alleles that display variable expressivity in the absence of genetic heterogeneity, because their activity is dependent on their epigenetic state. In mammals, retrotransposons become demethylated and active at the pre-implantation stage and then undergo co-suppression around gastrulation stage. Silencing is both stochastic and incomplete resulting in a variegated pattern (between cells) or variable expression (between individuals) (reviewed in (Rakyan et al. 2002)). Variegating repression of allelic TEs can cause differences in the silencing of proximal genes. One example is the differential methylation of the IAP present in the mouse agouti locus which is driving transcription of *A* as a cryptic promoter. Isogenic A^{vy} mice have coats that vary in a continuous spectrum from full yellow, through variegated yellow/agouti, to full agouti (Morgan, Sutherland et al. 1999). Similarly, in Axin-fused ($Axin^{Fu}$) mice the differential methylation of the LTR of an adjacent IAP transposon (inserted into intron 6 of *axin* gene) correlates with the severity of the kinked tail phenotype (Rakyan et al. 2003).

Changes in gene expression contribute importantly to evolutionary process. This hypothesis that was originally made by King and Wilson (King and Wilson 1975) has now received support from several studies (Carroll 2008). The polymorphic influence that TEs can have on endogenous gene expression suggests that indeed they could have contributed to speciation as proposed by Jurka (Jurka, Kapitonov et al. 2007). Their rapid mode of propagation through the genome could make them even provide a fast alternative for the host's evolution when subjected to stress (Hurst and Werren 2001). Without going that far, it is that TEs are acting as important "controlling elements" as coined by Barbara McClintock (MCCLINTOCK 1956) that have shaped and are shaping genome structure and function, thus contribute to evolution.

2.5 Dactylaplasia: A case study of the influence of transposable elements on gene expression and specific silencing mechanisms

2.5.1 *Dactylaplasia* mice

In the 1970s, a spontaneous mutation, with missing digits of the limbs was observed in the SM7B/SM inbred mouse strain of the Jackson Laboratory (Bar Harbor, Maine). This *Dactylaplasia* mutation is a semi-dominantly inherited mouse limb malformation characterized by the absence of phalangial bones in the middle digits of each hand and foot, accompanied by reductions or fusions of metacarpals and metatarsals (as shown in Figure 4) (first time demonstrated in (Chai 1981)). Later on, a second allele of *Dactylaplasia* arose spontaneously in the Jackson Laboratory (described in (Sidow et al. 1999)). These two alleles of *Dactylaplasia* were named as Dac^{1j} (homozygous with partial neonatal lethality) and Dac^{2j} (homozygous with complete prenatal lethality). They were mapped either within or

Figure 4. Skeletal preps of *Dactylaplasia* mice E18.5 forelimbs.

close to the F-box gene Fbxw4 on chromosome 19qC3. Both were suggested to be insertions of transposable elements (Sidow, Bulotsky et al. 1999). A more recent study precisely positioned these mutations for Dac^{1j} to Chr19: 45,723,779 (Mouse February 2006 assembly) and ~50kb further for Dac^{2j} to Chr19: 45,669,788 (Mouse February 2006 assembly) and that they corresponded to TEs of MusD family (Friedli et al. 2008).

2.5.2 Limb development and the molecular basis of limb malformations in *Dactylaplasia*

Dactylaplasia embryos show a defect in the maintenance of the **A**pical **E**ctodermal **R**idge (AER), which is most likely the explanation for the observed limb

malformations. When the vertebrate limb bud first appears (at E9.0 in the mouse), it consists of mesenchyme surrounded by an ectoderm-derived epithelial layer. Signals from the mesenchyme induce the overlying ectodermal cells at the tip of the limb bud to change morphology and form a thickened ridge called the AER. After limb buds form, their continued growth depends on the continuous presence of a functional AER. By day E13.5, the AER is no longer visible as a morphological structure, and expression of most AER genes is terminated (Guo et al. 2003).

.

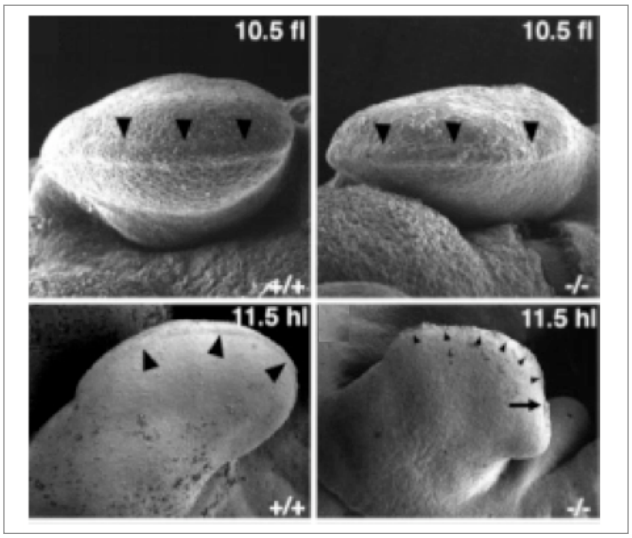


Figure 5. Loss of AER in *Dactylaplasia* mice, Scanning electron microscopy of limbs. Arrow heads represent AER, here it is shown that in homozygous mutants at E11.5 only posterior AER remains intact.

(adapted from Crackower, M.A., J. Motoyama, and L.C. Tsui, *Defect in the maintenance of the apical ectodermal ridge in the Dactylaplasia mouse*. Dev Biol, 1998. **201**(1): p. 78-89)

The AER removal experiments on chick wing buds demonstrated truncation of the distal limb elements or the complete absence of the limbs, depending on the developmental stage at removal (SAUNDERS 1948), (Summerbell 1974), showing that AER produces signals that are continuously required to maintain the proximodistal outgrowth of the limb. In *Dactylaplasia* mice, the AER was found to be morphologically normal at embryonic day (E) 10.5; however by E11.5 the central aspect of the AER degenerates (as shown in Figure 5) (Crackower et al. 1998). Thus, this premature degeneration of the AER as observed in *Dactylaplasia* embryos, in a late stage and predominantly in its central region accounts very well for the absent digits found in mutant mice. The signals produced by the AER to maintain limb outgrowth are encoded by **F**ibroblast **g**rowth **f**actor (Fgf) gene family members. *Fgf8* is expressed from the time of specification of the AER and onwards, and it is essential for limb development as demonstrated in mice with severe reduction of the limb size when *Fgf8* is inactivated specifically in the early limb ectoderm (Lewandoski et al. 2000), (Moon and Capecchi 2000). However, other Fgfs (*Fgff4*, *Fgf9*, *Fgf17*) are activated later in the posterior AER and compensate for *Fgf8* absence. Even

though the deletion of any one of these genes resulted in normal skeletal patterning, their deletion together with *Fgf8* led to more severe abrogation of limb formation, which is increasing with the number of deleted *Fgf*s (Mariani et al. 2008).

The induction of AER is directed by signaling pathways that operate within the ectoderm and between the mesoderm and ectoderm of the prospective limb bud. WNT/β-catenin signaling in the limb ectoderm is necessary to induce *Fgf8* expression, and therefore AER. The ligand that activates canonical WNT pathway is WNT3 and for AER formation, WNT3 signaling is required in the ventral ectoderm (Kawakami et al. 2001), (Barrow et al. 2003). The ability to induce an AER resides in the mesoderm and several studies showed that FGF10 is likely to be the corresponding factor (Min et al. 1998), (Sekine et al. 1999). Mice with a mutation of the FGF10 receptor, *Fgfr2b*, failed to induce *Wnt3* and did not form proper AER (De Moerlooze et al. 2000).

After initiation, the maturation of the AER is marked by the compaction of the AER cells over the dorsal-ventral tip of the limb bud. This process is under the control of many factors, including the homeodomain-containing transcription factor, *Engrailed1* (*En1*). In the absence of *En1,* the AER did not mature correctly and its anterior half expanded ventrally (Loomis et al. 1996). Once induced, the AER requires continuous maintenance signals from the limb mesoderm. This maintenance needs the balanced and positive inputs of WNT and FGF10 signaling from the ectoderm and mesoderm respectively, establishing an epithelial-mesenchymal (e-m) feedback loop between FGF8 and FGF10.

Growth and patterning of the vertebrate limb does not only rely on AER but also signals produced by another discrete signaling center: the **Z**one of **P**olarizing **A**ctivity (ZPA). The ZPA produces a signal, **s**onic **h**edge**h**og (*Shh*), which instructs limb bud mesenchymal cells with respect to their anteroposterior fates (Tickle 1981), (Riddle et al. 1993). AER is required to maintain the ZPA, and vice versa, indicating two interlinked signaling centers patterning the axes (Niswander et al. 1994). The BMP antagonist gremlin1 (GREM1) plays an important role in this crosstalk and is required to pass the SHH signal to AER. All together, these signals define the SHH-GREM1-FGF e-m feedback loop (as shown in Figure 6) (Zúñiga et al. 1999). The SHH–GREM1–FGF e–m feedback loop is self-terminated when the Shh expressing descendant population of cells expand. These cells are resistant to GREM1 signals and their expansion generates a gap that moves the *Grem1* expression domain where SHH signals cannot reach (Figure 6c) (Scherz et al. 2004). Finally, for the

regression of the AER, the bone morphogenetic protein (BMP) signaling is required. BMP controls apoptosis in the interdigital areas indirectly through their modulation of FGF signaling from the AER (Fernandez-Teran and Ros 2008).

Dactylaplasia has been mapped in proximity to the essential AER factor *Fgf8* (~70kb from Dac^{1j} and ~120kb from Dac^{2j}). Even though a lack of *Fgf8* limb expression has been observed in the mutants limbs (Crackower, Motoyama et al. 1998) (Figure 7), a down-regulation of *Fgf8* alone is not enough to explain the *Dactylaplasia* phenotype. Heterozygous *Fgf8-KO* animals are phenotypically normal and even limb-specific homozygous *Fgf8-KO* mice still have an AER (Lewandoski, Sun et al. 2000) and show mild defects compared to *Dactylaplasia* mice in homozygous. Thus, it suggests there is another reason for the observed *Dactylaplasia* phenotype. It has been demonstrated that there is a dramatic cell death or lack of cell proliferation in the AER of *Dactylaplasia* mice (Seto et al. 1997), (Crackower, Motoyama et al. 1998). The re-investigation of the mutants revealed increased level of apoptosis (after E10.25) and transiently increased level of proliferation (between E10.0-E10.5) (PhD thesis, Schwarzer W. 2010, Freie Universität Berlin) of cells in the *Dactylaplasia* limbs. The genes responsible for the apoptosis and proliferation differences in mutant limbs remain to be unknown. In fact, the absence of *Fgf8* is more likely to be the consequence of AER misfunction, and not its primary cause.

2.5.3 *Dactylaplasia*: A transposon insertion

The genomic basis of the *Dactylaplasia* mutations was analyzed by a massive PCR-based screening of the genetically mapped interval. This analysis suggested that both alleles were caused by insertion of additional sequences, from an ERV-like element (Sidow, Bulotsky et al. 1999). Subsequently, the insertions were characterized to belong in fact to the Young-MusD repeat family (Kano, Kurahashi et al. 2007), (Friedli, Nikolaev et al. 2008). MusD forms a group of murine ERV, which are highly similar at the level of the LTR, **P**rimer **B**inding **S**ite (PBS) and **P**oly**P**urine **T**ract (PPT) regions to the ETn elements (Baust et al. 2003) that were initially thought to be inserted in *Dactylaplasia* locus (Sidow, Bulotsky et al. 1999). Whereas ETn elements have no **o**pen **r**eading **f**rames (ORFs) or similarity to any known retroviral genes, MusD elements have the *gag*, *pro*, and *pol* genes with high resemblance to the primate type D viruses (Figure 8) (Mager and Freeman 2000). MusD elements are autonomously transposing sequences that lack the viral envelope gene (*env*;

which is needed for the budding off the cell). Therefore, they are members of the intracellularized retrovirus family (Ribet et al. 2007). The strong homology between MusD and ETn suggests that these latter ones are non-autonomous copies derived from an ancestral MusD element (Mager and Freeman 2000), (Ribet et al. 2004).

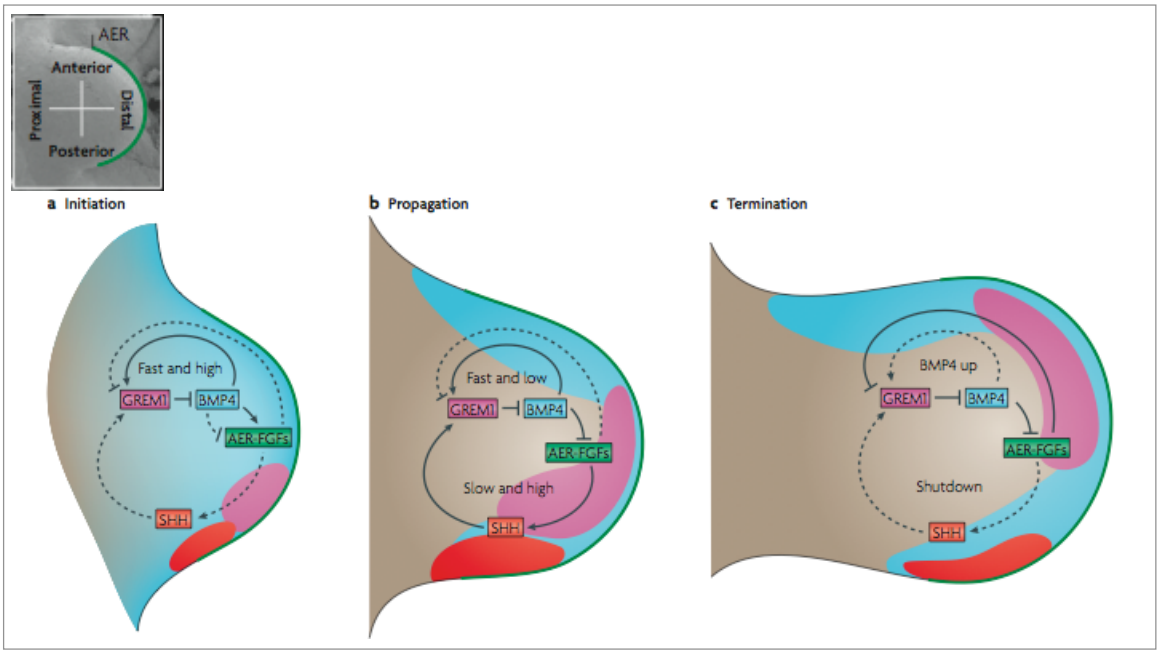


Figure 6. Interlinked feedback loops in limb development,
a) In the initiation phase, BMP4 up-regulates Grem1 expression and Shh expression is activated independently.
b) in the propagation phase, SHH up-regulates Grem1 expression and GREM1 reinforces AER-FGF and ZPA-derived SHH signaling by an e–m feedback loop
c) In the termination phase, the widening gap between expression domains terminates the signaling system.
(Figure is taken from Zeller, R., J. López-Ríos, and A. Zuniga, *Vertebrate limb bud development: moving towards integrative analysis of organogenesis.* Nat Rev Genet, 2009. **10**(12): p. 845-58 (Zeller et al. 2009))

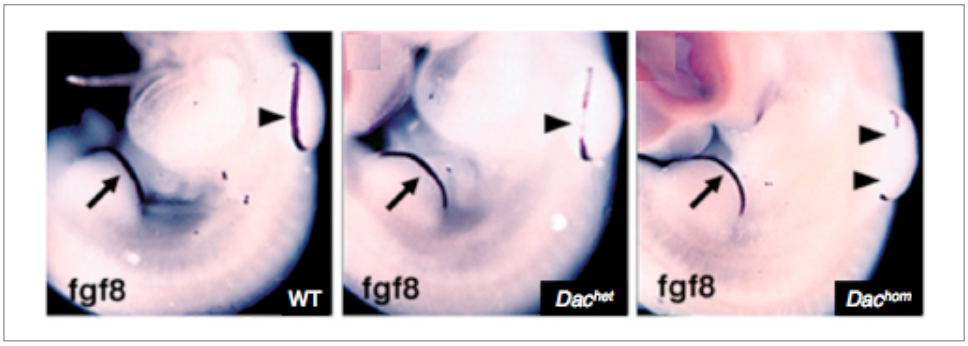


Figure 7. *Fgf8* expression in forelimbs and hindlimbs of E11 embryos
(Photos are taken from Crackower, M.A., J. Motoyama, and L.C. Tsui, *Defect in the maintenance of the apical ectodermal ridge in the Dactylaplasia mouse.* Dev Biol, 1998. **201**(1): p. 78-89)

The two alleles (*Dac*^{1j} and *Dac*^{2j}) of *Dactylaplasia* are insertions of distinct but highly similar (98% sequence identity) MusD elements. They occurred at different positions that are 50kb apart and in opposite orientation (Friedli, Nikolaev et al. 2008). The locus contains several genes around with *Fbxw4* gene being the nearest to both insertions (Figure 9). The *Dac*^{1j} insertion is 10kb upstream of *Fbxw4* and the *Dac*^{2j} insertion is in the 5th intron of this gene. Thus, it has been hypothesized that *Fbxw4* expression might be affected by these insertions that this could contribute to the phenotype. Indeed, Sidow and colleagues found out that *Dac*^{2j} insertion gives rise to an aberrant transcript (~9.5kb which is bigger than the wild type transcript of 2.8kb) with the absence of wild type transcript in homozygous mutants but in *Dac*^{1j} insertion carrying animals expression is not different from wild type levels (Sidow, Bulotsky et al. 1999). In the same study, both *Dac*^{1j} and *Dac*^{2j} mutants were crossed to a transgenic mouse (Krd mouse in (Keller et al. 1994)) with a large deletion on chromosome 19 (covering beyond *Tlx1* and *Fgf8* genes) and none of *Krd*^{het} progeny showed a limb phenotype. Thus, they concluded that the phenotype of *Dactyplasia* heterozygote was due to a gain of function or dominant negative effect. Despite these discrepancies, in several studies, *Fbxw4* is considered as the causative gene and even named as "*Dactylin".*

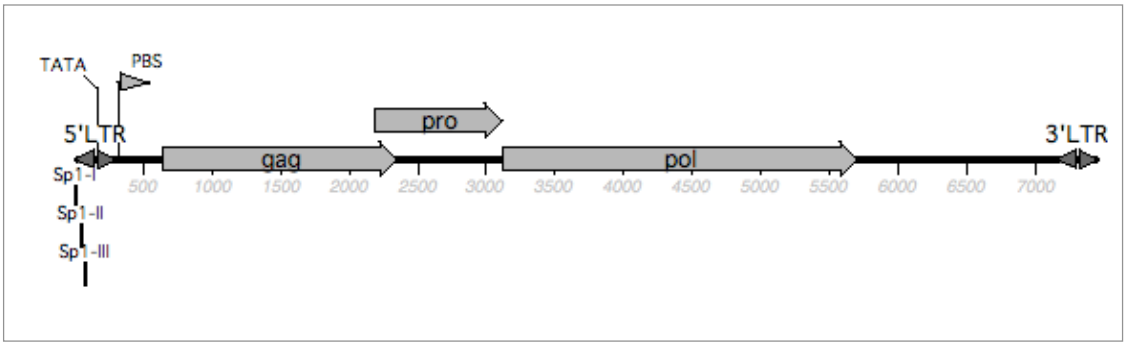


Figure 8. MusD retrotransposon characteristics
Proviral genes *gag* (for viral martix, capsid and nucleoproteins), *pro* (protease for gag-pol protein precursor maturation) and *pol* (polymerase for DNA and insertion) are found in MusD sequence. Primer binding site (PBS) for tRNA priming of reverse transcription (more specifically tRNA^{Lys}) is also predicted in the MusD sequence (Mager and Freeman 2000). Transcriptional activity of the 5'LTR is characterized and found to be regulated by Sp1 transcription factor binding and not affected by the absence of its noncanonical TATA box (Maksakova and Mager 2005).

2.5.4 *SHFM Type3*: A duplicated locus

***S**plit **h**and and **f**oot **m**alformation* (*SHFM*) is a term used to define heterogenous limb malformation that is characterized by a deep median cleft in hands and feet with the absence of the central rays. The genetic causes of this type of disorder have been mapped to several loci. Inheritance is mostly autosomal dominant, but forms with autosomal recessive and X-linked inheritance fashion have been described (reviewed in (Duijf 2003)). The severity of the phenotype is variable between individuals but also between limbs of a single individual (see figure 10a for the phenotype). *SHFM Type3* has been mapped to human chromosome 10q24, which is syntenic to the region identified for the mouse *Dactylaplasia* mutants (Nunes et al. 1995), (Ozen et al. 1999). In several *SHFM3* patients, tandem duplications covering ~500kb of this region have been identified (comprising genes from *TLX1* to *FBXW4*) (de Mollerat et al. 2003). Study of additional patients and mapping of the break points of the corresponding duplications enabled to define a minimal 325kb region that is duplicated in all patients. This interval comprises *βTRC, POLλ, DPCD* genes and 3'UTR of gene. In most patients, the telomeric break point is localized within *FBXW4*, while the centromeric one is more variable with the presence or absence of *LBX1* gene in the duplicated region. (see Figure 10b and 10c for identification of breakpoints) (Lyle et al. 2006). The genes present in the regions code for proteins with different biological functions: an NK-like subfamily homeobox transcription factor, *Tlx1*; a ladybird homeobox protein homolog required for muscle cell precursor migration, *Lbx1*; a component of E3 ubiquitin-protein ligase complex, *βTrcp*; a DNA polymerase, *Polλ*; an uncharacterized protein (potentially having a role in the formation or function of ciliated cells), *Dpcd*; a recognition protein for ubiquitin mediated protein degradation machinery, as well as afore mentioned *Fgf8* (see Figure 9 and 10 for the locus).

The consequences of this duplication on gene expression have been analyzed in lymphoblastoid cells of *SHFM3* patients. People observed an increase in *βTRC* (probably a simple dosage effect resulting from duplications) and *SUFU* (located further downstream of *FGF8* gene) expression levels (Lyle, Radhakrishna et al. 2006). *βTRC* is involved in the canonical Wnt signaling pathway regulating β-catenin levels through ubiquitin-mediated degradation and conditional loss of β-catenin results in loss of the AER (Barrow, Thomas et al. 2003), similarly *SUFU* regulates β-catenin signaling negatively (Meng et al. 2001). It is difficult to conclude whether the above-mentioned findings are biologically significant or not, since these

experiments were carried out using a cell type that is not directly relevant to the phenotype associated with the disease. However, considering the role of Wnt-signaling in AER maintenance, detected gene expression changes show a strong correlation with the phenotype.

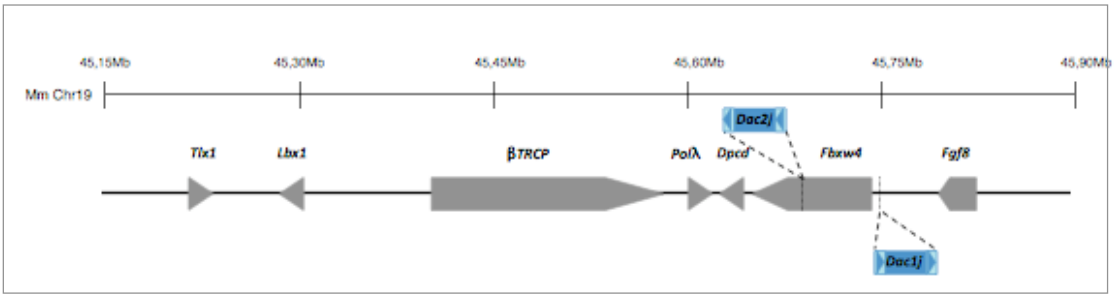


Figure 9. MusD element insertions in *Mus musculus* chromose19
(The chromosome positions are indicated according to Mouse July 2007 assembly)

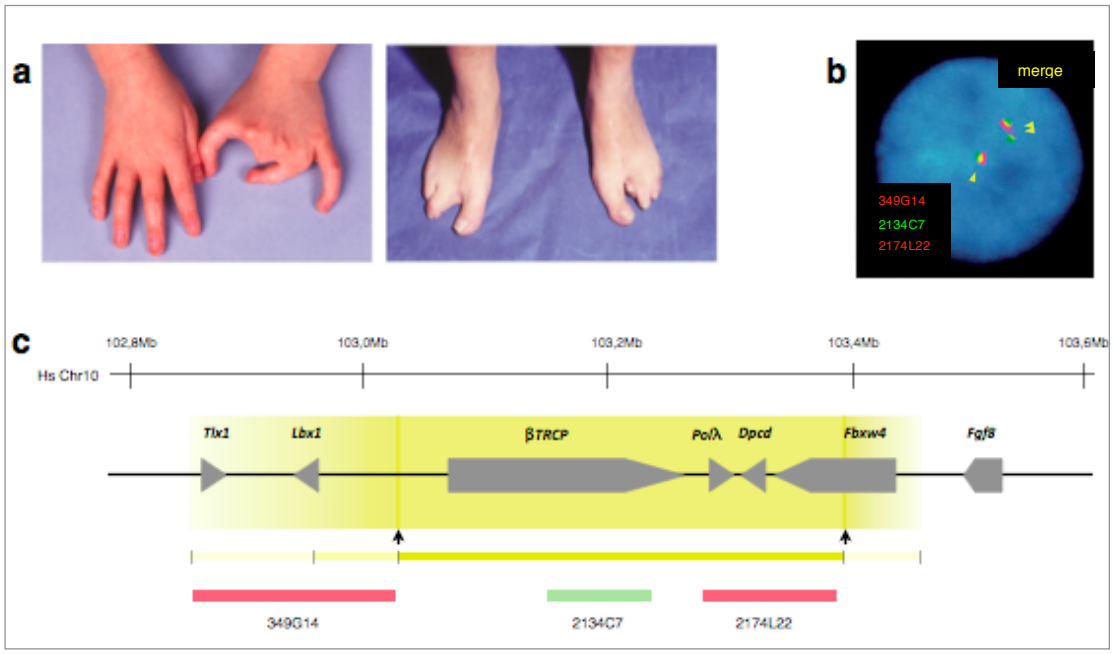


Figure 10. Human *SHFM3* caused by tandem duplication in the syntenic locus of *Dactylaplasia* mouse. **a)** Severity of the phenotype is variable between hands and feet of an affected individual. (taken from Duijf, P.,*Pathogenesis of split-hand/split-foot malformation.* Human Molecular Genetics, 2003. **12**(90001): p.51R-60.) **b** and **c)** Duplication is shown by probes on a sample from a patient. Breakpoints of the duplication and the frequencies are color coded by a darker yellow shade for an increased frequency.(adapted from Lyle, R., et al., *Split-hand/split-foot malformation 3 (SHFM3) at 10q24, development of rapid diagnostic methods and gene expression from the region.*Am J Med Genet A, 2006.**140**(13): p.1384-95.)

As mentioned earlier, several loci have been mapped for other forms of ectrodactyly. Interestingly, in one large family affected with split-hand/foot malformation, a missense mutation of *WNT10b* gene has been identified (Ugur and Tolun 2008), further suggesting a relationship between Wnt pathway and *SHFM*. However, the other known *SHFM* loci are pointing to the involvement of various genetic pathways. In a large number of cases, mutations of the transcription factor p63 have been identified. This gene is however involved in different malformation syndromes, some including limb ectrodactyly such as ectodermal dysplasia and cleft lip/palate syndrome (EEC, OMIM 604292) and limb mammary syndrome (LMS, OMIM 603543), other without limb malformations such as ankyloblepharon-ectodermal defects-cleft lip/palate syndrome (AEC, OMIM 106260) and Acro-dermato-ungual-lacrimal-tooth syndrome (ADULT, OMIM 103285) (Celli et al. 1999), (van Bokhoven et al. 2001), (Rinne et al. 2007) and *p63* deficient mice exhibit limb, craniofacial and ectodermal abnormalities (Mills et al. 1999), (Yang et al. 1999). The defects are primarily due to the key involvement of p63 in ectodermal differentiation and maintenance of the progenitor cells required for epithelial development. The *p63* gene produces multiple isoforms, through alternative promoter usage and splicing, with different transcriptional functions. This complexity could contribute to the variable spectrum associated with p63 mutations. Interestingly, some genotype-phenotype correlations could be identified with mutations leading to limb defects affecting primarily the DNA binding domain (Celli, Duijf et al. 1999), (Ianakiev et al. 2000). Transcription factors from DLX family that are expressed in the AER also have been associated with *SHFM* syndromes. These genes are organized in bigenic clusters and deletions removing *DLX5* and *DLX6* or putative *cis*-regulatory sequences have been found in *SHFM Type1* patients. Correspondingly, the double knock-out of *Dlx5* and *Dlx6* leads to ectrodactyly in mice (Robledo et al. 2002), (Merlo et al. 2002). Similarly, genetic mapping and functional similarities suggest that the *Dlx1-2* cluster could be a candidate gene for *SHFM5* (Duijf 2003). Interestingly, it was shown that *Dlx5* and *Dlx6* are direct transcriptional targets of p63 (Lo Iacono et al. 2008), suggesting that the p63/Dlx pathway could have a central role in the pathogenesis of *SHFM* syndromic and non-syndromic forms.

2.5.5 *Dactylaplasia vs. SHFM3*: one phenotype caused by two very different kinds of genomic mutations

The human disease and the mouse mutation are clearly caused by two very distinct mutation mechanisms, yet they result in the same limb malformation. There have been several hypothesis proposed to explain the phenotypes in human patients due to a duplication of a 300-500kb stretch and/or in mice due to insertions of ~7kb MusD elements. Importantly, careful quantification of gene copy numbers ruled out the possibility of an *SHFM3*-like duplication in *Dactylaplasia* mice (Friedli, Nikolaev et al. 2008). The duplications found in human patients have been proposed to cause a gene dosage imbalance (Lyle, Radhakrishna et al. 2006), similar to a mini-trisomy. Indeed, some genes of the locus (including some localized outside of the duplicated interval) showed mild up-regulation. In mice, the impaired expression of *Fbxw4* was thought to be responsible for the phenotype, even though the Dac^{1j} animals did not show any expression change for this gene (Sidow, Bulotsky et al. 1999), indicating its transcript levels are not essential for the phenotype.

An alternative hypothesis that I would further explore in this thesis is that both mutations could alter the action of remote gene regulatory elements. I have already detailed how TEs have a potential to impact gene regulatory mechanisms in multiple ways. Similarly, large chromosomal re-arrangements could lead to “position-effects” that could affect genes localized at a distance from the re-arrangement break points.

2.5.6 *cis*-regulatory mutations as the genetic cause of limb malformations

Deletions of regulatory elements or chromosomal re-arrangements involving regulatory sequences are known to be the cause of several disorders (reviewed in (Kleinjan and van Heyningen 2005)). This has been mostly demonstrated for re-arrangements such as deletions, inversions or translocations, which are moving critical enhancers away from their target genes, resulting in a tissue-specific loss-of-function of the corresponding gene. However, recent studies have exemplified that chromosomal duplications of regulatory elements could also lead to abnormal phenotypes. A microduplication in the ***S**onic **h**edge**h**og* (*Shh*) locus covering the **Z**PA **r**egulatory **s**equence, (ZRS) an enhancer for *Shh* posterior limb bud expression, (Lettice, Heaney et al. 2003)) causes triphalangeal thumb-polysyndactyly in humans (Klopocki et al. 2008), which is normally caused by the ectopic expression of *Shh* in the anterior region. Surely, *Shh* locus is not the only one affected by disruptions of long-range interactions. A microduplication of ~5.5kb conserved noncoding sequence

located ~110kb downstream of *BMP2* gene was detected as a causes of an autosomal-dominant brachydactyly type A2 (a limb malformation characterized by hypoplastic middle phalanges of the second and the fifth finger). *BMP2* ablation in the mouse limb does not cause a limb malformation; therefore, the underlying mechanism for the disease could be due to other means of gene expression changes, possibly an up-regulation of the BMP2 level (Dathe et al. 2009). Another example of a gene regulatory disruption is in *SOX9* (SRY-box containing gene 9) locus. *SOX9* is a dosage-sensitive, tissue-specific and temporally expressed transcription factor that is known to be involved in chondrogenesis through its interactions in Wnt–β-catenin pathway (Akiyama et al. 2002). Duplications in a minimal ~1.2Mb critical interval on chromosome 17q24.3 that is 5' of *SOX9* were found to cause brachydactyly and nail aplasia. This region encompasses a large gene desert between *KCNJ2* and *SOX9* involving putative regulatory elements of *SOX9,* suggesting a misexpression and/or overexpression of this gene as the reason for abnormal digit and nail development (Kurth et al. 2009).

2.5.7 Repression of MusD and consequences on *Dactylaplasia* phenotype

The *Dactylaplasia* mutation arose in a SM7B/SC strain stock in the Jackson Laboratory. Subsequently, the *Dactylaplasia* mice were crossed to SM/Ckc strain for a few generations to maintain the line, then breeding tests were initiated using SM7B/SC for intercrosses and different strains including SM/Ckc, LG/Ckc, C57BL/6J, DBA/2J, C3H/HeJ, SWR/J, AKR/J, BALB/cJ, and 129/J inbred strains for outcrosses. The outcrosses produced affected animals in the F_1 generation for crosses made with LG/Ckc, SM/Ckc, BALB/c, and 129/J but not with the crosses of remaining inbred strains. These criteria distinguished inbred strains as two groups: permissive and resistant to *Dactylaplasia* malformation. Intercrossing of F_1 generation suggested that the *Dactylaplasia* phenotype was genetically suppressed by a second locus (Chai 1981). This unlinked dominant modifier locus was called as ***M**odifier of **Dac**tylaplasia (Mdac)*.

Initially *Mdac* was mapped to a 27Mb region on mouse chromosome 13 by backcrossing hybrid strains and mapping for polymorphic microsatellite markers of origins (Johnson, Lane et al. 1995). The locus is refined to a 9.4Mb fragment containing 125 genes, including genes which are important for limb development such as *Ror2*, *Msx2*, *Fgfr4* and *Patched* (Kano, Kurahashi et al. 2007). In the same study a differential DNA methylation of the 5'LTR of the MusD inserted in *Dac^{1j}*

animals was found between resistant and permissive strains. Concomitantly, MusD expression was detected in the AER of mutant limb buds at E10.5, whereas this expression was not seen when the animals were crossed to C57BL/6J strain, suggesting that this expression could contribute to limb malformation.

3. Aims and Objectives

Ohno postulated that the generation of additional regulatory systems contributes to evolutionary changes more than the generation of new structural genes (Ohno 1972). Shortly after, King and Wilson demonstrated that the evolutionary changes between chimpanzees and humans could not be explained by sequence changes of proteins alone (King and Wilson 1975). Britten and Davidson proposed that the change and the increase of complexity in regulatory system as a consequence of repetitive sequence dispersion throughout the genome (Britten and Davidson 1971). As mentioned above, these models and hypotheses have received many recent experimental evidence which support that TEs could contribute to evolution of gene functions and modulation of gene expression. Yet, in many instances, the detailed mechanistic impact of TEs – apart from the disruption of elements – has not been really characterized.

The *Dactylaplasia* mutation is an excellent model system to study novel roles of transposable elements in gene regulation in mammals. The phenotypic resemblance between human patients and mutant mice makes this system especially intriguing, as it suggests that insertion of a small element could have the same impact as a much larger chromosomal re-arrangement disrupting an evolutionarily conserved locus. Moreover, the existence of a modifier locus (*Mdac*) that seems to contribute to the epigenetic control of the mutagenic TE in its inserted position opened possibilities of identifying new factors involved in controlling these elements and their activities.

In the course of this thesis, my first aim was to understand how MusD sequences influence interactions between genes and remote regulatory elements. In addition to the *Dactylaplasia* locus, I examined the *Olig2-Olig1* locus located on a different chromosome to see if the presence of a neighboring MusD could drive gene expression changes in multiple contexts. To investigate the mechanism associated with the regulatory impact of MusD elements, I developed *ex-vivo* approaches to test their silencing and insulating potential. To gain insights into the gene causing the cleft-limbs of *SHFM3* patients and *Dactylaplasia* mice, I used an *in-vivo* transgenic approach to examine the consequences of forced expression of candidate genes from this locus in the developing AER. This work is summarized in the first part of the “Results” section.

The two-locus model of *Dactylaplasia* was described more than 30 years ago, but the modifier gene and the mechanisms of its action have been enigmatic until today. Therefore, the second aim of this thesis was to identify the main player of this (potentially) MusD-specific silencing mechanism. Using genetic mapping, I narrowed down the *Mdac* locus to a 1.7Mb interval. I used different approaches to characterize this region in different strains and identified a polymorphic deletion corresponding to the Mdac/mdac condition. This characterization yielded potential candidates of transposon silencing. The deleted region in the *Dactylaplasia* permissive strains contains genes and pseudogenes of the KRAB Zinc Finger family proteins. Further functional analysis provided compelling evidence supporting the role of these genes in MusD silencing. This work is summarized in the second part of the “Results” section.

4. Materials and Methods

4.1 Materials

4.1.1 Instruments

Centrifuges

Name	Supplier
Microcentrifuge 5424	Eppendorf
Chilling centrifuge 5415 R	Eppendorf
Multifuge 3SRT	Thermo Scientific
Sorvall RC6T	Thermo Scientific

Thermo cyclers

Name	Supplier
PTC-200 DNA Engine Cycler	BIO-RAD
C1000 Thermal Cycler	BIO-RAD
ABI7500 Light cycler	Applied Biosystems

Microscopes

Name	Supplier
Leica MZ16F	Leica
Leica MZ16	Leica
Leica DM IL	Leica

Other

Name	Supplier
Mithras LB 940	Berthold
GenePulser Xcell	BIO-RAD

4.1.2 Chemicals

Unless indicated otherwise, all chemicals were supplied by Merck (Darmstadt) and Sigma-Aldrich (Steinheim).

4.1.3 Buffers

Unless specified otherwise, all solutions were prepared according to (Sambrook and William Russell 2006).

4.1.4 Kits

Name	Supplier
EPITECT Bisulfite Kit (Cat.No. 59104)	QIAGEN
QIAquick PCR Purification Kit (Cat.No. 28104)	QIAGEN
MinElute Gel extraction Kit (Cat.No. 28606)	QIAGEN
QIAEXII Gel extraction Kit (Cat.No. 20021)	QIAGEN
QIAprep Spin miniprep Kit (Cat.No. 27106)	QIAGEN
NucleoBond-XtraMidi /Maxi Prep (Cat.No. 740410.10/ 740416.10)	MACHEREY-NAGEL
illustra MicroSpin G-50 Columns (Cat.No. 27-5330-01)	GE Healthcare
pGEM-T Easy Vector System I (Cat.No. A1360)	Promega
Dual Luciferase Kit (Cat.No. E1960)	Promega
ProtoScript M-MuLV First Strand cDNA Synthesis Kit (Cat.No. E6300S)	NEB
PureLink RNA Micro Scale Kit (Cat.No. 12183016)	Invitrogen
PCR DIG Probe Synthesis Kit (Cat.No. 11636090910)	Roche
DIG RNA Labeling Kit (SP6/T7) (Cat.No. 11175025910)	Roche

4.1.5 Enzymes

Restriction enzymes were purchased from MBI Fermentas (St. Leon-Roth) or NEB.
DNA polymerase for genotyping was expressed from the construct provided by EMBL Protein Expression and Purification Core Facility. Other DNA polymerases which were used are: LA Taq (TaKaRa BIO Inc.), Long Range Expand (Roche), and Phusion (Finnzymes-Thermo Fisher Scientific).
Ligases and phosphatases were purchased from NEB.

4.1.6 Bacterial Strains

Name	Supplier
DH5α	Invitrogen
DH10B	Invitrogen
Stbl3	Invitrogen
JM110	DSMZ

4.1.7 Plasmids

Name	Supplier
pGEMT-easy	Promega
pGL4.23	Promega
pGL3	Promega
pRL-SV40	Promega
pCAGGS-Puro linker	M. Treier

4.1.8 Oligos

All oligos were purchased from Sigma-Aldrich.

4.1.8.1 For genotyping

Name	5'->3' sequence	target	# in DB
Dac2j_5'_R	ACTTTAAACCCTTTCTTCTTCCACCT	Dac2j/Dac1jMusD	164
Dac2j_5'F_UT	ATTCACCACTTTCTCACAAGGGTGGGA	Dac2j MusD	171
Dac2j_3'R_UT	TCCCTGAACCCCTCACTCTATCCCTCA	Dac2j MusD	173
Dac1_3'R_UT	CCTGACTTGAGATGTAACCATAAAT	Dac1j MusD	170
Dac1jGNTYP-fln	GACATTGAATTGAGAAGCTTCACTTAATAG	Dac1j MusD	468
Dac2j_3'F_UT	TTTCAGTTGTTTACTAGAAAGGACAGT	Olig2-1MusD	172
MusDolig2_1	GGTTCCAAGAAGCAGCTCGAAG	Olig2-1MusD	305
MusDolig1_1	AACACCAGACCTCAGCTTGACTTC	Olig2-1MusD	306
lentiWPRE5´	CCCGTATGGCTTTCATTTTCTCC	WPRE	92
lentiWPRE3´rev	AAGGAAGGTCCGCTGGATTGAG	WPRE	93

4.1.8.2 For Bisulfite sequencing

Name	5'->3' sequence	target	# in DB
Dac2j-nested-5'F	TGGTGTTTTTATTAGGTATTTGTGA	Dac2j-MusD 5'LTR	183
Dac2j-nested-5'R	CTAATATTTCTTCTTCCTTAAACCA	Dac2j-MusD 5'LTR	184
Dac2j_5'_F	ATTTATTATTTTTTTATAAGGGTGGGA	Dac2j-MusD 5'LTR	163
Dac2j_5'_R	ACTTTAAACCCTTTCTTCTTCCACCT	Dac2j-MusD 5'LTR	164
Dac2j-nested-3'F	GGATGAGAAAATTATTTGATTATTT	Dac2j-MusD 3'LTR	185
Dac2j-nested-3'R	CACACATTAAATACAAAAATCAA	Dac2j-MusD 3'LTR	186
Dac2j_3'_F	TTTTAGTTGTTTATTAGAAAGGATAGT	Dac2j-MusD 3'LTR	165
Dac2j_3'_R	TCCCTAAACCCCTCACTCTATCCCTCA	Dac2j-MusD 3'LTR	166
Olig-5LTR-out-F	TTTAAGGTGATTTAGAATTAATTAG	Olig2-1MusD 5'LTR	351
Olig-5LTR-out-R	TTCCACAATCTAATATTTCTTCTTCC	Olig2-1MusD 5'LTR	352
Olig-5LTR-in-F	GATTTGAGGATGGATTAGATTTTGTGG	Olig2-1MusD 5'LTR	353
Olig-5LTR-in-R	CTTTAAACCCTTTCTTCTTCCACCTAA	Olig2-1MusD 5'LTR	354
Olig-3fln-in-R	AAACATTCTCAAAAATATCCTCT	Olig2-1MusD 3LTR	358
Olig-3LTR-out-F	GGATGAGGGATGAGAAAATTATTTGATTATTT	Olig2-1MusD 3'LTR	359
Olig-3LTR-in-F	AGTTGTTTATTAGAAAGGATAGTT	Olig2-1MusD 3'LTR	360
Olig-3LTR-in-R	TATTTCTCTAAACCTTAAACTTAAA	Olig2-1MusD 3'LTR	361
ETnII-SH3bp4-out-F	GTGAAATATTTTTTTTTGGATGTTGG	ETnIIinSH3bp4 5'LTR	1053
ETnII-SH3bp4-in-F	GGTTTTTTTAGAATTTTTTTTTATAGG	ETnIIinSH3bp4 5'LTR	1054
ETnII-SH3bp4-out-R	CCCTTTCTTTTCTCACACCTCAT	ETnIIinSH3bp4 5'LTR	1055
ETnII-SH3bp4-in-R	CCATAACTTTAAACCCTTTCTTCTTTC	ETnIIinSH3bp4 5'LTR	1056
ETnII-Chr6-88_5-out-F	TTGATTTATGATTTTAAATTTTTTTTTTTTTTA	ETnIIonChr6 5'LTR	1057
ETnII-Chr6-88_5-in-F	GTTGGGTTAGTATTGATTTAA	ETnIIonChr6 5'LTR	1058
ETnII-Chr6-88_5-out-R	TCCAAAACTATAAAATCCAAT	ETnIIonChr6 5'LTR	1059
ETnII-Chr6-88_5-in-R	CAACACATAAATAACTCCATAAATATTTTA	ETnIIonChr6 5'LTR	1060
ETnII-Chr11_5-out-F	GGTTTGTTTTGGGTTAAATTTAGTG	ETnIIonChr11 5'LTR	1061
ETnII-Chr11_5-in-F	GATTTTAGTAAGGTAGTTGTAGT	ETnIIonChr11 5'LTR	1062
ETnII-Chr11_5-out-R	CCTTTCTTTTCTCACACCTCAT	ETnIIonChr11 5'LTR	1063
ETnII-Chr11_5-in-R	AACCCTTTCTTCTTCCACCTAA	ETnIIonChr11 5'LTR	1064
chr6:975-ETn2-5'LTR-out-F	ATAGAATATTTATTTATATAGATTGGAT	ETnIIonChr6_2 5'LTR	1173
chr6:975-ETn2-5'LTR-out-R	AAAATATATAAACTCTTCAAAAATAAAAA	ETnIIonChr6_2 5'LTR	1174
chr6:975-ETn2-5'LTR-in-F	AGATAAGATATAAAAGGTAATAGAGTAT	ETnIIonChr6_2 5'LTR	1175
chr6:975-ETn2-5'LTR-in-R	ATAAACTTATATAATCTCTCCTCCTCC	ETnIIonChr6_2 5'LTR	1176

4.1.8.3 For *Mdac* mapping

Name	5'->3' sequence	# in DB
snp54232729f	AAACTGCCTCTGTCTGAGAAGTGG	329
snp54232729r	GAACCCTGTTGCTAGAAACTCGG	330
snp55020106f	GGAGGTGCTTCCAAACAAACTG	331
snp55020106r	CATATTGTGGACTGCATCGTGG	332
snp76357515f	TAGGCAAGCACCATCCAGGAGTCC	333
snp76357515r	TGATGAAGGCATTTGTGGAAGCCC	334
snp52637261f	GCACTCTGTTTAACAGGGAGGAAG	339
snp52637261r	TGAGCCGTCCCTATTTTCCAG	340
snp52642361f	GGTTCTGGCAGTCTGTTCTGAAAG	341
snp52642361r	GCAACAAAGGAAAAGCCATCACAG	342
snp61819596f	CGCCAGCTCCATAAAGGTAACCTG	343
snp61819596r	TTCTCACTTGAGCCAAGCCACCTC	344
snp61814572f	GCCTGCTACAACAGTGGTTCTTTG	345
snp61814572r	TGTAGGAAACAATCAAGCCTCTGG	346
snp61822734f	TCATGGTGAGCATGGGACCCTAAC	390
snp61822734r	CCTGGAACGATATTTCCCCACAATC	391
snp61829035f	ACCTTATTCCCCCAACTCCGTTTG	392
snp61829035r	TCTCTTTGGTCAACCCAGGAGCTG	393
snp61817080f	TCTTCTCTGCCCTAGCAGGAAAC	481
snp61817080r	GCGTAAATGCAACTTTGAACCG	482
snp61815141f	AGAAGTCACAGCCTGAGGAGACAG	483
snp61815141r	AGCTTCACCGCATTCTTGGC	484
snp61819714f	CGGAAGCTATACAAAAGGAGACTGAATGT	486
snp61819714r	CGATGCTTACTGGCTGAGATGTGTGA	487
snp61823624f	ACAGAGCTCAAAAAGGCTGAAGACAAG	488
snp61823624r	TCTCCCTACTCATTCCCAGGAACAGAGTTC	489
snp61819045f	GGTGAGTAAGATTTGAAGCCAGAGC	501
snp61819045r	AGTGTCCTTTGCACCTGAGCAC	502

4.1.8.4 For *Mdac* locus structural variations characterization

Name	5'->3' sequence	# in DB
chr13:61730601-61731132F	GATACACTTCGGTAATTAGGATATGGA	607
chr13:61730601-61731132R	ATGAACTTGGAAAAGCAATTAATTTTA	608
chr13:61844624-61845085F	TAAAAGGTCTGATGCCTGTTTACATTT	609
chr13:61844624-61845085R	ACACAGTGAATATCTTCAGAGGGCATC	610
chr13:62004906-62005437 F	TCTCCTGCTTTTCTCCCAGGGTTGTCT	611
chr13:62004906-62005437 R	GGATGACCCTAGATTCAGATAGATGCA	612
chr13:62147244-62147995F	CCAGGCACCCAGGAACTCTGCCAATAT	613
chr13:62147244-62147995R	ATGCTGGGTCTACTCGCTTATGCATTT	614
chr13:62304685-62305178F	AAGTGGAGTAGGCCACAGAATTATATG	615
chr13:62304685-62305178R	TTACCTTCTCTGTCTCCTAGAACTTTG	616
chr13:62429958-62430522F	TCCCATGATCACAATATAGCAATTTAG	617
chr13:62429958-62430522R	AAGGAGAAGGAAAGTGAGCAGCAGAAG	618
chr13:62692278-62692812F	TTCATTGGATCAACCTAAATACCTCAA	619
chr13:62692278-62692812R	TCTTCAGAGGAGCAGTCCCGCTGTTGT	620
chr13:62889597-62890349F	GCCTGGAGGTATAACAGGTAGAATTAG	621
chr13:62889597-62890349R	ACATGTCCCAGCAACAGACTGATATAC	622
chr13:62013106-62013880F	CACTGCACTGCTGAGTTCTCCCTCCTG	639
chr13:62013106-62013880R	ACTACCGAGTCATGTTAGAAGGATACTTAA	640

4.1.8.5 For qPCR

Name	5'->3' sequence	target	# in DB
Olig1-RT-F	CCAAAGAGGAACAGCAGCAG	Olig1	1219
Olig1-RT-R	GTGGCAATCTTGGAGAGCTT	Olig1	1220
Olig2-RT-F	CACAGGAGGGACTGTGTCCT	Olig2	1221
Olig2-RT-R	GGTGCTGGAGGAAGATGACT	Olig2	1222
Fwd_SYBR_MmACTB	CTAAGGCCAACCGTGAAAAG	Actb	1319
Rev_SYBR_MmACTB	ACCAGAGGCATACAGGGACA	Actb	1320
Fwd_SYBR_MmPGK1	TACCTGCTGGCTGGATGG	Pgk1	1321
Rev_SYBR_MmPGK1	CACAGCCTCGGCATATTTCT	Pgk1	1322
Maksakova_ETnMusD514-s	GTGCTAACCCAACGCTGGTTC	ETn/MusD	2276
Maksakova_ETnII662-as	ACTGGGGCAATCCGCCTATTC	ETnII	2277
Maksakova_MusD690-as	CTCTGGCCTGAAACAACTCCTG	MusD	2278
Maksakova_ETnI-s	TGAGAAACGGCAAAGGATTTTTGGA	ETnI	2279
Maksakova_ETnI-as	ATTACCCAGCTCCTCACTGCTGA	ETnI	2280
MusD-qPCR-F	GATTGGTGGAAGTTTAGCTAGCAT	MusD	2723
MusD-qPCR-R	TAGCATTCTCATAAGCCAATTGCAT	MusD	2724
LINE-qPCR-F	TTTGGGACACAATGAAAGCA	LINE1	2725
LINE-qPCR-R	CTGCCGTCTACTCCTCTTGG	LINE1	2726
IAP-5'UTR-qPCR-F	CGGGTCGCGGTAATAAAGGT	IAP	2727

IAP-5'UTR-qPCR-R	ACTCTCGTTCCCCAGCTGAA	IAP	2728
Nanog-F	GCAAGCGGTGGCAGAAAA	Nanog	2922
Nanog-R	GGTGCTGAGCCCTTCTGAATC	Nanog	2923
Oct4-F	TGGCGTGGAGACTTTGCA	Oct4	2924
Oct4-R	GAGGTTCCCTCTGAGTTGCTTTC	Oct4	2925
1214-2-F	GAACATTGTCAAAGTTCTAGAAGAAACAGA	Tromer	2926
1214-2-R	GACTGTGAGCACATAAAGCAAAGGCT	1214-2	2927

4.1.9 Cultured cell lines

Name	**Origin**
HEK293T	Human embryonic kidney
NIH3T3	Swiss mouse immortalized fibroblasts
E14 ES cells	129/Ola mouse

4.10 Antibodies

4.10.1 Primary Antibodies

Antigen	**Source animal**	**Supplier/ Cat.No.**
HA	Mouse	Sigma-Aldrich/H3663
FLAG	Mouse	Sigma-Aldrich/F1804
Tubulin	Mouse	Sigma-Aldrich/T9026

4.10.2 Secondary Antibodies

Antigen	**Conjugated**	**Supplier/ Cat.No.**
α-mouse	HRP	GE Healthcare/NA931VS

4.11 Animals

Breeding and crossings were performed by Silke Feller and Andrea Schulz in the animal facility (LAR) of EMBL.

Dac1j animals were maintained from SM/Ckc-Fbxw4^{Dac}/J Jackson Laboratory line by crossing to BalbCj and Dac2j animals were from CBy.MRL-*Fbxw4*$^{Dac\text{-}2J}$/J of Jackson Laboratory.

4.12 Software

Program	Producer	Application
BIQ Analyzer	Max Planck Institut-Informatik	Bisulfite sequencing analysis
MacVector 11.0.2	MacVector Inc.	Sequence analysis, comparison
ABI7500 Software V2.0.5	Applied Biosystems	Analysis of qPCR data
ApplicationSuiteV3	Leica	To acquire embryo photos
MicroWin 2000	Berthold	To acquire luciferase reads

4.13 Internet Resources

Resource	Address
UCSC Genome Browser	http://genome.ucsc.edu/
Mouse Phenome Database (MPD, Jackson Lab)	http://phenome.jax.org/db
Mouse Genome Informatics (MGI, Jackson Lab)	http://www.informatics.jax.org/
National Center of Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/
BiSearch Primer Design Tool	http://bisearch.enzim.hu
Simple Modular Architecture Research Tool (SMART)	http://smart.embl-heidelberg.de/
Inter*activate*	http://www.shodor.org/interactivate/activities
GraphPad	http://www.graphpad.com/quickcalcs

4.2 Methods

4.2.1 Molecular Biology Methods

4.2.1.1 DNA Isolation

4.2.1.1.1 Plasmid or BAC DNA Isolation

Plasmid-DNA was isolated with a QIAprep Spin Miniprep Kit (for small scale) or the Nucleobond Xtra Midi/Maxi (for medium/maxi scale) according to the manufacturer's specifications. BAC DNA was isolatated using Nucleobond Xtra Maxi Kit using low copy plasmid purification protocol described in user manual provided.

4.2.1.1.2 Genomic DNA Isolation

Low Purity Isolation (from tails or membranes for genotyping PCR): Tails or membranes were lysed in 150-200μL of tail lysis buffer-ProteinaseK mix (ProteinaseK was diluted 1:100 from 10mg/mL stock solution) over-night at 56ºC.

Tail Lysis buffer: 50mM KCl, 5mM Tris pH8.0, 2mM $MgCl_2$, 0.1%w/v Gelatin, 0.45% v/v NP40 and Tween-20.

High Purity Isolation: Tissues or cells were lysed in 150-500μL of lysis buffer-ProteinaseK mix (ProteinaseK was diluted 1:100 from 10mg/mL stock solution) over-night at 56ºC. The lysate was mixed with 70% of its volume isopropanol and DNA was pelleted. Then the DNA pellet was washed with 70% ethanol.

Lysis buffer: 100mM Tris pH8.5, 5mM EDTA pH8.0, 0.2% SDS, 200mM NaCl.

4.2.1.2 RNA Isolation

RNA was isolated from embryonic tissues or cell pellets. Embryonic tissues (Brain, Neural Tube and limbs) were collected from either embryonic day E11.0 or E14.5 and were homogenized in TriZol reagent (volume of TriZol depending on the weight of the tissue, 10μL /mg). Cell pellets were obtained from trypsinization of adherent cells followed by two times PBS wash and were homogenized in TriZol (10μL/10^5 cells). After homogenization, tissue/cells were incubated at room temperature for 5 minutes, then mixed with Chloroform (amount determined by the 20% of the TriZol volume) by shaking the tube vigorously. Chloroform-homogenate mix was incubated at room temperature for 3 more minutes and then spun at 12,000g for 15 minutes at 4ºC. The aqueous phase was collected and mixed with an equal volume of 70% ethanol. The ethanol and nucleic acids mix was then loaded on to PureLink columns

and clean-up procedure was followed according to manufacturer's directions. On-column DNaseI treatment was performed using PureLink DNaseI. RNA was eluted in 30μL Nuclease free water.

4.2.1.3 cDNA Synthesis

First strand cDNA was synthesized using Protoscript MuMLV Kit. 500-1000ng of total RNA was used as template for random priming.

4.2.1.4 PCR

4.2.1.4.1 Standard PCR

For genotyping:

DNA amplification was performed with Taq-polymerase that was produced from the construct provided by the Protein Expression Purification Core Facility of the institute. The reagents were pipetted into a chilled 0.2 ml reaction tube and incubated in a thermocycler with following program:

Phase	**Temperature**	**Time (minutes:seconds)**	**Cycles**
Initial denaturation	94ºC	4:00	
Denaturation	94ºC	0:30	
Primer annealing	XºC*	0:30	
Elongation	72ºC	1:00	30
Final elongation	72ºC	10:00	
Hold	10ºC	∞	

*For *Dac*^{1j}-MusD, *Dac*^{2j}-MusD and *Olig2,1*-MusD genotyping X=56ºC

*For WPRE genotyping X=62ºC

Product sizes:

Dac^{1j}-MusD: WT allele is 974bp and MusD allele is 675bp

Dac^{2j}-MusD: WT allele is 378bp and MusD allele is 713bp

Olig2,1-MusD: WT(BalbC) allele is bp and MusD allele is 569bp

WPRE: 362bp

Reagent	Amount
DNA template	20-100ng of genomic DNA
10X PCR buffer	2μL
25mM dNTP mix	0.16μL
10μM forward primer	0.5μL
10μM reverse primer	0.5μL
Taq Polymerase	0.5μL
ddH_2O	15.34μL

For Mdac mapping using SNPs (followed by restriction enzyme digestion):
Densely available single nucleotide polymorphisms (SNPs) were selected from the SNP-database (Perlegen) of the Jackson Laboratory. The same PCR program conditions were used with the same reagent concentration set-up as described in 4.2.1.4.1, then 10μL of the product was digested with a restriction enzyme to distinguish the origin of the allele via SNPs resulting in a restriction enzyme site polymorphism. The selected enzymes for particular SNPs and the expected product sizes are as follows:

Primer pair#	SNP ID	Position on Chr13	Product size	cut in C57BL6J	Sizes after cut	Annealing Temperature
1	49111842	56852545	515 nt	EcoRI	361+154	58ºC
2	51082153	34203888	496 nt	BamHI	389+107	58ºC
3	51146017	103907426	540 nt	BamHI	306+234	58ºC
4	52656624	70427030	429 nt	BamHI	306+123	58ºC
5	61794681	50505420	490 nt	MunI	378+112	58ºC
6	61805475	53450102	455 nt	BamHI	293+162	58ºC
7	61827471	60357611	494 nt	BamHI	371+123	58ºC
8	61801304	54232729	539 nt	PvuII	359+179	58ºC
9	61807097	55020106	657 nt	PvuII	332+324	58ºC
10	52665098	76357515	543 nt	PvuII	246+296	58ºC
11	52637261	57562487	561 nt	KpnI	302+258	57ºC
12	52642361	59071041	585 nt	PvuII	291+293	56ºC
13	61819596	62881702	567 nt	HindIII	292+275	56ºC
14	61814572	64219061	517 nt	HindIII	264+253	56ºC
15	61822734	61653221	562 nt	EcoRV	267+295	56ºC
16	61829035	64837047	531 nt	BglII	271+260	56ºC
17	61817080	63441010	525 nt	BglII	259+266	58ºC

18	61815141	63984139	521 nt	PvuII	300+221	58ºC
19	61819714	62859021	592 nt	BglII	293+299	58ºC
20	61823624	61302220	600 nt	BamHI	151+449	58ºC
21	61819045	62981912	538 nt	SpeI	302+236	58ºC

4.2.1.4.2 Amplification of genes for expression constructs

High fidelity Phusion enzyme was used as the polymerase, PCR conditions were set according to user manual (for genomic DNA, 30 seconds elongation / 1000bp to be amplified, for low complexity DNA (such as plasmid and BAC), 15 seconds elongation / 1000bp to be amplified)

4.2.1.4.3 Amplification of sequences for probe preparation

Southern Blot probes were produced with Roche PCR DIG Labeling Kit (Cat.No. 11636090910) from 100pg-1ng plasmid DNA template with the suggested PCR program in the user manual.

4.2.1.4.4 Quantitative real-time PCR (qPCR)

SYBR-Green binds specifically to double stranded DNA and can be used for quantification of DNA amplification from a template in real time. Suitable primers were designed targeting a sequence at the exon/intron boundaries whenever possible, to exclude signal variation due to genomic DNA contamination. The reaction was carried out in 96-well plates in a volume of 20μl containing 1μL of cDNA (prepared as described in section 4.2.1.3), 1 μmol of forward and reverse primer mix and 50% 2x SYBR Green PCR Master Mix (Applied Biosystems). Gene expression analysis was performed by normalizing the test gene to multiple controls (Actb, Pgk1 and Gusb), using the 7500software V2.0.5 for analysis (Applied Biosystems).

4.2.1.5 Sanger Sequencing

Sequencing of plasmids or PCR products were performed by GATC Biotech. For plasmid sequencing on 96-wells, bacteria colonies were stabbed into ampicillin selection carrying agar media plates.

4.2.1.6 Southern Blot

1-2 μg of genomic DNA was digested over night at 37ºC with the restriction enzyme combinations PstI+XhoI, PstI+NotI and PstI+AfeI to distinguish between methylated and unmethylated positions of MusD/ETnII transposons (NotI and AfeI activities are

blocked by CpG methylation). Digested fragments were run on a 1% Agarose (ethidium bromide stained) gel for at least 5 hours at 100Volts. After bands on the gel were imaged, gel was denatured (using denaturation solution: 0.5N NaOH, 1.5M NaCl) for 30 minutes. Gel was neutralized (using neutralization buffer: 0.5M Tris-Base, 1.5M NaCl, pH7.2-7.4) until pH strips showed a value around 7.0-7.5 exchanging the solution every 30 minutes. Capillary transfer was set on a nitrocellulose membrane (Amersham Hybond+) overnight by using 10X SSC, pH7.0 (diluted 1:2 from 20X SSC stock: 3M NaCl, 1M Sodium citrate). The next day membrane is washed with 50mM NaPi, pH7.2 for five minutes, then baked/fixed at 80ºC in the oven for two hours. Pre-hybridization was performed in the hybridization buffer (0.5M NaPi pH7.2, 7% SDS, 1mM EDTA pH8.0) for at least thirty minutes at 65ºC(however up to four hours most of the time) then membrane was hybridized with probe (5ng/mL) containing hybridization buffer over night at 65ºC rotating. After over night probe hybridization the membrane was washed twice in Church wash buffer (0.08M NaPi pH7.2, 1%SDS) for ten minutes, then blocked in blocking buffer ,DIG2, (1xDIG1 buffer (0.1M Maleic acid, 0.15M NaCl) with Roche Blocking reagent (Cat.No.11096176001)) for thirty minutes and incubated with DIG antibody (Roche Cat.No.11093274910, used 1:20000) for thirty minutes more. After antibody incubation, the membrane was washed twice in 0.1% Tween-20 containing DIG1 buffer for twenty minutes. Once the unbound antibody was removed, the membrane was washed in DIG3 buffer (0.1M Tris pH9.5, 0.1M NaCl) for five minutes and the signal was detected by 6μL/mL CDP-star reagent (Tropix Cat.No.T2306-0705036 MSC050) in DIG3. The exposure to the X-Ray film varied from 1 hour to over night.

4.2.1.7 Bisulfite sequencing

4.2.1.7.1 Chemical treatment and clean-up

For treatment and cleaning up the treated genomic DNA, QIAGEN EPITECT kit was used according to manufacturer's instructions. 1000ng of genomic DNA was treated and eluted twice in 20μL water after cleaned up.

4.2.1.7.2 PCR set-up and sequencing

Nested PCR was performed using the listed primers in section 4.1.8.2 and LA-Taq (TaKaRa) on 0.7μL of second elution from bisulfite treatment clean up (section 4.2.1.7.1) with a PCR program as follows:

Phase	Temperature	Time (minutes:seconds)	Cycles
Initial denaturation	94ºC	4:00	
Primer annealing	55ºC	2:00	
Elongation	68ºC	2:00	2
Denaturation	94ºC	1:00	
Primer annealing	55ºC*	1:00	
Elongation	68ºC	2:00	35
Final elongation	72ºC	10:00	
Hold	10ºC	∞	
PRIMARY PCR			

Phase	Temperature	Time (minutes:seconds)	Cycles
Initial denaturation	94ºC	4:00	
Denaturation	94ºC	1:00	
Primer annealing	55ºC*	1:30	
Elongation	68ºC	2:00	35
Final elongation	72ºC	10:00	
Hold	10ºC	∞	
SECONDARY PCR			

*50ºC was used for 3'LTR amplification

4.2.1.8 BAC targeting

The BAC clone which was desired to be modified is transformed into EL250 bacterial strain and this strain is heat induced for the expression of recombinase. The competent cells were made freshly from induced bacteria and the targeting construct was electroporated into them. Targeting construct is designed with 50bp long homology arms for the site to be targeted in the BAC clone and contained neomycine selection under both prokaryotic and eukaryotic expression promoters.

4.2.2 Cell culture

4.2.2.1 Culturing conditions

HEK293T and NIH3T3 cells were generally cultivated in DMEM (high glucose, with glutamine, Gibco Cat.No.41965) supplemented with 10% heat-inactivated SerumSupreme (Lonza BioWhittaker, Cat.No.BW14-492F), 1% L-glutamine, 1% penicillin/streptavidin (Gibco, Cat.No. 25030-081 and 15070-063 respectively), which was warmed up to 37°C prior to cell seeding of feeding. The cells were kept in a humidified incubator at 37°C and 5% CO_2 and medium cells were passed every second or third day. E14 ES cells were cultivated in DMEM (Gibco, Cat.No.41965) with 15% FBS (PAN Biotech GmBH, Cat.No.2602), 1% L-glutamine (Gibco, Cat.No.25030-081), 1% penicillin/streptavidin (Gibco, Cat.No.15070-063), 1% Non-essential amino acids (Gibco, Cat.No.11140-050)), 1% Sodium Pyruvate (Gibco, Cat.No. 11360-070), 1% 2-mercaptoethanol (diluted 1:100 from the stock solution which was made using 35μL of Sigma-Aldrich M7522 in 50mL water stock solution) and 1000U/ml of Leukaemia inhibitory factor (LIF, Chemicon, ESG1107). ES cells were plated on feeder containing gelatinized (0.1% gelatin incubated on plates for one hour) plates and were fed every 24 hours.

4.2.2.2 Thawing cells

The cells were thawed quickly in a 37°C water bath and immediately transferred into a 15mL polypropylene tubes containing 4mL standard medium. After the suspension was centrifuged at 1300 rpm for 5 min, the cells were resuspended in 10 mL fresh medium and plated on 10cm cell culture plates (Nunc).

4.2.2.3 Splitting cells

When the cultured cells reached 80-90% confluency, the cells were split into a new culture plates. For this purpose, HEK293 and NIH3T3 cells were were washed with PBS and trypsinized for three minutes (Trypsin-EDTA, Sigma, Cat.No. T3924) at 37°C until they detached from the plates. Immediately after, the cells were taken up in medium, centrifuged, and resupended 1:6 in fresh medium. ES cells were trypsinized for five minutes and plated 1:4 on fresh feeder containing plates.

4.2.2.4 Cryopreservation

All types of cells were frozen in 1mL of freezing medium (90% serum, 10%DMSO) in Mr.Frosty (Nalgene Cat.No.5100-0001) that ensures ~1ºC change per minute, at minus 80ºC.

4.2.2.5 Cell number determination

When a defined cell number was required for an experiment, cells were counted in Neubauer-chamber (four of the middle-sized squares around four sides of the central big square, in total sixteen squares are counted and divided by four which gives the number of cells in 1mL when multiplied by 10^4) and plated out accordingly.

4.2.2.6 Cell transfection

4.2.2.6.1 Transient transfection

Cells were seeded one day in advance to transfection with Lipofectamine-2000 reagent (Invitrogen), which was used as described in user's manual, in 1:6 (DNA to Lipofectamine) ratio for HEK293T cells and 1:3 ratio for NIH3T3 cells (70-80% confluency of both cell types at time of transfection). Cells were harvested for luciferase assay (see section 4.2.3.2) twenty four hours after transfection.

4.2.2.6.2 Stable transfection

HEK293T cells were transfected with expression constructs under the control of CAG promoter, using Lipofectamine-2000 reagent (described in section 4.2.2.6.1). Cells were split in two different ratios (1:5 and 1:10) 24 hours after transfection and Puromycine selection (2.5μg/mL) was performed 24 hours after cell splitting for four days. Then the surviving population was further controlled for expression of the transfected gene, using Western Bloting (as explained in section 4.2.3.1).

4.2.2.7 Cell electroporation

ES cells were grown to 70-80% confluency and fed four hours prior to electroporation. Trypsinized $2x10^6$ ES cells were electroporated in ES cell electroporation buffer (Millipore) with 10μg of linearized BAC clone under 240Volts, 500μFarad in 4milimeter cuvettes. After electroporation cells were plated on feeder containing plates. 250μg/mL Neomycine (G418, Gibco, Cat.No. 10131-027) selection was performed 48 hours after electroporation for 7 days. Resistant colonies were picked, expanded and screened further, using PCR.

4.2.2.8 Lentivirus production

4.2.2.8.1 Cell transfection for virus production

10 times 10 cm dishes with 4 x 10^6 HEK 293 T cells were plated in order to have 70-80 % confluency after 24 hours of growing. 2 hours before transfection the medium was exchanged with the fresh medium. (Transfection reagents were provided by Clontech (Mammalian Transfection kit (Cat.No.631312)))

(the concentration of each plasmid: 1 μg/μl)

100μL of plasmid with gene of interest was mixed with 35μL of pMD2G plasmid (envelope plasmid), 65μL of pCMVR8.74 plasmid (packing construct) then 620μL of 2M $CaCl_2$ solution and 4180μlL water. Everything was mixed by vortexing slowly and while mixing the tube on the vortex, 5000μL of 2X HBS was added drop wise. This mixture was incubated at room temperature for 20 minutes. The incubated mix was added drop wise to HEK293T plates (1mL per plate).

4.2.2.8.2 Virus production

After cells were incubated at 37^o C for 24 hours, the medium was changed to OPTI-MEM. Media were collected at 48 hours and 72 hours after cell transfection and pooled. The supernatant was filtered through Millipore Stericup 0.22μm to remove dead cells. Maximum 60mL of the filtered medium was transferred into Centricon Plus 70 centrifugal Filter Devices (Millipore UFC710008) and centrifuged at 2500 g for 15 to 20 minutes. The flow-through was discarded and the remaining volume of the medium was transferred to the same filter for the next centrifugation. Finally, the virus particles were collected from the filter with the leftover of the medium/buffer. The amount of concentrated virus was between 200 and 300 μL. To purify the virus as much as possible, an additional centrifugation step was performed using Ultrafree MC Filter Device (Millipore). The virus was aliquoted and frozen initially at -20ºC and when it was transferred to -80ºC.

4.2.2.8.3 Viral transduction of primary cell lines

10^8 infectious particles per milliliter containing virus was spread on mouse embryonic fibroblasts (MEF) or ES cells which are 40-50% confluent. Then cells were kept in culture for 72 hours to enable the gene expression from the transfected virus, then harvested for RNA, DNA and protein preparations.

4.2.3 Biochemical Methods

4.2.3.1 Western Blotting

4.2.3.1.1 Gel run and transfer

Invitrogen precast gel system, NuPAGE, was used for SDS-PAGE. The 4-12% Bis-Tris gradient gel was loaded with samples (cell pellets) prepared (resuspended) in NuPAGE sample buffer (diluted from 4X stock and DTT added to a final concentration of 20mM as the reducing reagent, then denatured at 75ºC for 10 minutes). Gels were run according to manufacturer's specifications (NuPAGE MOPS running buffer was used). Transfer was performed to PVDF membranes (Immobilion, Millipore) with Invitrogen Xcell Lock blot module according to manufacturer's directions (NuPAGE transfer buffer was used).

4.2.3.1.2 Detection

After blotting, the membranes were blocked in 5% milk containing PBS +0.03% Tween-20 (PBST) mix for at least one hour at room temperature. After blocking membranes were incubated overnight with primary antibody (section 4.10.1, for HA and FLAG antibodies 1:1000, and for Tubulin antibody 1:10000 dilution was used) in blocking solution at 4°C. The next day, the membranes were washed three times for 5 minutes in PBST. After washing the unbound primary antibody, membranes were incubated with a suitable HRP-coupled secondary antibody for one hour at RT. Finally, the membranes were washed and protein was visualized with enhanced chemoluminescence (ECL) solution (Immobilion, Millipore) on a developed X-Ray film (Kodak RP X-OMAT processor).

4.2.3.2 Luciferase assay

Dual Luciferase Kit (Promega) was used according to provided manual by the producer. The delay step was removed from the measuring program and 100μL of LARII and Stop&Glo substrates were used per well of a 96-well plate. Both reaction was read for one second.

4.2.4 Animal sample preparation

4.2.3.1 Embryo dissection

Gestating mice were sacrificed at stage E10.5 and the uteri were removed. Embryos were dissected and placed in 4% PFA/PBS for overnight fixation at 4°C. If necessary, the embryonic amnions were collected and processed for genotyping. The next day embryos were washed in 0.1% Tween-20 containing PBS (PBS-T)

three times for 5minutes, then dehydrated gradually in 30%, 50%, 70% and 100% methanol containing PBS-T. Dehydrated embryos were stored in -20ºC.

4.2.3.2 RNA *in-situ* hybridization (ISH)

4.2.3.2.1 generation of DIG-labeled probes

Probe sequences were cloned into pSKII(+) vector, which contains either a T7 or T3 recognition sequence at each side of insertion. Depending on probe orientation in the vector, either T3- or T7-RNA polymerases were used for the in vitro transcription (on linearized and Klenow-blunt plasmids) for digoxygenin-labelled complementary probe synthesis, which was conducted with DIG RNA Labelling Kit (Roche) according to the manufacturer's protocol. After the reaction was stopped, the RNA was cleaned with a G-50 column (GE Healthcare) and eluted in ~100 μL RNase-free water. The success of the reaction was tested with an agarose gel and the probes were stored at - 20°C until further use.

4.2.3.2.2 Whole mount ISH

Embryos were re-hydrated and then washed three times in PBS-T for five minutes. After washing, the embryos were bleached for an hour in 6% H_2O_2 (diluted from 30% stock in PBS-T) and washed three times in PBS-T for five minutes. Bleached embryos were permeabilized by Proteinase-K treatment at room temperature for five minutes. The treatment was stopped by washing in 2mg/mL Glycine solution on ice, then washed three times in PBS-T for five minutes. The embryos are post-fixed in 4% PFA for 20 minutes at room temperature and washed five times in PBS-T for five minutes. After PBS-T washes, embryos are washed in W1 (5X SSC pH4.5, 50% ionized Formamide, 1% SDS, 0.1% Tween-20) for ten minutes at 65ºC and W1 was replaced by H2 (W1+ 5mg/mL Torula yeast RNA and 25μL of 100mg/mL Heparin) in order to pre-hybridize for at least 2 hours at 65ºC, then the probe containing H2 was incubated overnight with embryos rocking (in tubes) at 65ºC. For second day washes pre-heat the wash buffers. Embryos were washed three times in W1 and three times in W2 (2X SSC pH4.5, 50% ionized Formamide, 0.1% Tween-20) for thirty minutes at 65ºC. Then, embryos were washed in W3 (2X SSC pH4.5, 0.1% Tween-20) for 15 minutes and equilibrated to room temperature for 15 minutes. After embryos reached to room temperature, they were washed three times in 1% Tween-20 containing Tris buffered saline (TBS, 137mM NaCl, 20mM Tris) and blocked in blocking solution (TBS-T with 20μL fetal calf serum and 20μL of 100mg/mL BSA) for at least 2 hours. After the blocking, embryos were incubated with DIG antibody (1:3000 dilution in

TBS-T) overnight at 4ºC. On the third day of ISH protocol, embryos were washed in TBS-T initially three times for 5 minutes then five times for 90 minutes. On the last day of the protocol, embryos were first washed to get rid of the unbound antibody in NTMT (100mM Tris pH9.5, 100mM NaCl, 1% Tween-20) three times for 10 minutes and the color was developed using staining solution (NTMT with 3.4μL/mL of 100mg/mL NBT and 3.5μL of 100mg/mL BCIP). Staining reaction was stopped by exchanging the staining solution with PBS after the required pattern/staining is obtained (depends on the gene which was being probed).

4.2.5 Animal transgenesis

4.2.5.1 Lentivirus injection

Female mice were super-ovulated by injecting 5 IU PMSG then 47 hours after by injecting 5 IU HCG and were mated on the same day. Embryos were collected from plugged females on the day of the plug and incubated in hyaluronidase (Sigma-Aldrich Cat.No.H3884) solution (0.3mg/ml) to have the zygotes released easier as the digestion removes the sticky cumulus cells. Then the hyaluronidase solution was washed and zygotes were incubated in KSOM medium (Millipore Cat.No.MR106D) in the humidified incubator. Freshly pulled Harvard 1.0mm OD x 0.78 capillaries were filled with ~2μL of virus (see section 4.2.2.8.2 for production) and few picoliters of virus was injected into perivitelline space of the zygotes. After injection zygotes were incubated in KSOM medium at 37ºC incubator (5% CO_2) and kept for 3 days in culture until they reached the blastocyste stage.

4.2.5.2 Embryo transfer

The embryos were transferred into timed pseudopregnant CD-1 fosters (2.5 days). Foster mice were anesthetized using 10μL per gram of mouse of the 1mL Ketamine, 0.8mL Xylaxidin in 9.2 mL PBS mix.

5. Results and Discussion

5.1 Impact of MusD sequence on gene expression

The phenotype observed in *Dactylaplasia* mice, the genetic pattern of inheritance and the genomic context of the mutations suggested that it could be due to an alteration of gene regulation, involving possibly long-range elements from the *Fgf8* locus. In this study, I aimed to investigate if and how MusD elements could modify the communication of regulatory sequences and the normally associated promoters.

5.1.1 Testing the previous models explaining the *Dactylaplasia* and *SHFM3* conditions

1. A role of *Fbxw4* and *Fgf8*

Previous analysis of Dac^{2j} mice had suggested that a down-regulation of *Fbxw4* (observed by Sidow et al.) or of *Fgf8* (because of its function in the AER) could be involved in the altered development of the limbs. In order to investigate the role of *Fbxw4* gene in this limb malformations, we (together with Sandra Ruf) engineered mice with a chromosomal deletion between *Lbx1* and *Fgf8* in the locus (*DEL(Lbx1-Fgf8)* Figure 1) using *lox*P sites targeted to these genes and *in-vivo* TAMERE strategy (Hérault et al. 1998). The heterozygous animals for the deletion did not show any limb malformation. By mating *DEL(Lbx1-Fgf8)* animals with Dac^{2j} animals, I obtained *DEL(Lbx1-Fgf8)*/Dac^{2j} animals. The limbs of these animals were phenotypically similar to Dac^{2j} heterozygous, and did not show the more severe monodactyly observed in surviving Dac^{2j} homozygous animals. From these observations, we concluded that neither the loss of *Fgf8* nor of *Fbxw4* in the mutant limbs was importantly contributing to the *Dactylaplasia* phenotype. The role of *Fbxw4* was further questioned as Dac^{2j} animals bred to C57BL/6J background has normal limbs despite still having a down-regulation of *Fbxw4* caused by the insertion of MusD element into one of its introns (Kano, Kurahashi et al. 2007). Altogether, these data argued that neither the down-regulation of *Fbxw4* nor *Fgf8* was sufficient to cause the limb phenotype or even modulating its severity, and suggested that other genes were implicated most likely through a gain-of-function mutation.

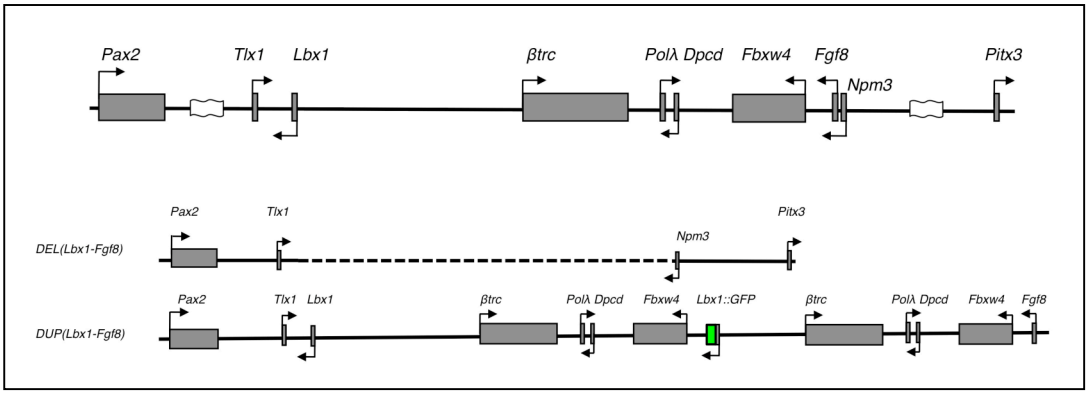


Figure 1. The view of chromosomal re-arrangements introduced into locus
For simplicity genes between *Pax2-Tlx1* and *Npm3-Pitx3* are not shown.

2. The toxicity of MusD expression

Intriguingly, Kano et al. observed an ectopic expression of MusD in the AER of Dac^{1j} animals prior to or concomitantly of loss of *Fgf8* expression in the AER (Kano, Kurahashi et al. 2007). They implied that this ectopic expression could have an effect on the survival of the cells in the AER. In theory, this toxicity could be due to an over expression of MusD retroviral proteins or to the mutagenic consequences of a highly increased transposition of ETn/MusD, via these proteins. However, ETn and MusD transposons are broadly expressed in different tissues from E7.5 to E13.5 (Loebel et al. 2004) in normal mice without apparent problems. To further examine this model, I carried out in-situ hybridization studies on whole E10.0 Dac^{1j} and Dac^{2j} embryos with the RNA probes specific for *MusD/ETn* transcripts. These experiments confirmed that MusD transcripts were expressed in the AER of *Dactylaplasia* animals. In addition, I detected an ectopic expression of MusD in the forebrain domain of Dac^{1j} and Dac^{2j} animals, as well as in the mid-hind brain boundary domain of Dac^{2j} animals (see Figures 2 and 3, MusD panel). Like for the AER, these domains coincided strikingly with the normal expression domains of *Fgf8* in these structures. However, brains of *Dactylaplasia* heterozygous animals seemed to develop normally in contrast to what is observed with a hypomorphic allele of *Fgf8* (Meyers et al. 1998), suggesting the absence of massive cell death in this crucial signaling center.

Thus, while we have confirmed that MusDs were ectopically expressed in *Dactylaplasia* animals, the possible toxic effects evoked by Kano et al. would have to be tissue-specific. They are not apparent in other domains of MusD ectopic and therefore AER is particularly sensitive to them. Furthermore, as MusD elements are

mouse-specific, a model putting forward the ectopic expression of MusD as primary cause of *Dactylaplasia* cannot explain the striking similar conditions resulting from duplications in human *SHFM3* patients.

3. *SHFM3*: a mini-trisomy syndrome leading to gene up-regulation?

When initially discovered, the duplications found in *SHFM3* patients led to the proposal that is genomic defect could be due to an altered gene-dosage, such as a mini-trisomy condition. To test this hypothesis we have engineered duplications along the locus. Mice with a duplication of the critical interval found in *SHFM3* patients (between *Lbx1* and *Fgf8* genes (*DUP(Lbx1-Fgf8)*) and with a much larger interval encompassing the largest duplication found in human (*DUP(Pax2-Pitx3)*) did not show any limb malformations, either in heterozygous or homozygous animals for the duplications (not shown). Therefore, the triplication or quadriplication of gene copies in the locus in mice did not lead to the phenotype observed in human. In order to test if other genes in the locus were expressed in *Fgf8* domains, like MusD, we used in-situ hybridization on E10.0 *Dactylaplasia* embryos to look at the distribution of mRNAs from the genes present in the critical interval in the *Dactylaplasia* models. We were unable to detect any up-regulation or ectopic expression of the genes tested (*Lbx1*, *BTrc*, *PolL*, *Fbxw4* and *Dpcd*).

5.1.2 *Dactylaplasia:* mutations causing complex regulatory alterations

1. Effects on *Fgf8* expression

This study however revealed several additional changes for *Fgf8* expression. In particular, *Fgf8* was importantly down-regulated in the forebrain of Dac^{2j} homozygous animals, concomitant to the ectopic expression of MusD in this domain. However, in Dac^{1j} animals, the expression of *Fgf8* was still robust in the forebrain (see Figures 2 and 3, *Fgf8* panel). These additional effects could explain why Dac^{2j} homozygous were never recovered from Dac^{2j} heterozygous crosses. In fact, there might be other effects at later stages that were overlooked and contribute to additional phenotypes in homozygous mutants, due to *Fgf8* loss of function in late domains. Furthermore, such loss of expression could account for the additional defects found in patients with a syndromic form of *SHFM*3 (Dimitrov et al. 2010). Interestingly, these defects (hearing disorders, micrognathia/microcephaly) are observed in regions which developed under the influence of the *Fgf8* gene as

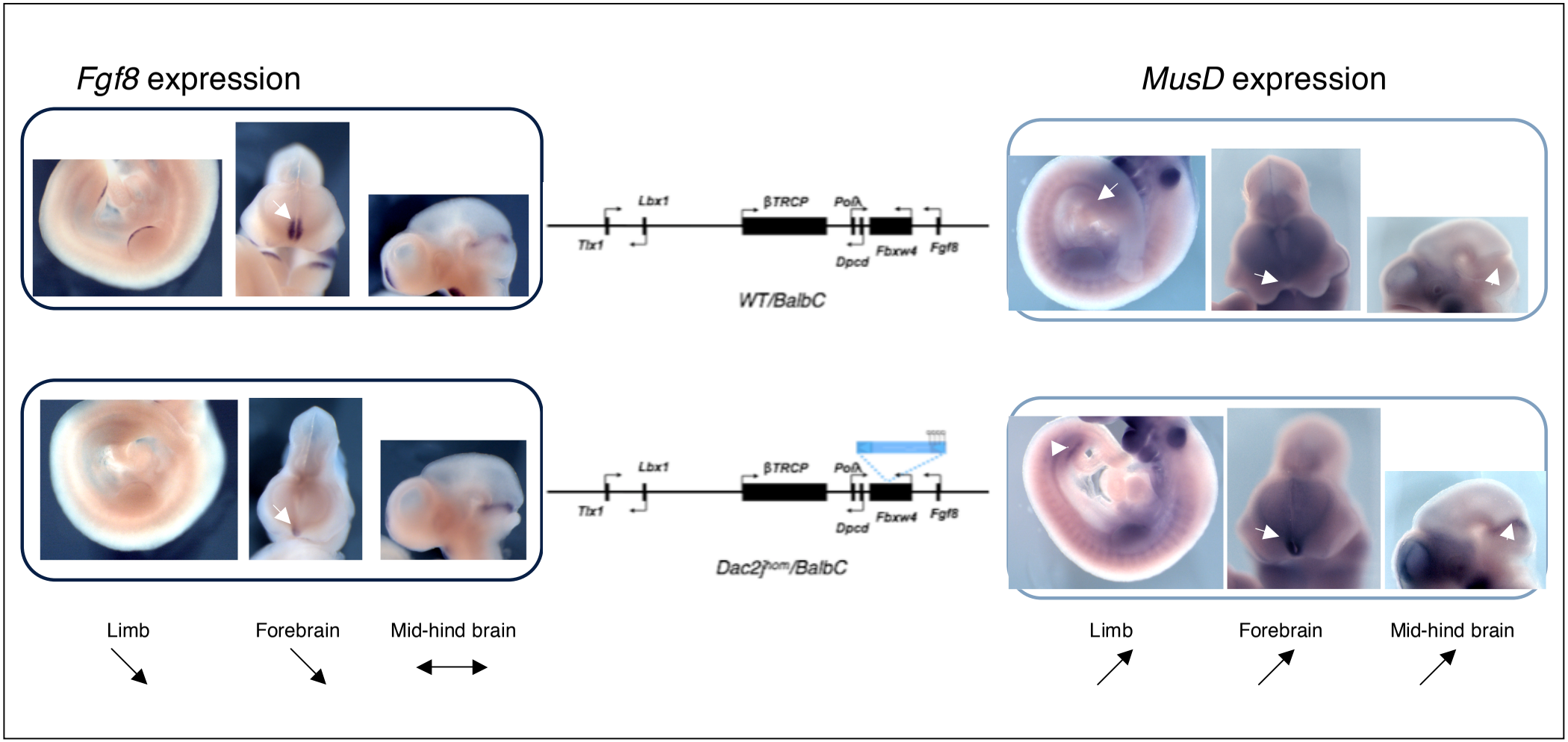


Figure 2. Whole Mount in-situ Hybridization on *Dac*^{2j} */BalbC* E10.0 embryos

The circles represent the unmethylated CG positions in 5'LTR

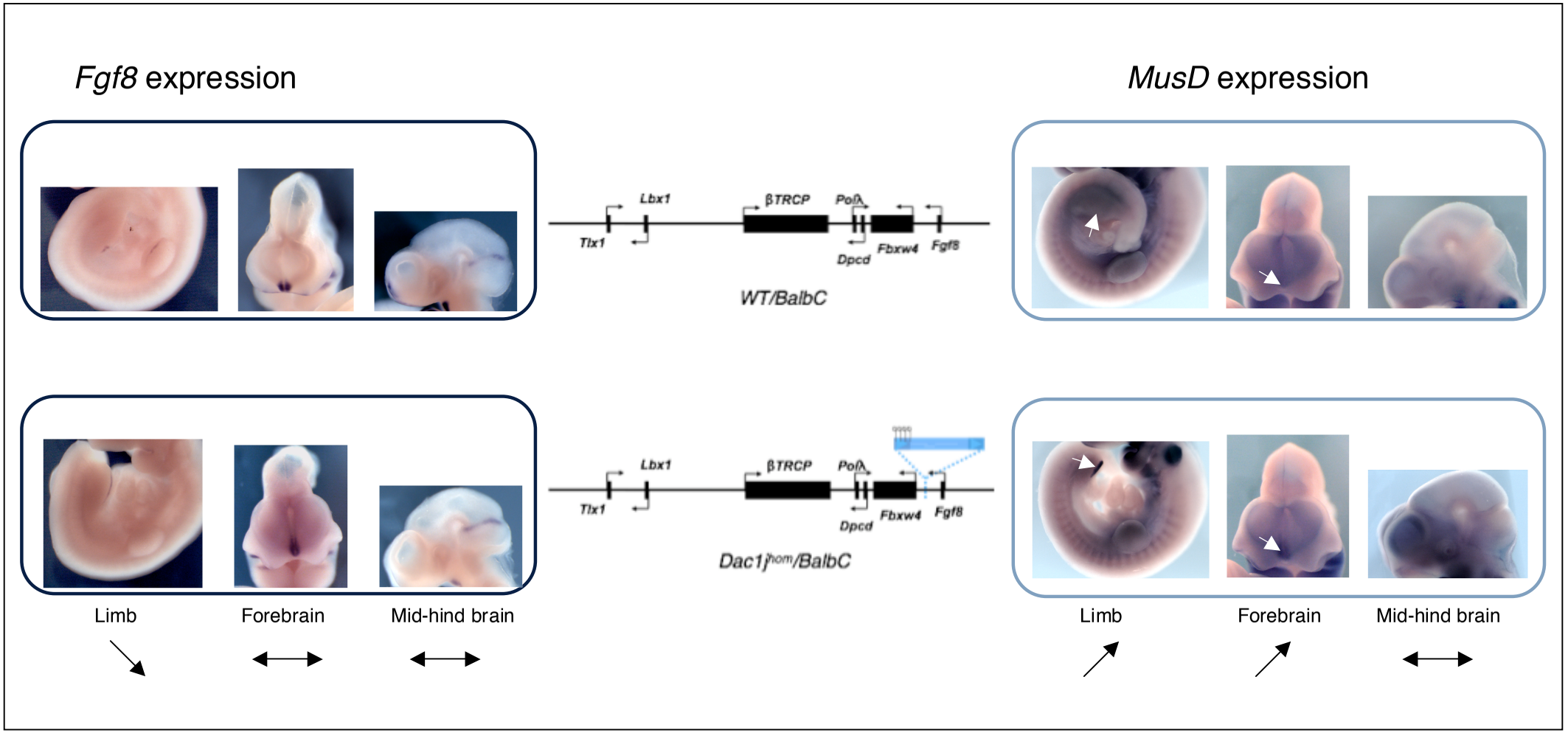


Figure 3. Whole Mount in-situ Hybridization on *Dac*^{1j} */BalbC* E10.0 embryos

The circles represent the unmethylated CG positions in 5'LTR

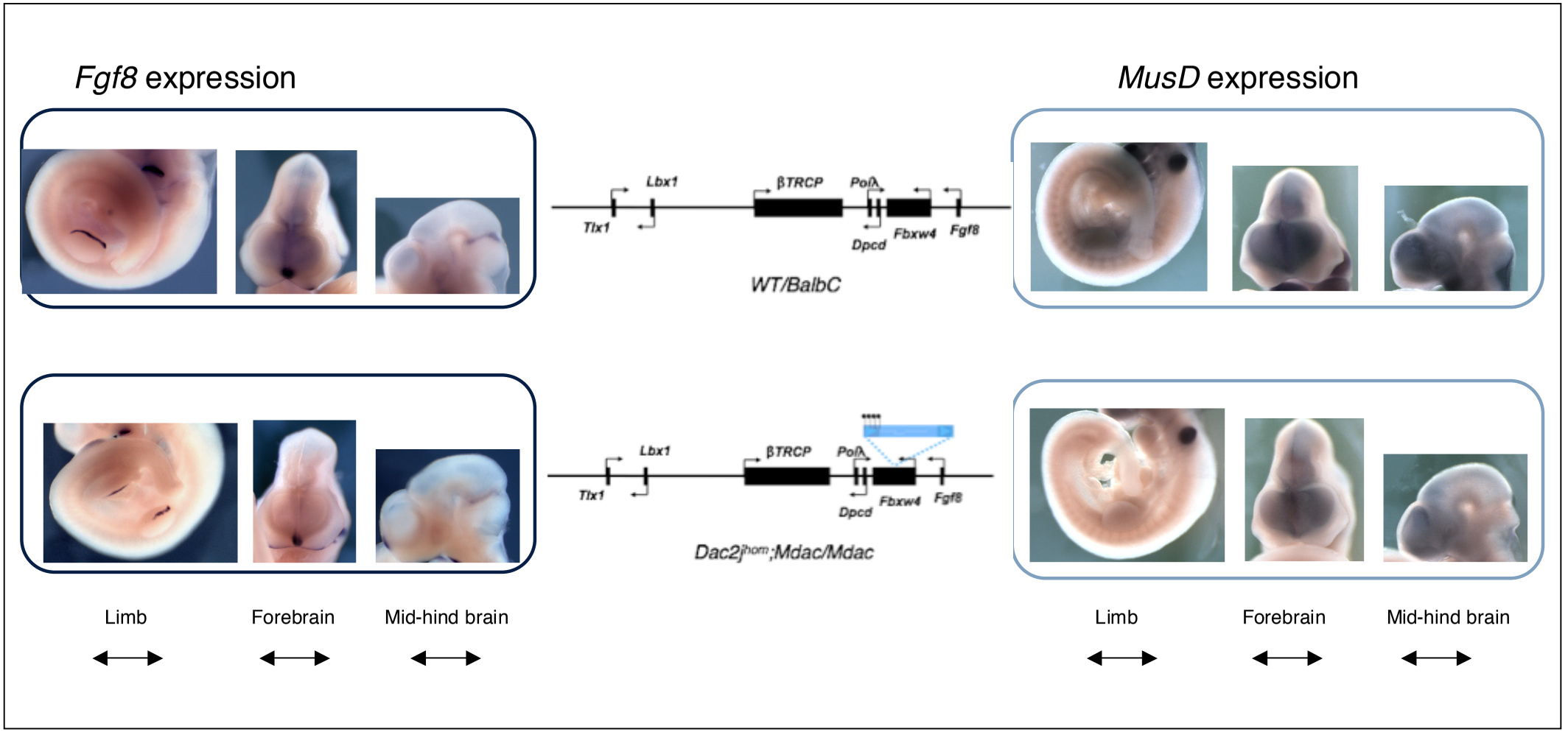


Figure 4. Whole Mount in-situ Hybridization on *Dac*^{2j} */Mdac* E10.0 embryos

The circles represent the methylated CG positions in 5'LTR

well. Importantly, this down-regulation in the forebrain was not seen in *Dac*^{2j} homozygous animals over the C57BL/6j background (see Figure 4). Hence, the effects of the insertions of MusD on the *Fgf8* expression domains correlated with the transcriptional status of *Dac*^{2j}-MusD and the modifier locus, *Mdac*, is counter-acting the consequences of the insertions both at the morphological and molecular level, in the limbs and in the forebrain.

2. Comparing the regulatory effects of the insertions and the position of enhancers

In a parallel study, a PhD student in the group, Mirna Marinić, has identified and localized the position of multiple regulatory elements in this locus that are controlling *Fgf8* (summarized in Figure 5).

She found that a BAC covering about 200kb of *Fgf8* 3'flanking sequence contained the elements that could drive the expression of the reporter gene in an *Fgf8*-like manner (Figure 5a,b and c). Within this region, she identified several individual enhancer modules (size in the range of 800-2000bp) that each drove reproducibly the expression of a reporter gene under the control of a minimal promoter in a different subset of *Fgf8* expression domains. In fact, some expression domains were represented with more than one element, indicating a redundancy or synergy of distinct elements for *Fgf8* control. In particular, she found a total of five AER enhancers, four of which were located in introns of *Fbxw4*; three enhancers driving expression at or around the mid-hind brain boundary and as well as a forebrain enhancer. These elements are indeed controlling *Fgf8* expression, as she showed a deletion from *Pol λ* gene to the middle of *Fbxw4,* including most of them was allelic to *Fgf8* (Mirna Marinić, unpublished data).

Amongst these different elements, one of them was possibly directly disrupted in a *Dac* mutant. The *Dac*^{2j} insertion is in the middle of a conserved non-coding element (Figure 6) even though it is only next to the most conserved portion. This element was shown to drive the reporter gene in forebrain and mid-to-hind brain (element is depicted with the yellow line in Figure 5a and the transgenic for the LacZ reporter is shown in Figure 5e), as well as in the neural tube and the developing kidneys. This insertion of a MusD next to the core-conserved part of this module could have impaired its regulation function and have contributed to the down-regulation of *Fgf8* in *Dac*^{2j} animals. For some domains (e.g. the kidney) where the expression of *Fgf8* is under the control of multiple redundant modules, these effects are difficult to see and we did not find kidney aplasia in the *Dac*^{2j}/*Fgf8*^{null} new borns,

in contrast to the *DEL(Pol λ-Fbxw4)/ Fgf8*^{null} animals (Mirna Marinić, François Spitz not shown). Nevertheless, the transgenic analysis suggested that this element play a

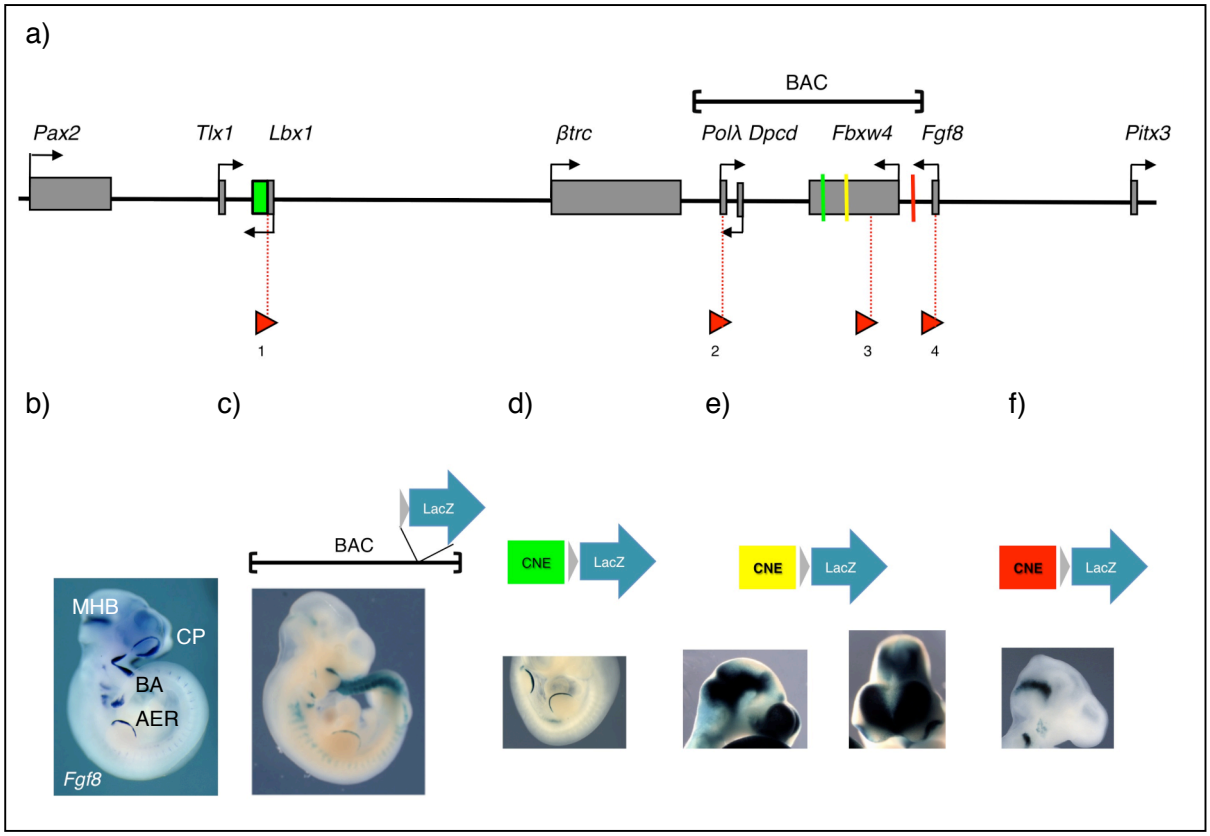


Figure 5. The *cis*-regulatory architecture of *Fgf8* locus

a) The representations of targeted *lox*P sites in the locus (*lox*P into *Lbx1* gene replaces *Lbx1* with GFP). Four different lines that were used to engineer duplications or deletions in the locus are shown here. b) in-situ hybridization for *Fgf8* c,d,e,f) The regulatory elements characterization by transgenic assays, using LacZ as a read-out. E10.5 embryos are shown here. g) The observed phenotypes of engineered re-arrangements in the locus. MHB: Midbrain hindbrain boundary, BA: Branchial arches, CP: Commissural plate, AER: Apical ectodermal ridge.

major role in the expression of *Fgf8* in the forebrain. Supporting an effect of the insertion of its enhancer activity, as mentioned above, we found that *Fgf8* was down-regulated in the forebrain of *Dac*^{2j}/*Dac*^{2j} embryos. However, we detected also MusD expression in this domain in the place of *Fgf8,* and when the *Dac*^{2j} mutation was in the C57BL/6J the expression of *Fgf8* in the forebrain was unaffected (Figure 4). Thus, the *Dac*^{2j} insertion does not seem to disrupt the regulatory potential associated with the conserved region next to where the element is inserted. Instead, as illustrated by the in-situ hybridization, both *Dac*^{2j} and *Dac*^{1j} animals showed reciprocal changes in MusD and *Fgf8* expression (see Figure 2 and 3, MusD panel),

corresponding to a competition mechanism. Interestingly, the comparison of the effects of *Dac*^{1j} and *Dac*^{2j} insertions suggested that MusD might be interfering differentially with the *Fgf8* expression depending on its insertion site. For several domains of expression the positive effects on MusD expression and negative effects on *Fgf8* expression correlated well with the proximity of MusD to the corresponding regulatory elements; and particularly when MusD is inserted in-between a regulatory module and *Fgf8*. For instance, expression domains depending on elements localized proximal to *Fgf8* promoter (e.g. tailbud, somites) seemed to be unaffected in *Dactylaplasia* mice. The contribution of a midbrain enhancer that is more proximal to *Fgf8* (Mirna Marinić, unpublished data and (Inoue et al. 2008)) (shown by the red line in Figure 5a) could also explain why *Fgf8* expression in that domain is mostly unaffected, while other mid-hind brain boundary elements localized in *Fbxw4* introns could activate MusD (Figure 2).

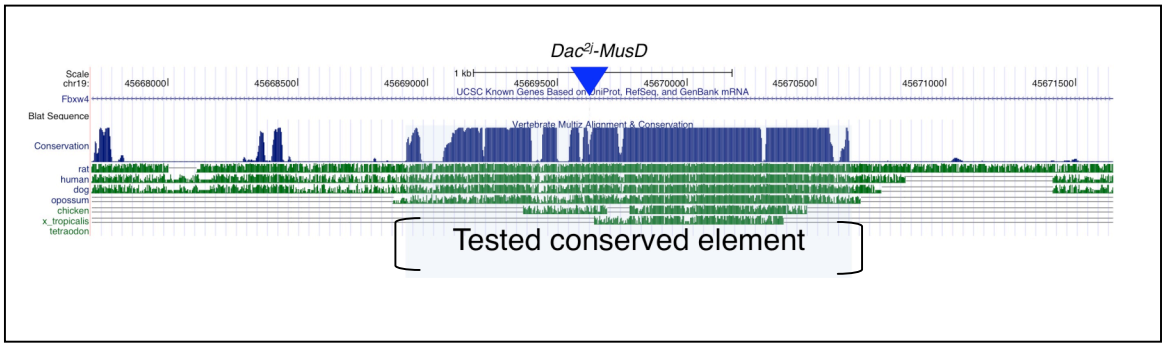


Figure 6. The position of Dac2j insertion relative to the closest conserved sequence. MusD does not split the region with the highest conservation

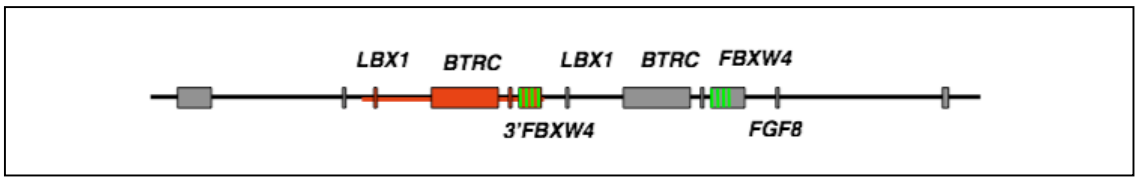


Figure 7. The duplicated region in *SHFM3* and relative positions of identified AER enhancers

In brief, this set of experimental evidence suggested that the MusDs inserted in the *Dactylaplasia* mice were hijacking several regulatory elements of *Fgf8,* leading to down-regulation of *Fgf8* at different degrees, in a tissue-specific manner (Figure 8). This hijacking is more robust when MusDs are located between the regulatory elements and the promoter, suggesting that the relative position of the different elements contribute to the outcome of this competition between *MusD* and *Fgf8*. From an *Fgf8* perspective, these MusDs are acting as enhancer blockers.

Interestingly, the region that is duplicated in *SHFM3* contains several of the modules shown to regulate *Fgf8*, notably four elements associated with the AER expression. As a consequence of the duplication, a set of these enhancers are moved away from *Fgf8* (Figure 7). This raises the possibility that this change in their position relative to their normal target genes could lead to a change in their gene specific activity, hence they could rather act on another gene and drive it in a subset of *Fgf8* specific domains (AER). Thus, we propose that a common cause underlying the limb phenotype in *Dactylaplasia* mouse and *SHFM3* patients could be the ectopic expression of an endogenous gene from the locus, under the control of regulatory elements normally associated with *Fgf8* expression in the AER.

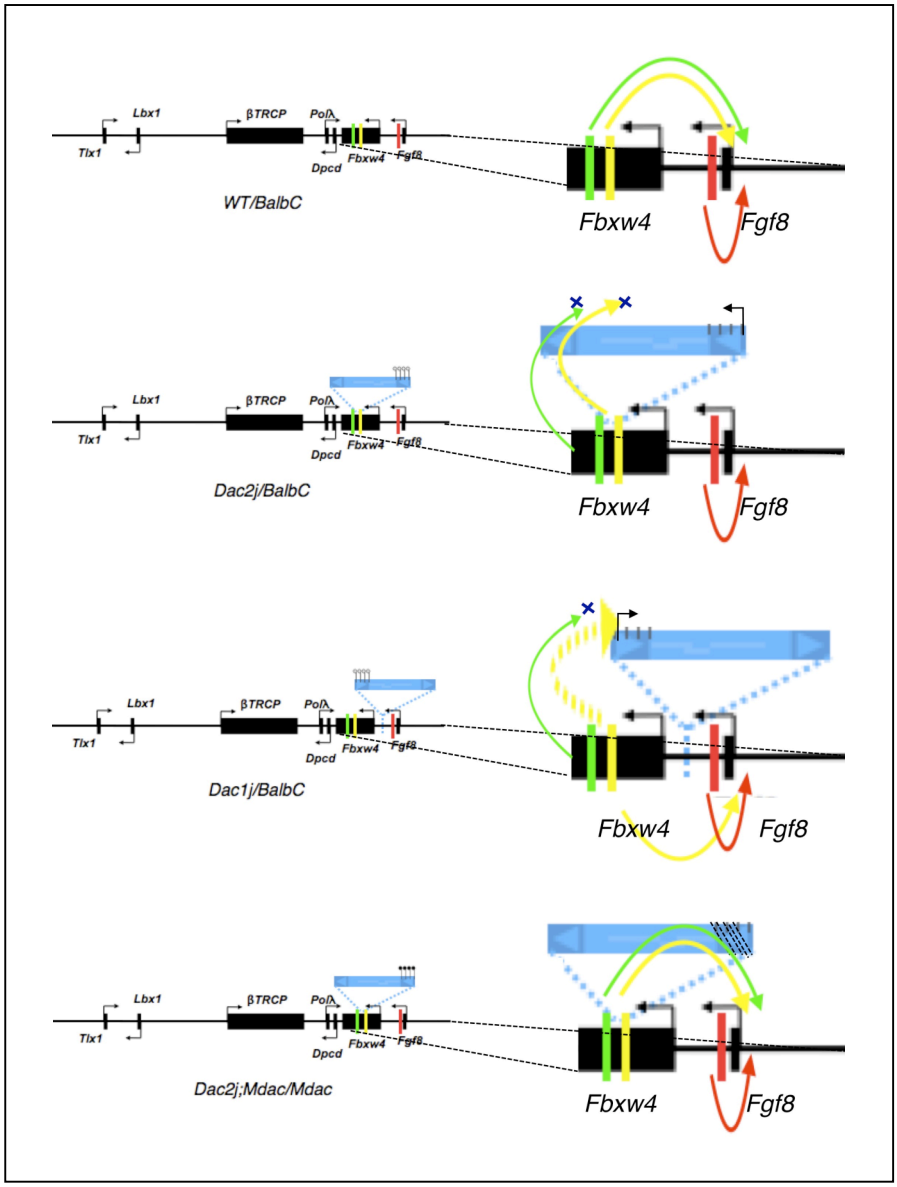


Figure 8. The model for enhancer blocking activity of MusD

(Green:AER, Yellow:Forebrain, Red:Mid-hind brain boundary)

The ectopic expression would be caused by preventing the normal interactions between *Fgf8* and its enhancers, either by structurally moving them away from the gene (as in *SHFM3*) or by functionally preventing these interactions through enhancer blocking/competition activity (by MusD as in *Dactylaplasia*). We have not been able to directly detect any ectopic expression of additional genes, besides MusD in the AER of *Dactylaplasia* mice*,* which would have supported our model. However, as the limb malformation is caused by massive cell death in the AER, it could be difficult to detect such up-regulation, as the cells where it occurs could be rapidly lost.

5.1.3 Characterization of MusD as an insulator using *ex-vivo* assays

To test the role of MusD as an insulator/silencer, I used an *ex-vivo* system similar to what has been used previously to characterize other insulator elements (Chung et al. 1993). I chose the luciferase as a reporter gene in my experimental set-up. Given the size of MusD element, I subdivided the full length MusD into large overlapping 9 sequences (Figure 9a). All test constructs were obtained from a methylation deficient bacterial strain to avoid epigenetic changes that could impair the activity of the tested elements. The different constructs were linearized prior to transfection. These constructs were transfected into two different cell lines, **H**uman **E**mbryonic **K**idney (HEK) 293T and NIH3T3, and the transfection efficiency was normalized using a renilla luciferase expression vector as an internal reference. These experiments were carried out together with Daria Shlyueva, a summer student under my supervision.

We first investigated the potential of MusD as a silencer sequence by comparing the activity of constructs where MusD fragments were cloned upstream of the enhancer-promoter region (Figure 9b). We did not observe any reduction in the reporter gene activity for any of the fragments tested (Figure 10). Therefore, MusD fragments do not seem to act as silencers. Next, in order to test the insulator potential of MusD sequences, we inserted the MusD fragments between the SV40 enhancer element and the promoter and checked firefly luciferase reporter (Figure 11a). This classical method of insulator testing method is called **E**nhancer **B**locking **A**ssay (EBA, defined in the pioneer studies on the 87A7 locus in *Drosophila* for hsp70 and flanking scs (specialized chromatin structures) (Kellum and Schedl 1992)).

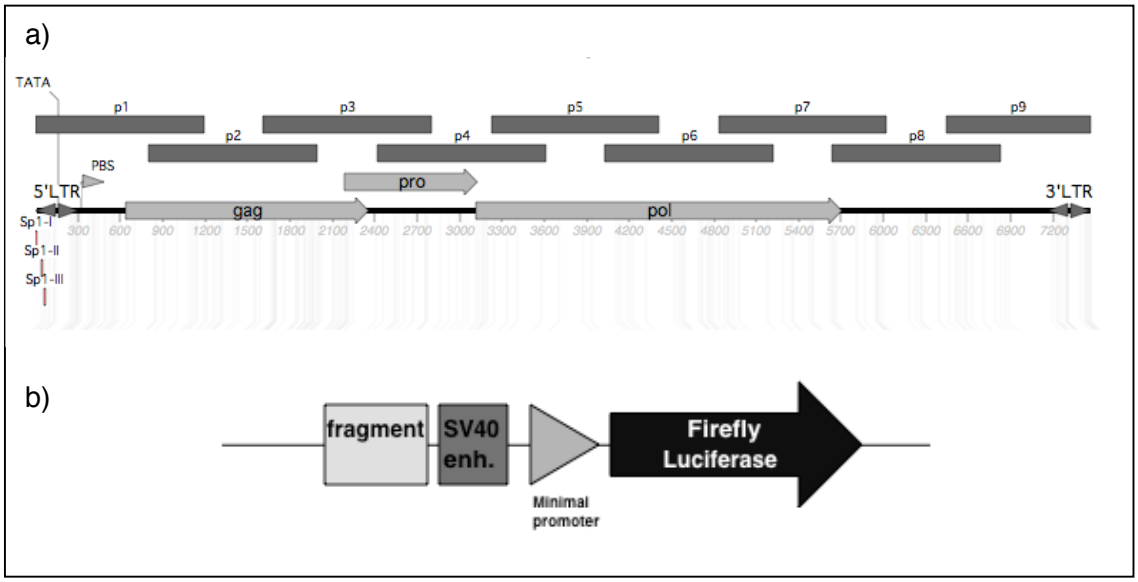


Figure 9. Experimental set-up for *ex-vivo* assays, a) The overview of the test fragments. Each line below the annotated sequence represents a CG dimer. Full length MusD is partitioned into nine ~1.2kb fragments with ~0.4kb overlaps. These pieces of MusD are shown here as pn (n is [1,9]). First piece (p1) covers 5'LTR and last piece (p9) covers 3'LTR of MusD. b) SV40 (Simian virus 40) early enhancer drives the firefly luciferase gene through an HSV-TK (Herpes simplex virus thymidine kinase) minimal promoter

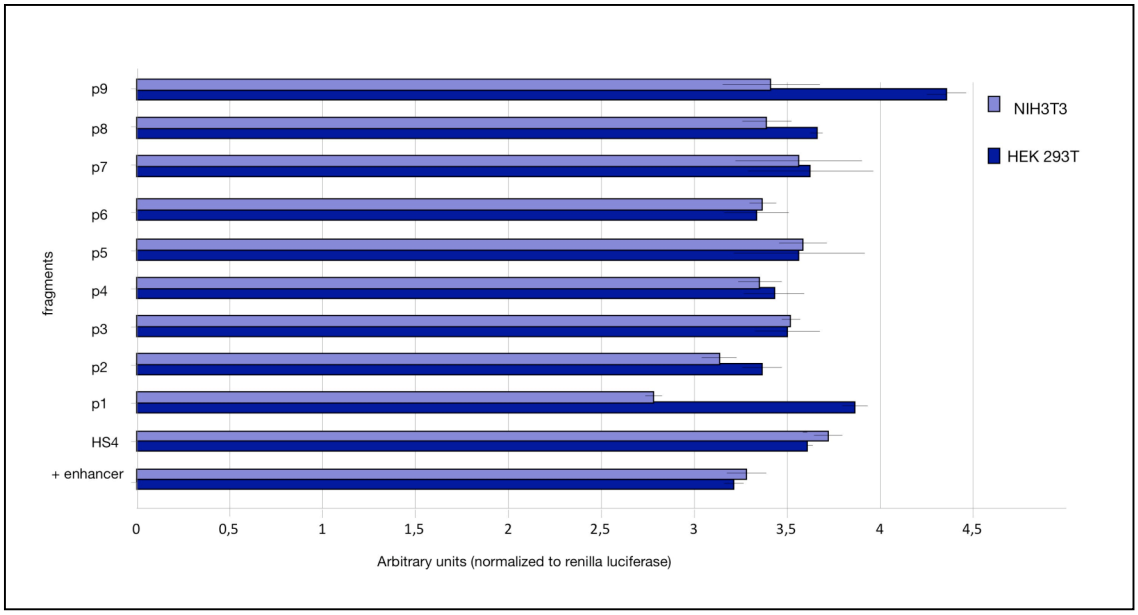


Figure 10. The silencer test of the MusD fragments

The arbitrary units representation of luciferase reporter (firefly luciferase normalized to renilla luciferase) in HEK293T and NIH3T3 cell lines. Error bars represent the standard deviation between experimental triplicates.

We included β-globin locus HS4 insulator element (Chung, Whiteley et al. 1993) as a positive control. To control for effects that would be due to the increased distance between the enhancer and the promoter but not with *bona fide* insulator activity, we cloned a piece of the *E.coli* Kanamycin resistance gene of the same length (~1.2kb). Any fragment that was reducing the reporter gene expression significantly more than the neutral Kanamycin sequence and to a level comparable to HS4 insulator was considered as a potential insulator sequence. According to these criteria, I found in both cell lines that MusD fragments 2, 3, 4 and 6 reproducibly and significantly acted as insulator sequences (Figure 11b).

To further confirm this activity and to complement the classical enhancer blocking assay, I repeated this assay with a different experimental design. In this set-up the enhancer is localized between two distinct copies of the same promoter but cloned in opposite orientation. Each promoter drives expression of a different luciferase reporter, either renilla or firefly. The MusD fragments to be tested are placed between the enhancer and the promoter sequence driving firefly luciferase (Figure 12a). Changes in the ratio between renilla and firefly luciferase activities could indicate that the shared enhancer activity was re-distributed differently between the two promoters. Thus, this design allows identification of elements that could influence promoter competition. As a proof of principle, we first tested the HS4 sequence and found that it showed strong insulator activity when compared to the Kanamycin resistance gene fragment (~20 folds reduction by HS4 and ~4 folds reduction by Kanamycin). With this assay, we found again that the MusD fragments 2,3 and 4 acted as insulators in both HEK293T and NIH3T3 cells (Figure 12b). The overall values of luciferase activities were not significantly different between the constructs (Figure 12c), suggesting that there were not major silencing effects associated with the fragments (confirming the previous analysis). However, the values obtained with the minimal promoter used (HSV-TK) in NIH3T3 cells were very low, and close to background level, this suggests that this promoter is not eliciting a strong expression in these cells and explains potentially the greater variability obtained in NIH3T3 cells.

Altogether, these experiments showed that several MusD regions have robust insulator/enhancer blocking activity mostly sequences 2, 3, and 4. These fragments are only partially overlapping; suggesting that more than one region contains insulator activity. In contrast, we did not identify any silencer elements within MusD sequence.

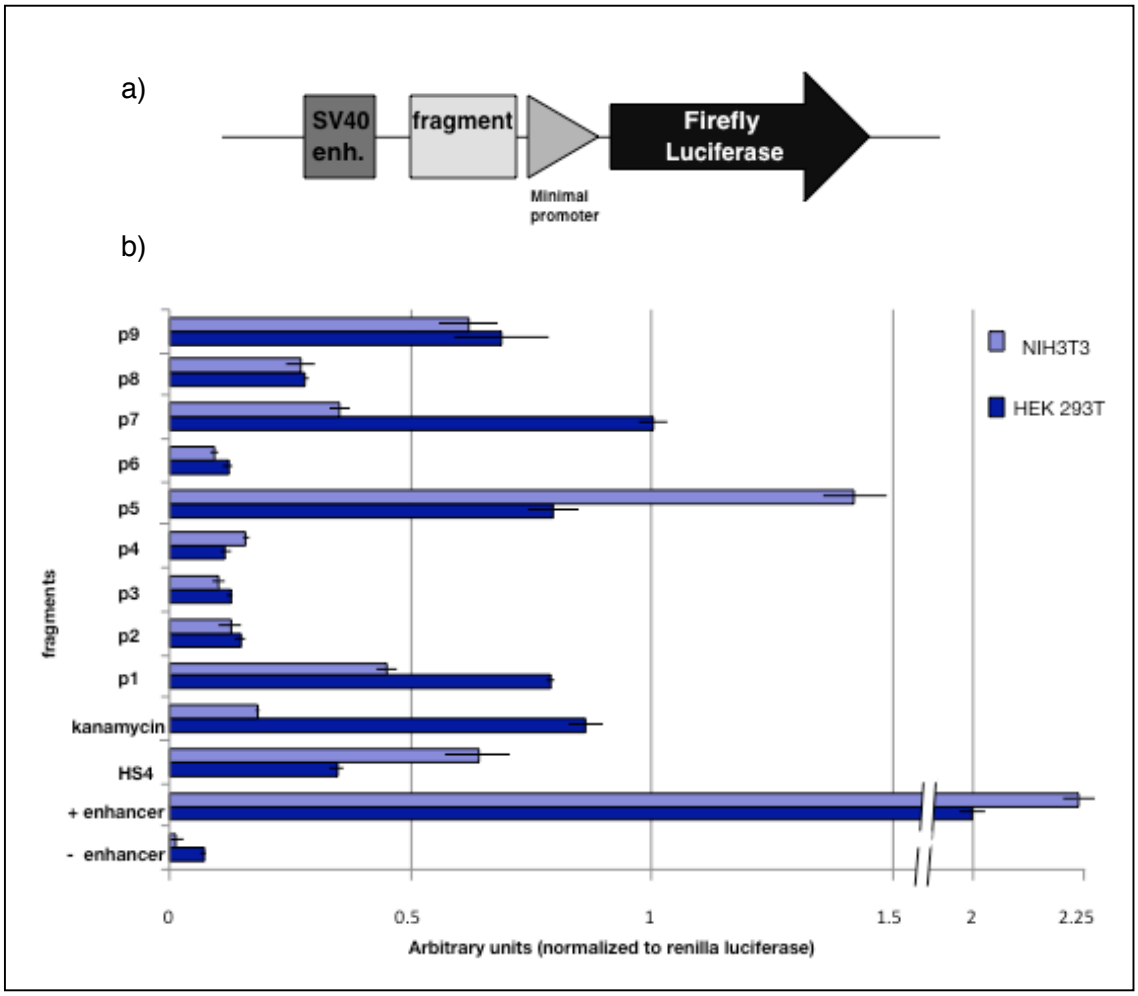


Figure 11. The enhancer blocking test in two cell lines

a) Design of the construct for EBA. b) The arbitrary units representation of luciferase reporter (firefly luciferase normalized to renilla luciferase) in HEK293T and NIH3T3 cell lines. (-)enhancer represents the transfection construct without SV40 enhancer sequence (i.e. reporter gene expression driven only by the minimal promoter). Error bars represent the standard deviation between experimental triplicates.

To further narrow down the enhancer blocking sequence(s), embedded in this 2kb long region formed by these 3 pieces altogether (excluding the parts overlapping with fragments 1 or 5 without insulator activity) I split it into eight fragments with 100bp overlaps (Figure 13a). Their enhancer blocking potential was tested with the previously used strategies. As expected from the first analysis, I found multiple fragments displaying an enhancer blocking activity. From both strategies fragments 1_2, 6_2 and 8_2 showed insulator activities of similar strength to chicken HS4 prototypic insulator (Figure 13b,c).

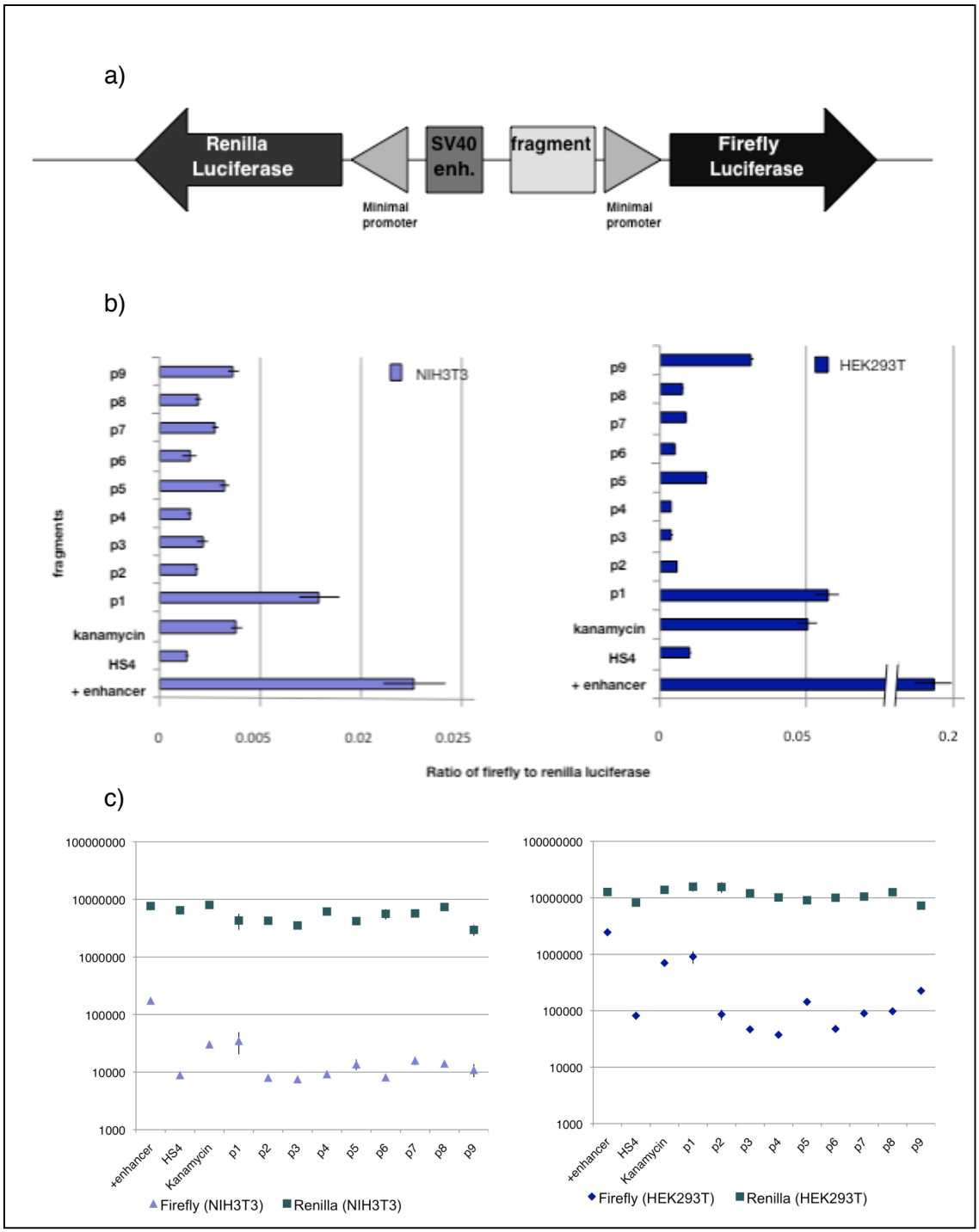


Figure 12. The enhancer blocking test by two reporters

a) Design of the construct for transient transfection assays. b) The ratio of the firefly luciferase activity to renilla luciferase activiy in NIH3T3 and HEK293T cell lines. Error bars represent the standard deviation between experimental triplicates. c) The absolute reads of luciferase and renilla activities (shown in log scale)

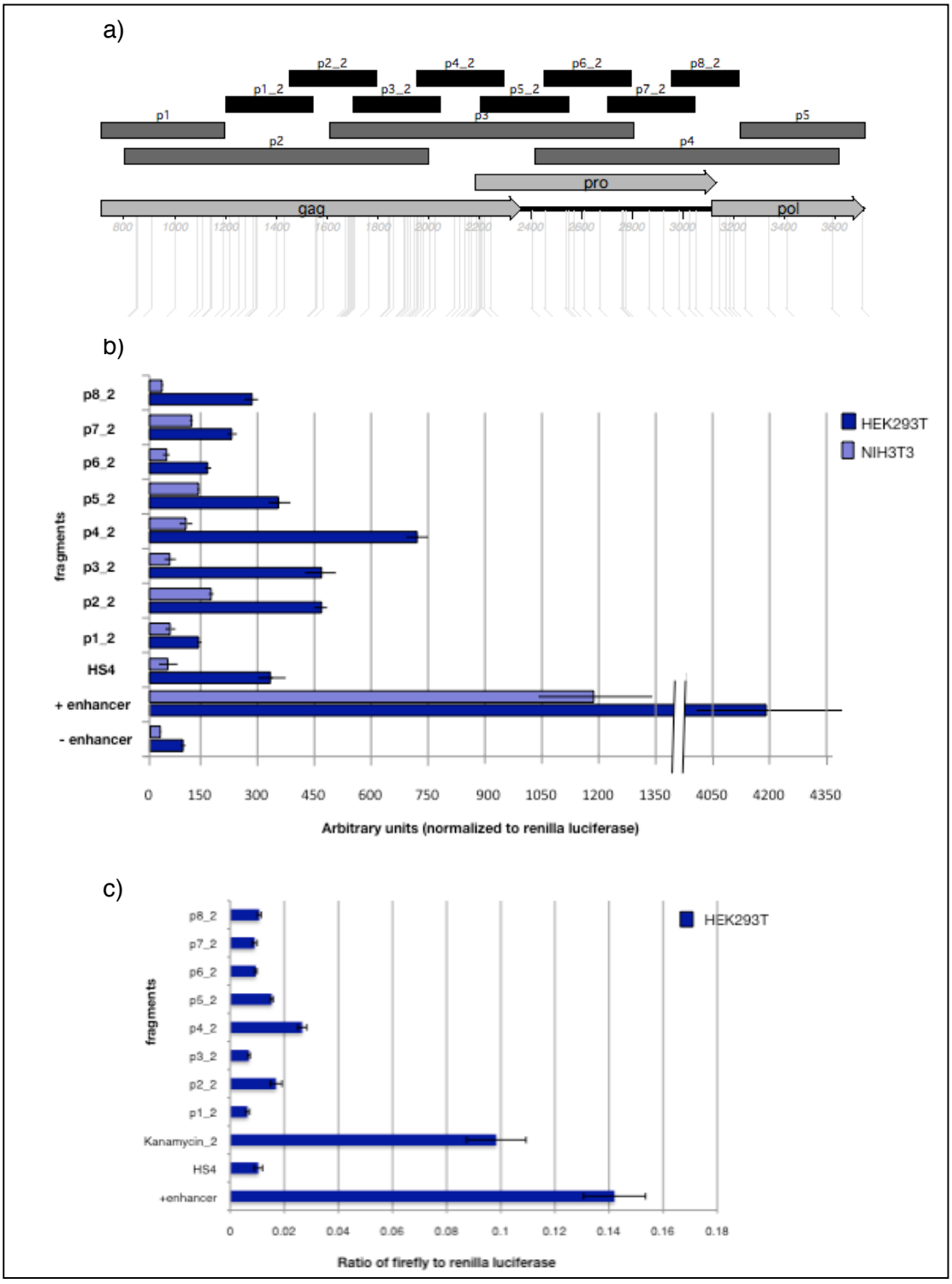


Figure 13. The enhancer-blocking test for further divided fragments

a) The division of new fragments (with a close-up to MusD element represented in Figure 9a). b) The arbitrary units representation of luciferase reporter (firefly luciferase normalized to renilla luciferase) in HEK293T and NIH3T3 cell lines. c) The ratio of the firefly luciferase activity to renilla luciferase activiy in HEK293T cells. Error bars represent the standard deviation between experimental triplicates.

In vertebrates, several proteins have been shown to bind to insulator elements, such as NF-Y/YY1 binding for regulation of Hoxb4 (Gilthorpe et al. 2002) or USF proteins that are required for HS4 barrier activity (West et al. 2004). However, the CCCTC-binding factor (CTCF) is the only protein identified so far that both binds directly to insulator sequences and is required for their enhancer-blocking activity (Gaszner and Felsenfeld 2006). CTCF is an evolutionarily conserved DNA-binding protein, which binds to different DNA sequences through its 11-Zinc Fingers (Ohlsson et al. 2001). ChIP-chip or ChIP-seq arrays have identified consensus binding for CTCF (Kim, Abdullaev et al. 2007), (Chen et al. 2008). To see if CTCF could be involved in MusD fragment insulator activities, I looked for the presence the consensus sequence within fragments (using insulator database, http://insulatordb.uthsc.edu). Interestingly, I did not find any significant CTCF binding site within the three insulating fragments 1_2, 6_2, and 8_2. I analyzed sequence composition similarities of these three fragments (analysis by pairing the fragments one to another and searching for at least 90% of similarity for at least 15 bases). This analysis showed a short sequence with high similarity between p1_2 and p6_2 with a core AAAGGACAGAATA sequence (Figure 14). I performed a preliminary experiment to test the insulator function of this similarity region and few nucleotides of the flanking sequence from the p1_2 and observed a mild blocking activity, suggesting that it could indeed contribute to the insulating activity of the whole fragment. However, additional experiments would be required to further confirm this hypothesis and identify the possible protein associated with this region(s).

In summary, we found that MusD elements could behave as functional insulators, confirming that, as suggested by the effects on *Fgf8* expression *in-vivo*, this activity could account for the observed changes in the *Dactylaplasia* mice. We characterized multiple regions in the MusD sequence that could contribute to this activity. Interestingly, these elements do not contain CTCF binding sites, in contrast to most vertebrate insulators characterized so far, suggesting that it could correspond to new mechanisms of enhancer blocking.

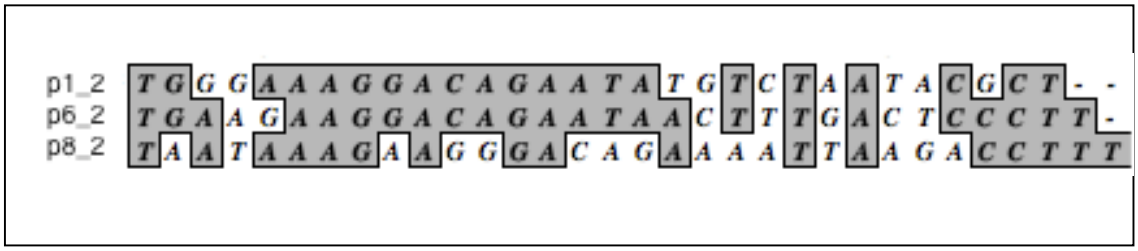


Figure 14. The sequence comparison of sequences with blocking activity

5.1.4 The characterization of the factors that contribute to *Dactylaplasia* and *SHFM3* phenotypes

The model proposed in section 5.1.2, suggests that the enhancers that control normally *Fgf8* expression could be re-directed to other genes, upon structural changes of the locus. We, and others, have shown that this is leading to MusD expression in an *Fgf8*-like manner in *Dactylaplasia* mice. We hypothesized that this could be followed by the activation of another gene in the AER, through the action of distal AER enhancers that are not able to act on *Fgf8* anymore, MusD acting as an insulator. This model provides also a simple explanation for the human *SHFM3* relying on the same mechanism. The duplication found in *SHFM3* patients includes several AER enhancers; the duplicated set of enhancers are moved further away from their normal target gene *Fgf8,* a situation that could free them from normal interaction rules and enables them to act on another gene.

In order to determine if such a regulatory re-allocation is possible after such a structural change we have engineered a tandem duplication of this locus using in vivo mediated Cre (Hérault, Rassoulzadegan et al. 1998) recombination between loxP sites targeted to *Fgf8* (allele from (Meyers, Lewandoski et al. 1998)) and *Lbx1* (allele from (Vasyutina et al. 2005)). The duplication brought a copy of the 3'flanking region of *Fgf8,* which contains many regulatory elements contributing *Fgf8* expression, far from the remaining copy of *Fgf8,* but next to the duplicated copy of *Lbx1* (Figure 15). This duplicated copy of *Lbx1* is not functional, due to the insertion of a GFP reporter, under the control of endogenous *Lbx1* promoter. Importantly, while the expression of *Lbx1::GFP* is normally restricted to *Lbx1* expression domains (migrating myoblasts and a few neurons), in the context of the duplication, GFP was detected in *Fgf8* expression domains such as AER, forebrain, mid-hind brain boundary and branchial arches (Figure 15). This experiment demonstrated that upon chromosomal re-arrangements in this locus, new enhancer-promoter associations could take place leading to ectopic gene expression.

Lbx1 is coding for transcriptional factor, which determines migratory routes of muscle precursor cells (Schäfer and Braun 1999), (Brohmann et al. 2000) and contributes to neuronal patterning (Jagla et al. 1995). However, it is not known to be involved in limb patterning and growth, apart from its role in limb skeletal muscle formation. Interestingly, Marc Friedli (University of Geneva) found that *Lbx1* was mildly up-regulated in homozygous *Dactylaplasia* limb buds, using quantitative real time PCR on stage E11 whole limbs. However, this analysis also revealed a general

up-regulation of genes specific for the proximal limb bud and respectively a down-regulation of genes specific for distal limb bud (Marc Friedli, personal communication). Thus, we concluded these changes could be mainly due to the truncation of the distal limb observed in these animals, leading to a relative enrichment for proximal cells (including *Lbx1*-expresing muscle precursors). The duplication between *Lbx1* and *Fgf8* genes covers the duplicated interval in *SHFM3*, but did not lead to limb malformations in transgenic mice. However, in this allele any possible effect of *Lbx1* in *SHFM3* would be functionally masked, as the duplicated *Lbx1* gene has been replaced by GFP (which we indeed observed expressed ectopically).

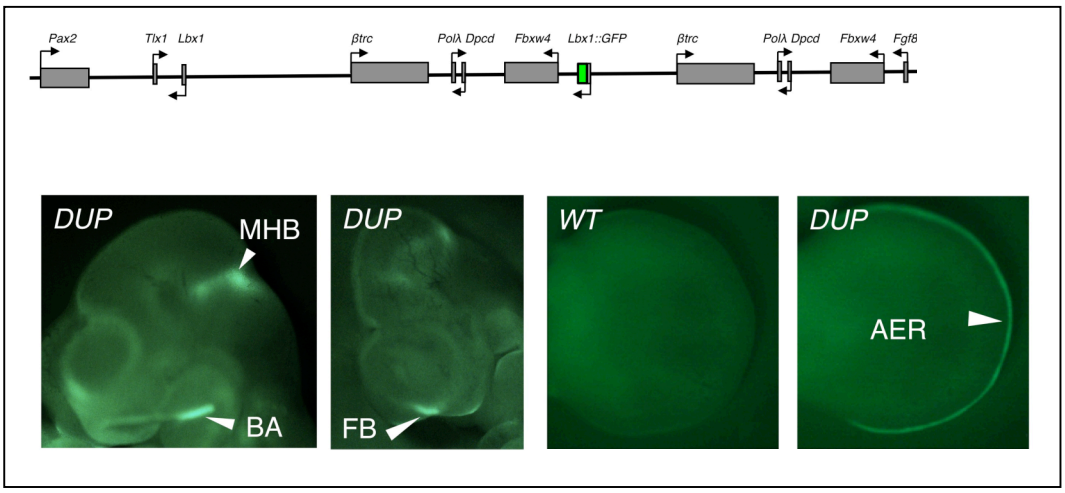


Figure 15. The response of GFP to position of enhancers

The duplication between *Fgf8* and *Lbx1* brings the 3' flank region of *Fgf8* close to *Lbx1* promoter driven GFP. As the regulatory elements that are sufficient to form *Fgf8* expression domains are in this locus, expression of the reporter is gained in the same tissues.

To test the consequences of ectopic expression of *Lbx1* in the AER, I cloned the cDNA of this gene under the control of the promoter region of *Msx2* gene, which is specially driving expression in the AER of the limb buds (Liu et al. 1994). This transgene was cloned into a lentiviral backbone to allow efficient production of transgenic animals by lentivirus-mediated transduction (Lois et al. 2002). The production of the virus and injection into mouse embryos was performed by Katja Langenfeld. Out of the 14 transgenic embryos collected at stage E14.5 and E16.0, we found four had duplicated or enlarged thumbs (Figure 16). This preaxial polydactyly was mostly observed in the hind limbs and was not fully penetrant as some animals only showed it in one of their appendages. Interestingly, this

malformation was similar to the polydactylous morphology observed in *Msx1*$^{-/-}$;*Msx2*$^{-/-}$ double null forelimbs (Lallemand et al. 2005). This phenotype correlated with the delay in AER regression, as *Msx* genes help control cell death in the anterior apoptotic domain (Lallemand et al. 2009). As *Lbx1* is structurally similar to *Msx* and has a similar binding sequence (ATTA), we hypothesize that *Lbx1* could compete with *Msx* genes and in a dominant negative way, interfere with the regulation of the *Msx*-downstream target genes. Thus, we found that ectopic expression of *Lbx1* led to abnormal limbs, but instead of an ectrodactyly due to premature death of the AER, we observed an additional digit, suggesting rather a maintenance/extension of the AER. Interestingly, *SHFM3* patients have also frequently proximally placed thumbs and/or triphalangeal thumbs (TPT) and preaxial polydactyly, in about 50% of the cases (Elliott et al. 2005), (Everman et al. 2006). These features were more frequently observed for patients with a duplication break point between *LBX1* and *βTRC*, while patient with a duplication that is extending up to the *LBX1* gene had classical central longitudinal deficiency. However, these comparisons are difficult, given the variability of the *SHFM* phenotypes and the limited number of patients with detailed mapping of the break point. Nevertheless, our analysis suggests that an over-expression of *Lbx1* could be responsible for the observed preaxial polydactyly, while another gene or factor is required for the ectrodactyly.

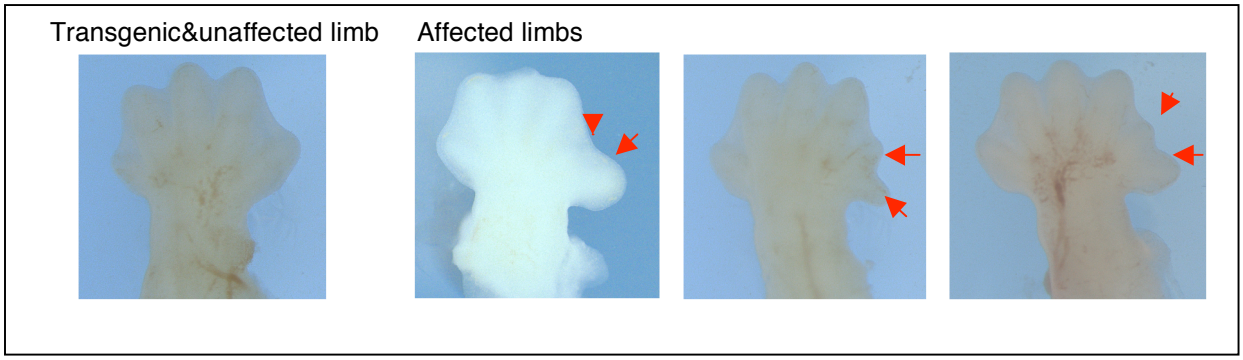


Figure 16. Limb phenotypes of *Msx2::Lbx1* transgenic embryos at E14.5
Arrows point the duplicated thumbs.

As mentioned in the introduction, a disruption of the Wnt/β-catenin pathway is leading to impaired AER maintenance due to increased apoptosis (Barrow, Thomas et al. 2003). *βTrc*, which is encoded by a gene located in the *Dactylaplasia* locus, binds to the phosphorylated N-terminus of β-catenin. This is causing the

ubiquitination of β-catenin by an E2 ligase and over-expression of *βTrc* promotes the down-regulation of β-catenin (Hart et al. 1999). In addition to this, β-catenin and a downstream mediator of the Wnt-pathway were reported to be down-regulated in limbs of *Dactylaplasia* embryos (Schwarzer W., PhD thesis, Freie Universität Berlin, 2010). Considering all these findings, we tested if *βTrc* expression in the AER could be the gene causing ectrodactyly in *Dactylaplasia* mice and *SHFM3* patients. We used the same strategy as for *Lbx1* expression in the AER via lentiviral-transduced transgenesis. Both the *Msx2* promoter and a compound driver including an AER *Fgf8* enhancer, which was characterized Mirna, were used for the forced expression experiments. With both of the drivers we did not observe ectrodactyly in E14.5 transgenic animals (thirteen and thirty-eight embryos respectively). Knowing that different isoforms of *βTrc* exist and they display differential activities in the regulation of Wnt signaling, we repeated the experiment with a cDNA for the isoform that was shown to have the strongest inhibition of the Wnt signaling in (Seo et al. 2009). We collected embryos at different stages (E14.5 and E18.0) and did not observe any limb malformations (thirty-nine and one embryos respectively). These results suggested that either the expression stage of *βTrc* in our assay differed from the real situation in mutant animals or this *Dactylaplasia* limb phenotype is caused by another gene.

In brief, we showed that chromosomal re-arrangements, similar to the one that is leading to *SHFM,* could alter the regulatory landscape of the locus and enable expression of the other genes in the locus in place of *Fgf8*. We showed that the ectopic expression of *Lbx1* in the AER could explain an aspect of the phenotypic spectrum observed in *SHFM3* patients. However, there seems to be a complex relationship between other genes causing variation of limb phenotypes and other syndromes. This diversity of phenotypes could be due to different combinations of interactions between genes and enhancers depending on the break points, which alter their relative locations. Therefore, changes in the enhancers' capacity to act on genes that could have detrimental consequences upon misregulation in some tissues might cause the syndromic forms of the disease.

5.1.5 Investigation of the influence of MusD on neighboring genes' expression in *Olig2/Olig1* locus

In order to investigate if the MusD-mediated gene regulatory changes were specific to the *Fgf8* locus, we aimed to examine possible gene expression changes associated with MusD insertion/expression in different loci. For this purpose, I collected a list of MusD insertions close to genes in the C57BL/6J reference genome using the BLAT algorithm with a full length MusD as an input. Only few full length MusD elements were found in vicinity of genes. This paucity of MusD elements in proximity to genes has already been reported by Maksakova et al. (Maksakova et al. 2009). It may indicate a preferential insertion of MusDs outside of these regions or reflect the consequence of a purifying selection against MusD insertions close to genes (i.e. especially if they interfere with the precise way of gene control). However, this search revealed the presence of a MusD element inserted on chromosome 16, in between the *Olig2* and *Olig1* genes (see Figure 17 for the locus). *Olig* genes are dynamically expressed from E9.0 to E14.5 in the spinal cord and are associated with neurogenesis (Lu et al. 2000) (Sun et al. 2001). These two genes are closely related bHLH transcription factors (*Olig1* is 98% similar to *Olig2* in the bHLH domain) and it is considered that this bigenic cluster has arisen via tandem gene duplication. Interestingly, these genes are expressed in the same cells and *Olig1* could complement the effects of a loss of *Olig2* for oligodendrocyte development in the brain (Lu et al. 2002), underscoring further their functional likeness. The similarities of expression of these two genes suggest that they might be controlled by same regulatory elements. Indeed regulatory elements, which drive expression of reporter genes in co-expression domains of *Olig* genes, were identified in the locus (Figure 17a,c) (Sun, Hafler et al. 2006), (Friedli, Barde et al. 2010). The MusD inserted there belongs to the same young-MusD family that includes the Dac^{1j}-MusD and Dac^{2j}-MusD elements. Since MusD insertions are rather polymorphic amongst mouse laboratory strains (Zhang et al. 2008), I used PCR to look at its distribution in different strains. I found that if this element is present in the C57BL/6J strain, it is absent in BALB/cJ and 129 strains. Since the effects of MusD in *Dactylaplasia* strains were strain-dependent and notably different between C57BL/6J and BALB/cJ because of the *Mdac* gene, I considered that this modifier gene could also eventually modulate the consequences of the presence of a MusD. For that purpose, I have backcrossed C57BL/6J mice with BALB/cJ and obtained the MusD element inserted in the *Olig2,1* locus (from C57BL/6J) with the mdac condition (from BALB/cJ). Both alleles were

genotyped by PCR, using the polymorphic SNPs for Mdac/mdac (from genetic mapping) and the region that flanked the MusD in *Olig2,1* locus. Thus, I obtained a stock of animals that were homozygous for the MusD insertion in *Olig2,1* locus and for the permissive modifier allele (*Olig2,1-MusD*^{hom}*; mdac/mdac,* depicted as mdac). Importantly, bisulfite sequencing analysis showed that the Olig2,1-MusD element is methylated in C57BL/6J, while it is unmethylated in the mdac/BALB/cJ background (see Figure 20 later).

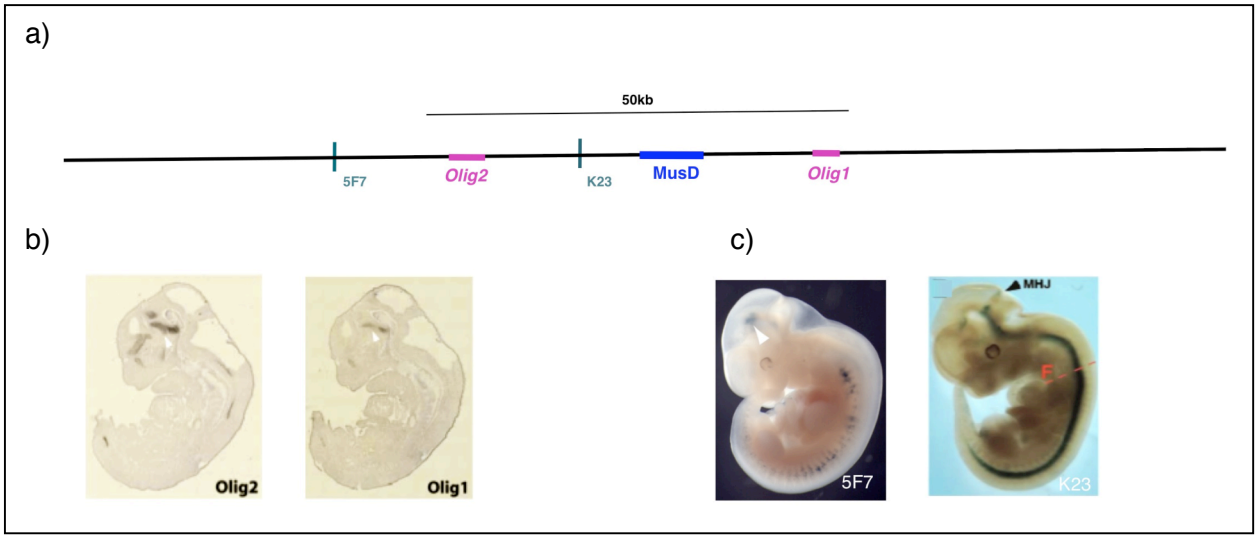
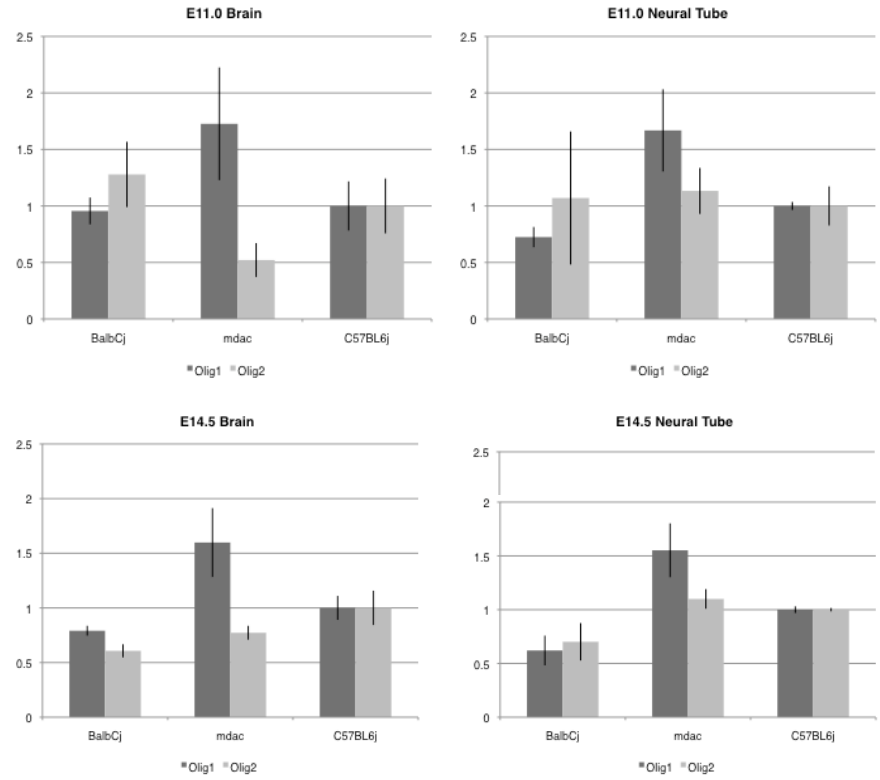


Figure 17. *Olig2/Olig1* locus

a) The representation of the genes and the MusD in the locus. b) Expression domains of Olig2 and Olig1 genes detected by in-situ hybridization on E11.5 embryo sections (pictures taken from (Friedli et al. 2010)). c) Two of the conserved sequences characterized as regulatory elements in the locus by transgenic assays (pictures taken from (Friedli, Barde et al. 2010) and (Sun et al. 2006)) The elements are depicted as lines in a).

Then, I compared *Olig* genes expression levels when MusD was absent (BALB/cJ), was present and methylated (C57BL/6J) and was present but unmethylated (mdac). Quantitative real time PCR was carried out using E11.0 and E14.5 brain and neural tube tissues by normalizing the test genes to multiple housekeeping genes (*Pgk1, Actb* and *Gusb*). This analysis revealed an overall increase of *Olig1* gene expression for all the tissues and stages of mdac embryos checked but most significantly for E11.0 for brain tissue as *Olig1* gene level has increased whereas *Olig2* gene level has decreased in comparison to embryos from parental strains (Figure 18a). This led to a change in the ratio between the expression level of these two *Olig* genes at this stage, in case of the presence of an unmethylated MusD element in the locus, while I observed no significant difference between two parental strains (Figure 18b).

a)



b)

Stage & Tissue	*Olig2/Olig1* ratio		
	BALB/cJ	mdac	C57BL/6J
E11.0 Brain	13.85	3.02	9.58
E11.0 Neural Tube	17.91	7.96	12.68
E14.5 Brain	1.75	0.94	2.31
E14.5 Neural Tube	1.04	0.73	1.04

Figure 18. *Olig* genes expression levels in E11.0 and E14.5 embryos

a) *Olig1* and *Olig2* were normalized to multiple internal controls for each sample, then all genotypes were compared to C57BL6j. Error bars represent the standard error of the mean between biological replicates (four different RNA preps for mdac, three different RNA preps for BALB/cJ and C57BL6j) b) ratios of two genes for each strain.

These findings suggested that the MusD insertion altered the expression of the *Olig* genes, depending on the methylation status of the MusD, in a manner analogous to what has been observed for the *Fgf8* locus. In collaboration with Ben Martynoga and François Guillemot (MRC, London), we have investigated the enhancer potential of four conserved sequences in the locus that are bound by transcription factors specific to neural cell differentiation. In addition to one element from our assays there was another element from VISTA enhancer browser that is downstream of *Olig1* gene and both elements were driving the expression of the reporter gene (LacZ) partially in

the brain expression domains of *Olig* genes (not shown). MusD element is located in between these elements and the *Olig2* gene, indicating an enhancer blocking activity of MusD when taken into account together with the down-regulation of *Olig2* gene. The changes in relative levels of *Olig* genes for mdac embryos were less significant at E14.5 (Figure 18b). This could be explained by a compensatory mechanism in ratios of two genes during development, similar to the one which has been described in the *Olig2* knock-out situation (Zhou and Anderson 2002), (Lu, Sun et al. 2002). Therefore, an alternative hypothesis would be that MusD is only diminishing the expression of *Olig2* and *Olig1* is up-regulated due to an independent effect.

In addition to these three strains I have tested a forth one obtained by the cross of *Olig2,1-MusD^{hom}; mdac/mdac* animals with C57BL/6j animals and collected embryos from this breeding at same stages as for other strains. These embryos are heterozygous for *Mdac* locus and were expected to be equivalent to C57BL/6j embryos in this locus, as MusD would be silent. However, we found that this MusD was methylated in the brain tissue in a heterogenous manner (see Figure 20 later), and accordingly the *Olig* gene levels in these embryos were more similar to levels in mdac embryos.

This analysis of a different locus showed that the impact of MusD elements on endogenous gene regulation is a prevalent feature and that their potential to interfere with gene expression modules is not restricted to the *Fgf8* locus. It also emphasized that these effects are linked to their transcriptional and epigenetic status, which appeared to be under the control of the *Mdac* locus.

5.2 Epigenetic control of MusD silencing

5.2.1 *Mdac* dependent differential methylation of MusD 5'LTRs

1. Effects of *Mdac* on MusD

The *Dactylaplasia* mice were bred to several other strains, some of the breeding gave only offspring with normal limbs (Chai 1981), suggesting that the manifestation of the *Dactylaplasia* phenotype is depending on a second unlinked locus (*Mdac*). Later on, this second locus was mapped to mouse chromosome 13 (Johnson, Lane et al. 1995). This correlation of the modifier locus with the *Dactylaplasia* phenotype could be explained by two different mechanisms. According to first model, *Mdac* may act downstream of the *Dactylaplasia* mutation, for example by counteracting the factor(s) that leads to cell death in the AER. In the second one, *Mdac* could act in parallel to the *Dactylaplasia* mutation and help the survival of the AER in a

Dactylaplasia dependent manner. As reported before, *Mdac* does not only control the manifestation of the ectrodactyly in *Dactylaplasia* mice, but also affects the expression of the inserted MusD causing the mutation and the antiparallel changes of *Fgf8* (in Results part 1). These findings indicated that *Mdac* is acting prior the down-regulation of *Fgf8* and limb phenotype. As summarized in the "Introduction", retroviral sequences are controlled by many different epigenetic mechanisms, which play an important role in controlling their transcriptional activities. Therefore, we considered that *Mdac* might be acting on MusD via an epigenetic silencing mechanism targeted either on the element itself or that could alter the degree of specificity of enhancer-promoter communications within the locus. In order to address this question, I used bisulfite sequencing to explore the cytosine methylation levels of MusD LTRs on DNA samples from permissive (*Dac^{2j}-MusD^{het};mdac/mdac*) and resistant (*Dac^{2j}-MusD^{het};Mdac/mdac*) strains. I performed this experiment with two tissues (brain and limb) where MusD transcripts are detected in the homozygous mutants (see Results part 1) and one tissue where no MusD expression is not detected (heart). All tissues were isolated from E10.5 embryos. I saw a very significant different degree of CpG methylation of MusD-5'LTR between permissive (mdac) and resistant (Mdac) embryos, that is extending into MusD sequence. Resistant (Mdac) embryos had excessively methylated 5'LTR regions whereas permissive (mdac) embryos almost completely lacked DNA methylation inside and around the 5'LTR (Figure 19). This differential methylation also extended into the MusD sequence downstream of the 5'LTR. However, the 3'LTR region of the MusD was only mildly methylated, and to the same extend in resistant (Mdac) and permissive (mdac) samples. This finding provided an internal control for our assay and also showed that the spreading of the DNA methylation found in resistant (Mdac) animals over the Dac^{2j}-MusD was limited to its 5'end. Altogether, we found that the 5'end of the Dac^{2J}-MusD was differentially methylated for all the tissues tested depending to the allele of the modifier locus. Importantly, this was also true in tissues where MusD was not expressed (heart) or in tissues where the proportion of cells showing MusD expression (e.g. AER for limbs) was low in comparison to the whole tissue size. This is suggesting that the lack of MusD expression observed in resistant (Mdac) animals was due to the epigenetic silencing of the transposon, and not the opposite. Conversely, it suggests that the capacity of MusD to interfere with *Fgf8* expression was blocked by this Mdac-dependent

Figure 19. DNA methylation status of Dac2j-MusD LTRs

Empty boxes represent unmethylated and full boxes respresent methylated positions. Blue rectangle marks the CG positions in LTRs

epigenetic modification, as its silenced LTR could not titrate or block *Fgf8* remote enhancers. Noteworthy, during the course of this study, Kano et al. showed differential methylation of *Dac*^{1j}*-MusD* between permissive and resistant strains, further supporting my observations (Kano, Kurahashi et al. 2007). In order to see if

there was a locus-specificity for Mdac dependent methylation of MusD-5'LTR, I examined the degree of cytosine methylation in the 5'LTR for the MusD inserted in the *Olig2-1* locus, in the three types of embryonic tissues used before. This example was particularly interesting because in this case, the MusD insertion occurred in a non-permissive strain (C57BL/6J). In this strain, similar to what was found in *Dac^{1j}-MusD* and *Dac^{2j}-MusD*, the 5'LTR of the MusD element was strongly methylated in resistant (Mdac) animals (e.g. in the limbs) (Figure 20). However, when this MusD was brought in a permissive (mdac) background, by breeding it to BALB/cJ, its 5'LTR became unmethylated. In two other tissues (heart, brain), I found the same differential DNA methylation, even though some significant unmethylated LTRs were present in these tissues of resistant (Mdac) embryos, suggesting some heterogeneity, either in the cell population, or in the epigenetic marks present over this LTR (Figure 20). Yet, these results showed that Mdac silencing action on MusD element is mostly position and cell-type independent. Furthermore, this methylation process seems to be occurring somatically, as embryos obtained from the breeding of permissive strain to the resistant strain had methylated the MusD-5'LTR.

2. Comparison of the effects of *Mdac* on MusD and ETnII methylation

Non-autonomous ETnII elements share similar LTR regions with MusD elements (Mager and Freeman 2000). Therefore, LTR methylation status of ETn elements would be informative for the specificity of MusD methylation. I examined the epigenetic status of the 5'LTR of three ETnII sequences that were common to both mdac (BALB/cJ) and Mdac (C57BL/6J) strains and another ETnII that is specific to C57BL/6J that was brought to mdac (BALB/cJ) background via the backcrossing with BALB/cJ strains (next to *Frmbd4* gene). We did not consider ETnIIs localized in regions corresponding to constitutive heterochromatin, but only ones that are in similar genomic context (around active genes) to the MusD elements analyzed previously. One of these selected EtnII sequences resides inside a protein-coding gene (Chr1, *SH3bp4*) and the other three are around other protein-coding genes (Chr6 and 11). The element found only in C57Bl/6J was brought into mdac (BALB/cJ) background via outcrossing with BALB/cJ strains, like describe for Olig2,1-MusD. Indeed, I used the very same DNA samples that showed differential methylation for MusD for this analysis. In contrast to the MusD, none of these ETnII sequences were differentially methylated on their 5'LTR between permissive (mdac) and resistant (Mdac) embryonic tissues (Figure 21). Thus, Mdac seemed to discriminate between

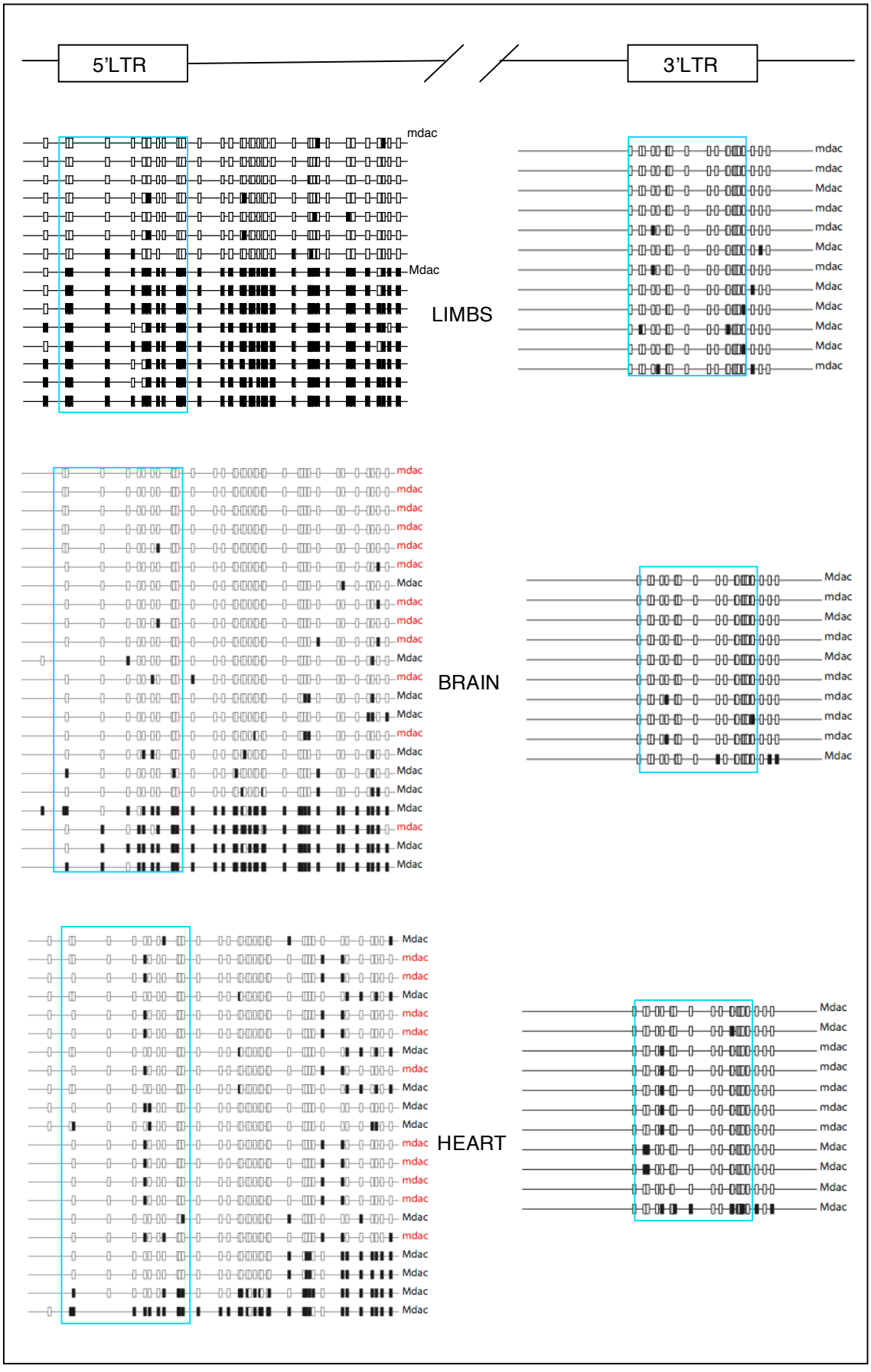


Figure 20. DNA methylation status of Olig2-1-MusD LTRs

Empty boxes represent unmethylated and full boxes respresent methylated positions.

Blue rectangle marks the CG positions in LTRs

ETnII and MusD, despite their almost identical LTR sequences. As the methylation extended further inside the MusD element, it is suggesting that Mdac could initiate the methylation of or possibly recognize specifically the regions that are present in MusD (e.g. *pol, pro, gag*) but have been lost in their ETnII derivatives. Accordingly, Maksakova et al. showed a difference in the extent of cytosine methylation between ETnII and MusD elements (Maksakova, Zhang et al. 2009), suggesting that these elements were controlled differently.

To compare more globally the methylation of ETnII and MusD in BALB/cJ and C57BL/6J mice, we compared the Southern Blot profile of samples digested with methylation-sensitive restriction enzymes and hybridized with specific probes. This method was used for MusD and ETnII repeats in cell lines (Maksakova, Zhang et al. 2009). I have used the same method on two DNA samples from BALB/cJ and C57BL/6J E10.5 brain tissue. The first combination of enzymes (PstI+XhoI) was not blocked by cytosine methylation, therefore just released DNA fragments of MusD and ETnII at any location they are inserted and represented the input. The second (PstI+NotI) and third (PstI+AfeI) sets of enzymes only cut if the cytosine in corresponding restriction site in the 5'LTR region or respectively in the common region was unmethylated. Importantly, the fragment sizes generated for MusD and ETnII were different, allowing to distinguish them even when the common region of these two repeat sequences was used as a probe. Altogether, these three combinations provide global information regarding the methylation status of MusD and ETnII at two cytosine positions in the 5'LTR and the common region. They revealed that BALB/cJ and C57BL/6J tissues showed differential methylation at 5'LTR for MusD, albeit affecting only a small portion of elements (Figure 22). The majority of MusD/ETnII elements was methylated independently of the strain, considering many elements are away form genes and probably found in heterochromatin regions this is not unexpected. Therefore, Mdac allele does not control DNA methylation as globally as DNMT1 (in *Dnmt1* mutants, MusD underwent a dramatic loss of methylation in ES cells (down to ~20%) (Dong, Maksakova et al. 2008)). However, the specific band corresponding to unmethylated MusD elements (as shown in Figure 22, second half of the blot) was much stronger in BALB/cJ than in C57BL/6J, despite a similar number of elements in the two strains, as judged by the intensity of bands in the first lane. ETnII elements seemed also less methylated in BALB/cJ embryos, as the bands obtained with enzyme combinations 2 and 3 digestions are more intense compared to C57BL/6J

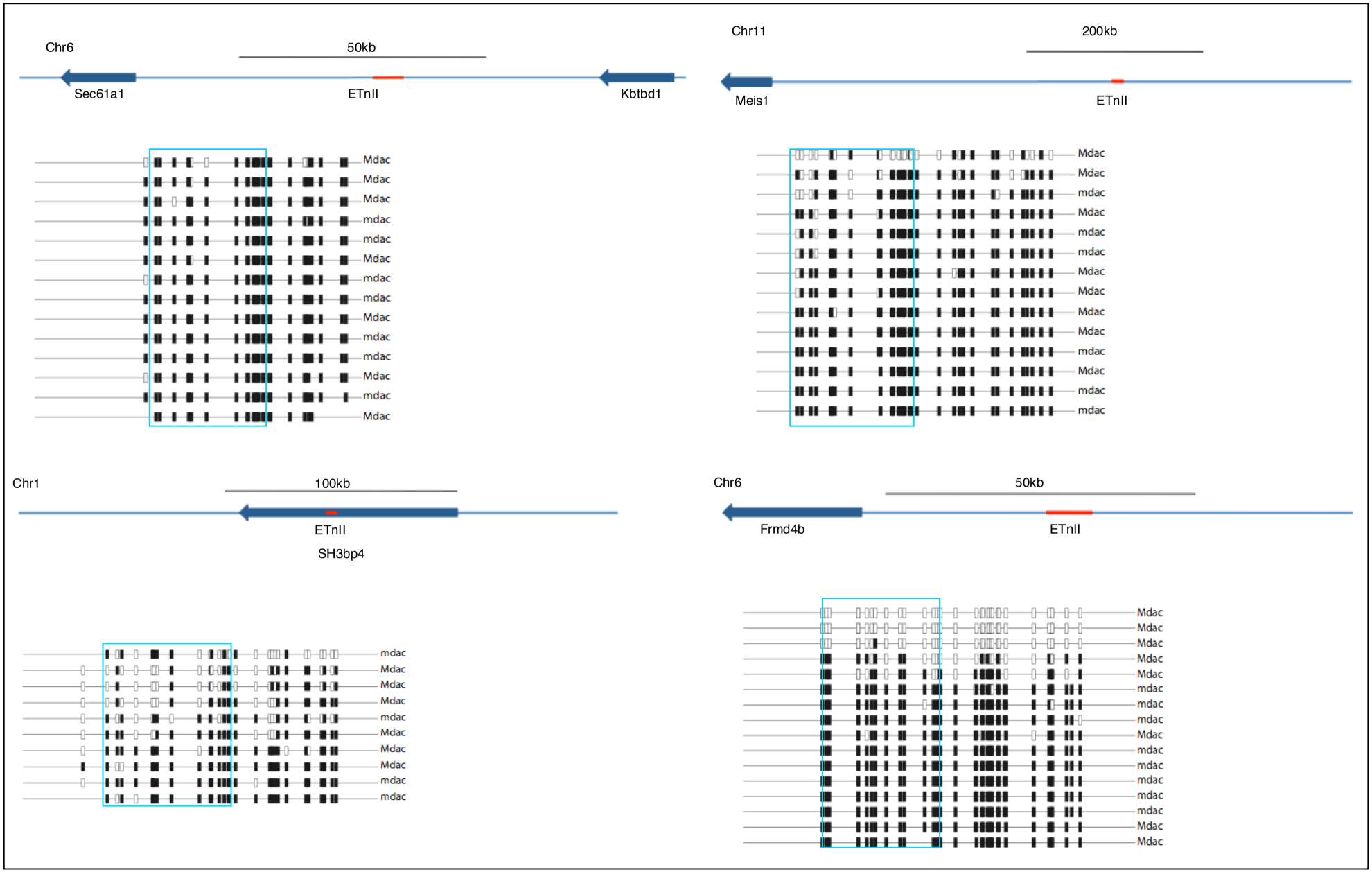


Figure 21. Methylation status of ETnII 5'LTRs in resposnse to *Mdac* locus

LTR region is marked with blue rectangle

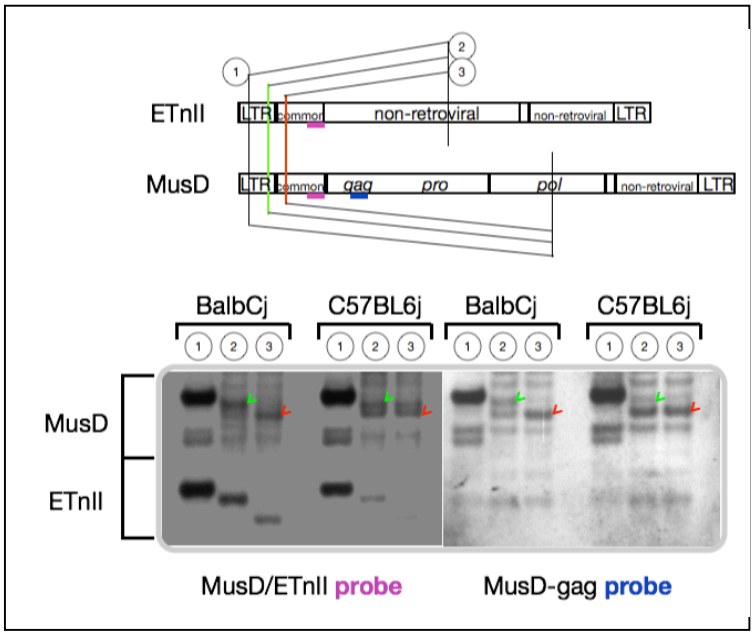


Figure 22. Southern Blotting for DNA methylation detection DNA was extracted from E10.5 brain tissue (and the rest of the embryo as well, which did not differ from the result shown here) and digested with three different combination of enzymes. Combination number 1, releases the MusD/ETnII elements, thus lane1 is for input. Combinations 2 and 3 cut only if cytosine is unmethylated.

lanes; however in this case the first lane (input) showed also more intense signal, reflecting probably the higher copy number of ETnII elements in BALB/cJ strain.

Next, I explored the differences in the expression of MusD or ETnII elements between C57BL/6J (resistant) and BALB/cJ (permissive) strains by quantitative real time PCR on E10.5 embryos (same developmental stage that was used for DNA methylation analysis). MusD was not differentially expressed in these two strains, most likely due to majority of the silenced population at an earlier stage of development (Figure 23). In contrast, ETn family transposons were significantly more expressed in BALB/cJ strain (Figure 23). Furthermore, this ~10 fold more expression of ETnII elements was halved in the hybrid embryos, that are obtained by crossing BALB/cJ and C57BL/6J strains (Figure 23). This effect was probably a reflection of the changes in copy numbers, which is consistent with the observations on the Southern Blot (Figure 22, BALB/cJ vs. C57BL/6J ETnII intensity comparison in the first lane). Indeed, ETns are known to differ in copy number and insertion site among different inbred strains (Zhang, Maksakova et al. 2008). Therefore, the differences in the expression levels of ETns may correlate with the activity of enhancers that they can trap in the loci they are inserted into. Moreover, it is known that ETn elements use MusD retrotransposition machinery (Ribet, Dewannieux et al. 2004) and therefore in the permissive strain (BALB/cJ), where some MusD elements are unmethylated and transcribed, probably had an increased chance of retrotransposition and expanding.

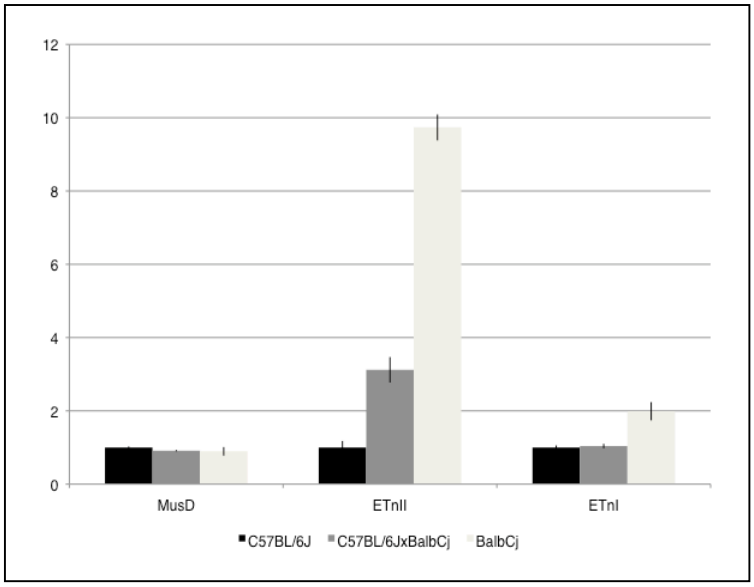


Figure 23. Expression of MusD/ETnII elements at E10.5 Quantitative real time PCR was performed on whole embryonic extracts of three biological replicates. MusD, ETnII and ETnI expression was normalized to multiple houskeeping control genes (Actb and Pgk1), then each genotype was normalized to C57BL/6J.

From this analysis, we concluded that *Mdac* allele is not a factor counteracting the cell death in the AER or helping AER maintenance and survival in modifying *Dactylaplasia* phenotype. Instead, the primary target of Mdac is the MusD elements themselves. Mdac controls the epigenetic status of MusDs, which is reflected by the methylation of their 5' LTR. In doing so, Mdac influences the capability of MusD to trap enhancer elements and prevents them from interfering/blocking the expression of surrounding genes. Furthermore, dominant mode of inheritance of *Mdac* suggests that it is a gene that is present in the resistant C57BL/6J strain but not in the permissive BALB/cJ strain. Importantly, the resistant allele from one of the parents could silence a MusD insertion provided by the other parent in all tissues, therefore Mdac is acting on MusDs somatically. These effects of Mdac seemed to be more prominent on a subset of MusDs, as the majority of MusDs was likely to be silenced already via other mechanisms that are probably redundant to the action mechanism of Mdac. This also suggests that Mdac has a more specific function to silence potential escapers of a more global silencing action, so it may immediately act on MusDs somatically. Another possible explanation for the general silenced status of MusD elements is that they could be mostly localized in heterochromatin portion of the genome.

5.2.2 Mapping of the *Mdac* locus

From this study and another one (Kano, Kurahashi et al. 2007), *Mdac* emerged as a new gene controlling the silencing of a specific type of ERV. Despite

its broad mapping to chromosome 13, the nature of *Mdac* was still unknown. In order to identify *Mdac*, we designed a strategy to map this locus finely using a genetic segregation approach. For this purpose, we selected SNPs from the Perlegen SNP database (Jackson Laboratory, http://phenome.jax.org/) that are associated with restriction enzyme site differences between BALB/cJ and C57BL/6j. I have designed 21 pairs of primers spanning these SNPs covering slightly more than the locus mapped by Johnson et al (Johnson, Lane et al. 1995). The breeding scheme for mapping involved parents that are heterozygous for the MusD insertion over a resistant background (*Dac2j*^{het}*; Mdac*$^{C57BL/6j/BALB/cJ}$) and that are permissive (BALB/cJ) to *Dactylaplasia* phenotype, so none of the parents had malformed limbs. The progeny was screened initially for limb malformation, then the presence of the *Dac*^{2j}-MusD. The origin (C57BL/6J or BALB/cJ) of the chr13 was deduced from the segregation patterns of the specific SNPs allowed to distinguish between the two strains (summarized in Table 1).

Generation	Genotype	Phenotype
Parent	*Dac2j*^{het}*; Mdac*$^{C57BL/6j/BALB/cJ}$	WT
Parent	*Mdac*$^{BALB/cJ/BALB/cJ}$ *(mdac)*	WT
F1	*Dac2j*^{het}*; Mdac*$^{C57BL/6j/BALB/cJ}$	WT
F1	*Dac2j*^{het}*; Mdac* $^{BALB/cJ /BALB/cJ}$	*Dactylaplasia*

Table 1. Breeding scheme for the mapping of *Mdac* interval.
Parents and the screened offspring listed here.

Genomic DNA from 215 Dac^{2j}-MusD^{het} animals was examined for recombination in the *Mdac* interval by controlling initially for SNPs rs61801304 and rs52665098 (Chr13: 54232729-76357515, shown as light orange in Figure 24). We found that the limb phenotype was 99% penetrant with the status of the mapped *Mdac* interval considering these markers. Twenty-one of 215 offspring exhibited recombination between rs61801304 and rs52665098, and were further genotyped with the additional SNPs located in between to narrow down the corresponding break point. Among these, two independent recombination events placed *Mdac* within a 1.3 to 1.7Mb interval (Figure 24). This new interval excluded the previously considered strong candidates based on their involvement in limb development, such as *Fgfr4*, *Ror2*, *Msx2* and *Ptch1*. According to genome annotations, we initially found that the

new *Mdac* interval contained a cluster of *Cathepsin* genes, two fructose bisphosphatase genes and a large cluster of tandem Zinc finger proteins that contributed highly to repeated nature of the interval (Figure 25).

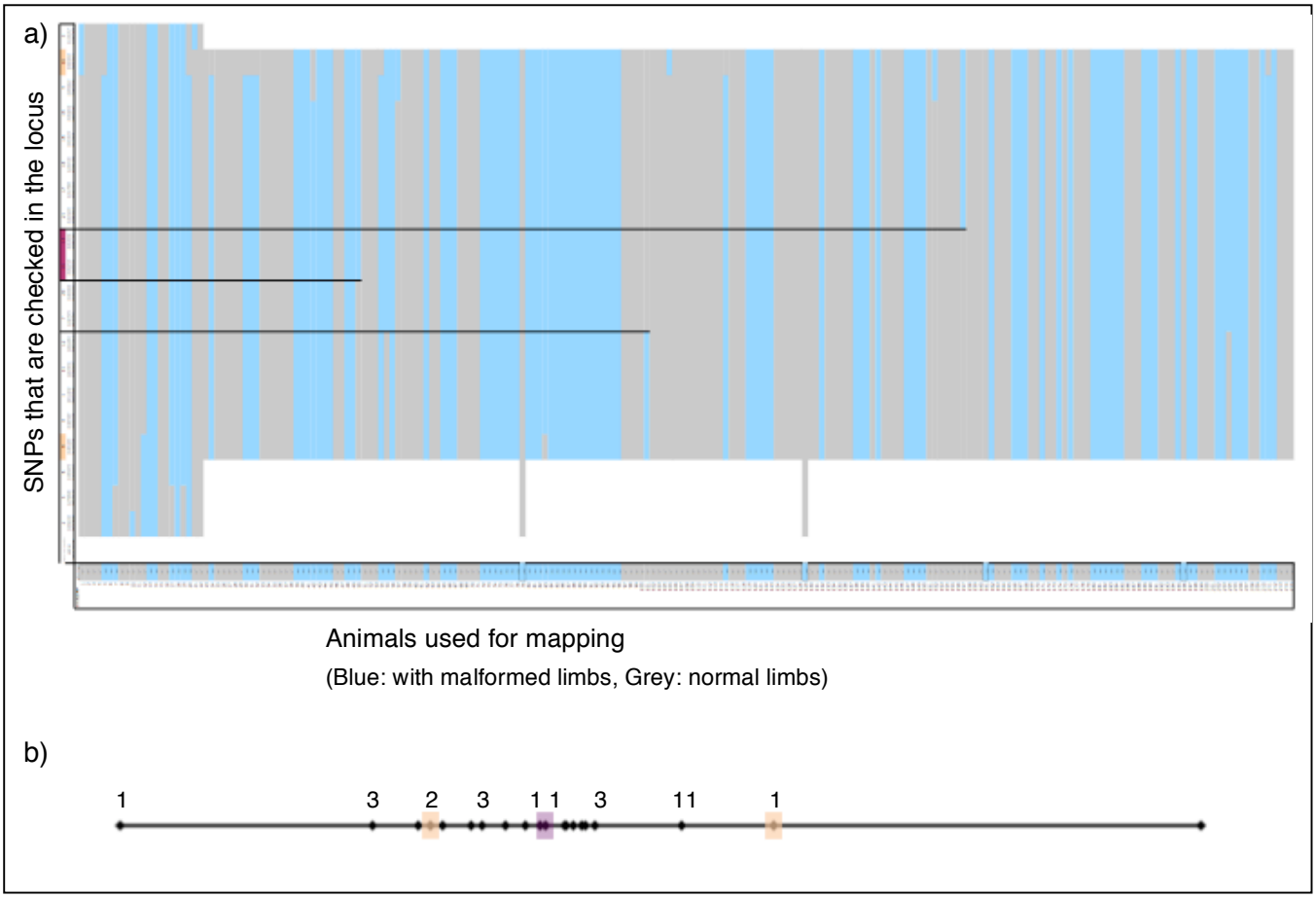


Figure 24. *Mdac* locus mapping

a)In total 215 *Dac*^{2j}*-MusD*^{het}; *Mdac*^{het} animals were mapped. Two animals formed the smallest boundary for the locus (marked with lines). Starting positions for mapping included SNPs at Chr13: 34203888 and Chr13: 103907429 (~70Mb), the positions for the smallest locus included SNPs at Chr13: 61653221 and Chr13: 62981912 (~1.3Mb). b) relevant positions of the SNPs that were used in the mapping and the numbers above represent the animals that had break points at corresponding positions.

Orange rectangles in b) represent the SNPs that were used for initial screen and if there was a break point between these SNPs, the mapping was continued until the break point was found. Pink rectangles represent the SNPs that gave the smallest interval.

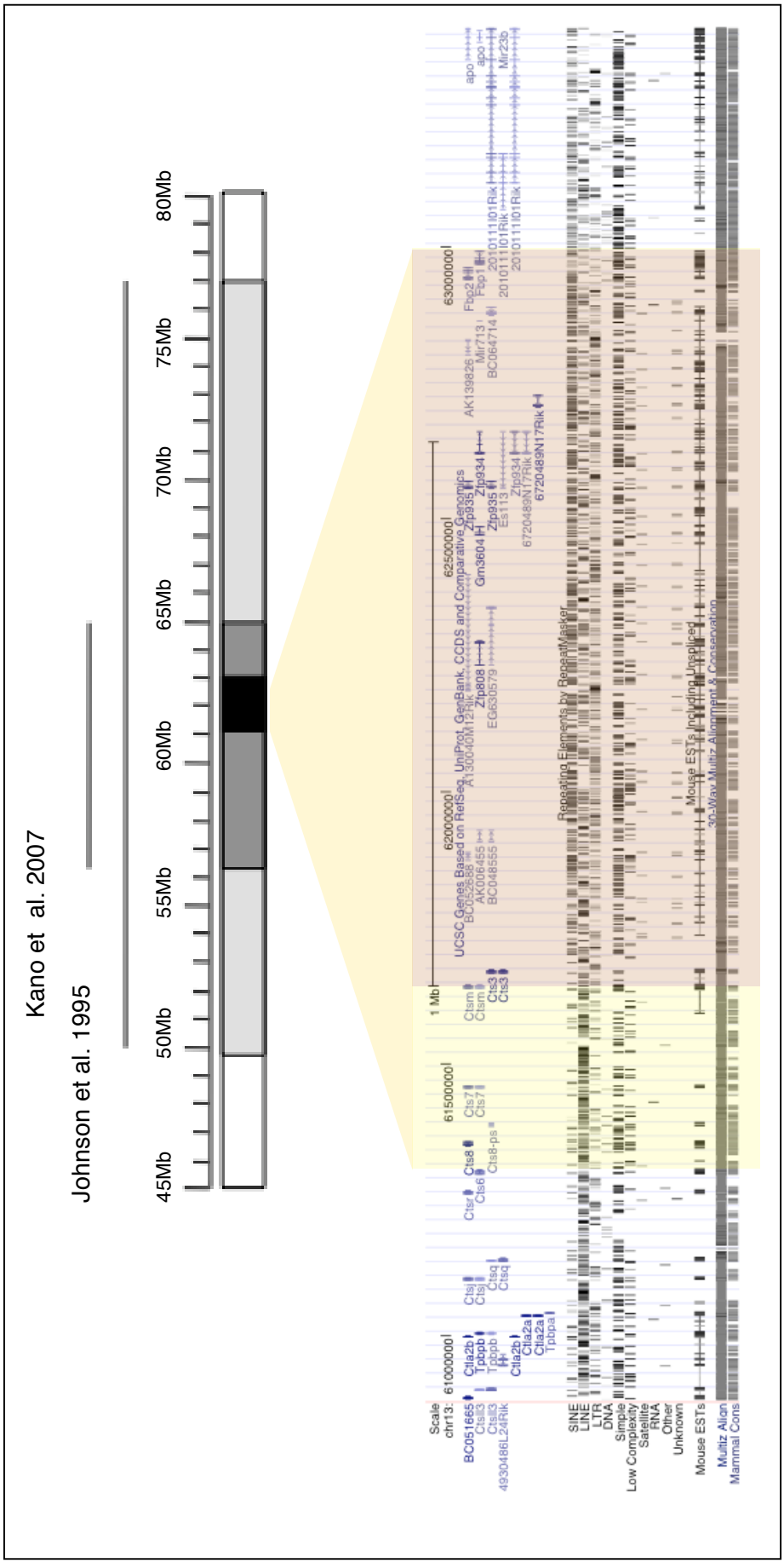


Figure 25. Mapped *Mdac* locus

Light yellow color represents the biggest interval and the pink overlap represents the smallest interval for *Mdac*. Therefore the size of the mapped locus is 1.3-1.7Mb. The mapped interval from previous studies is shown as grey lines above the genomic scale (Johnson et al. 1995), (Kano et al. 2007).

5.2.3 Characterization of the *Mdac* locus

To better characterize the *Mdac* interval I used other genomic resource to compare its organization in other known resistant and permissive inbred strains. Johnson et al. reported SWR/J, AKR/J, C3H/J, DBA/2J strains as resistant in addition to C57BL/6J and 129/S1ImJ, NZW/LacJ, SMc/Ckc and LG/Ckc strains as permissive in addition to BALB/cJ (Johnson, Lane et al. 1995) and NOD/ShiLtJ was later shown to be a permissive strain (Schwarzer W., PhD Thesis, Freie Universität Berlin, 2010) (summarized in Table 2). In the wild derived mouse strain CAST/EiJ the inheritance of the phenotype was found to be variable, suggesting a partial activity by *Mdac* in this strain (Johnson, Lane et al. 1995). Firstly, the coverage from a 129S7/SvEv BAC library over this interval was specifically much lower in comparison to BAC library from the C57BL/6J genome or to the immediate surrounding genes (Figure 26, Ensembl data). A comparative study probing for segmental duplications and/or CNV regions amongst mouse genomes from different strains, in which a microarray platform that contained probes against previously identified variable regions was used (She et al. 2008), generated a valuable resource for structural variations for inbred strains. I have compared several of the known permissive and resistant strain on the browser provided by this study. The *Mdac* interval seemed to be partially deleted in the permissive strains 129S1/SvJ, A/J, BALB/cJ and NOD/LtJ (Figure 27).
All these data pointed out consistently a copy number variation in the locus where *Mdac* was mapped that also correlated with the permissiveness of the strain for *Dactylaplasia* phenotype. It corresponded to ~0.5Mb of sequences that were present in the C57BL/6J reference genome, but deleted in a number of other strains.

Inbred strains	**status for *Dactylaplasia* permissiveness**
SWR/J, AKR/J, C3H/J, DBA/2J	Resistant, normal limbs.
129/S1ImJ, NZW/LacJ, SMc/Ckc, LG/Ckc, NOD/LtJ	Permissive, malformed limbs.
CAST/EiJ	Resistant F1 hybrids, backcrosses lower penetrance than expected.

Table 2. Mouse inbred strains Mdac allele status.

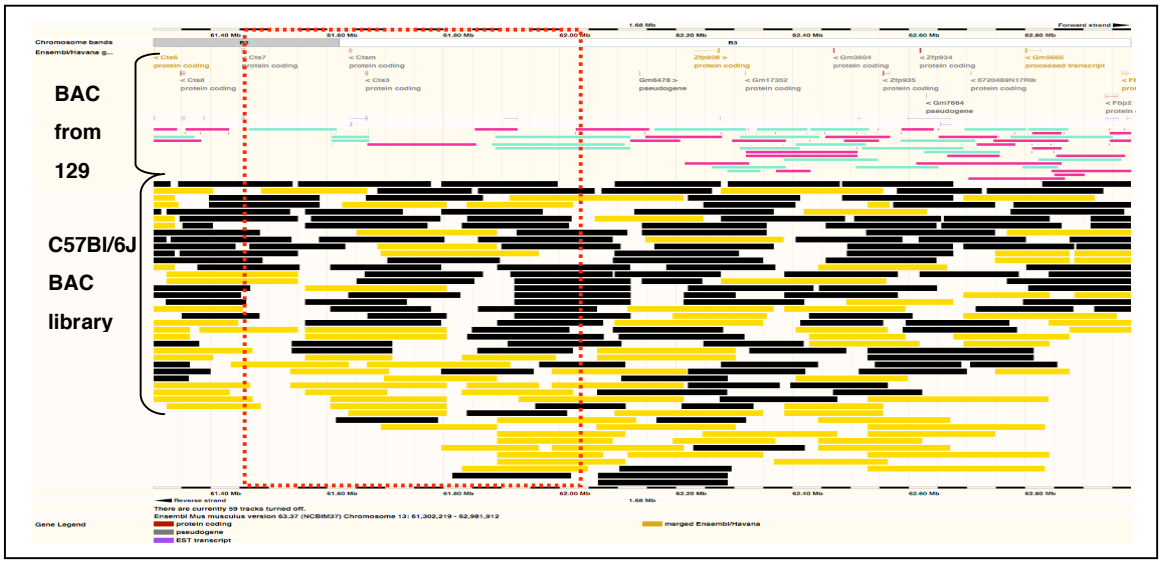


Figure 26. *Mdac* interval BAC coverage from 129 strain to C57BL/6J

Low BAC coverage of the permissive strain in the mapped *Mdac* interval marked with the red frame.

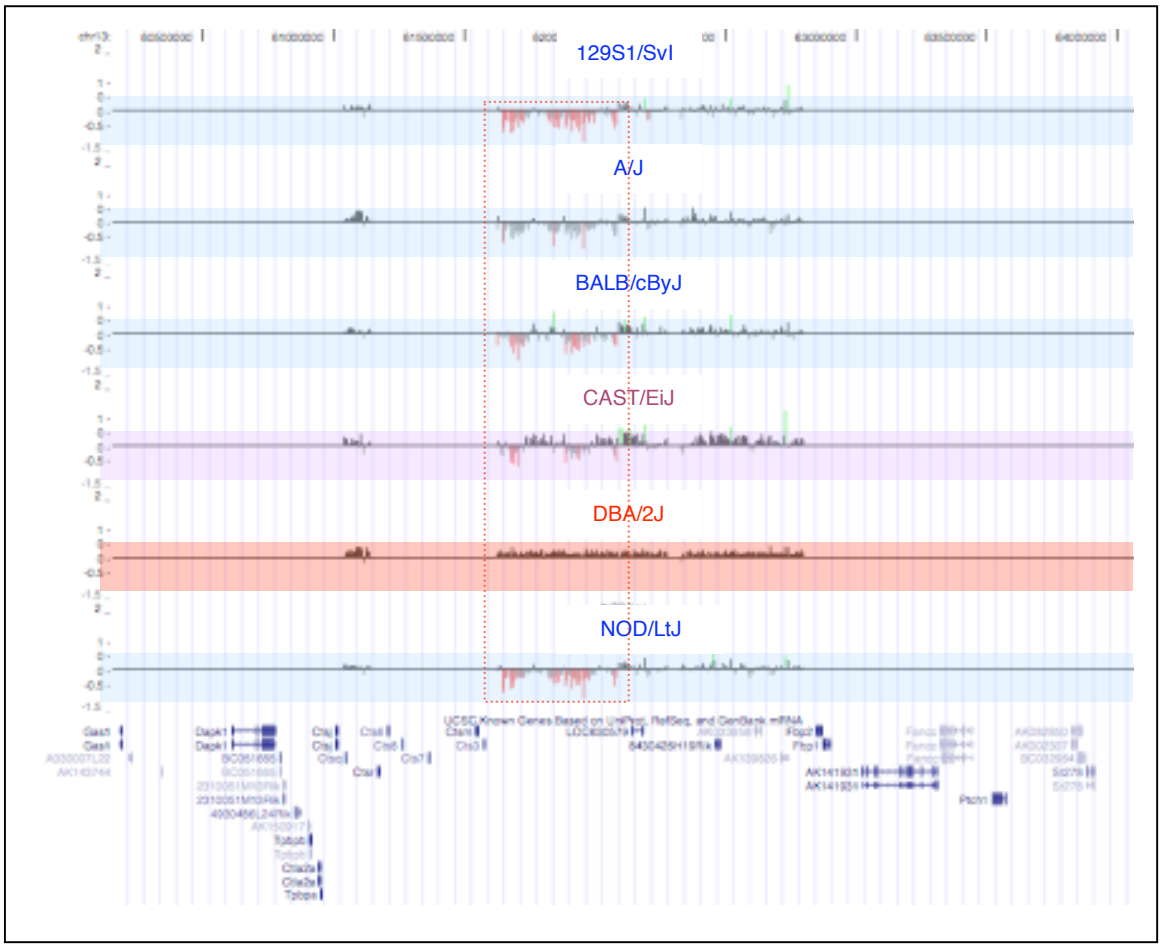


Figure 27. Comparison of mouse inbred strains for copy number variations (CNVs)

Resistant strain shown by red shade and the permissive strains with blue shade, all strains are compared to C57BL/6J genome (resistant strain). Red bars on the genome bar indicate missing positions in that particular strain. (assembly MGSCv3 (mm2) mouse paralogy)

If this interval was predicted to be deleted in the genome of the inbred strain, that strain was permissive to *Dactylaplasia* phenotype. Given the repetitive characteristics of the interval, it was difficult to design specific probes or oligos detecting a single part of the locus, therefore I designed only one pair of oligos that amplified a single product and I found that was missing in the permissive 129 and BALB/cJ genomes (Figure 28). Then, taking this as the start point (Chr13:62013106-62013880), the presence of PCR products that could be amplified from C57BL/6J genome was controlled in BALB/cJ products. I performed the detailed analysis of the boundaries for the deleted region from permissive strain (BALB/cJ) using oligos that could amplify multiple products spanning the locus. Each product was distinguished after sequencing by analysis of product-specific polymorphic nucleotides. Initially, three pairs of locus specific oligos, which amplify multiple products, were used on genomic DNA from both strains and the PCR products were directly sequenced. I examined the chromatogram results of the sequencing at polymorphic bases that were specific to one product out of the pool of products that are predicted. I confirmed the presence of a product by characterizing three or more polymorphic bases unique for this product. Some of the predicted products were not amplified in either strain. The presence/absence call of the specific markers was quite consistent and identified a contiguous interval. This direct genotyping analysis indicated the presence of a deletion in the permissive BALB/cJ and 129 strains, between Chr13:61748793-62269482 interval (Figure 29b). Interestingly, this interval overlapped with the proposed deletion based on micro-array-based analysis (She, Cheng et al. 2008).

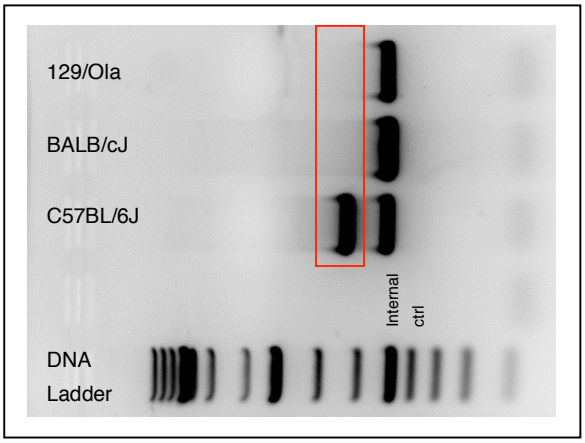


Figure 28. Missing region in permissive strains

The start point of PCR based analysis for deletion in the *Mdac* interval. PCR was performed with two airs of oligos at the same time, internal control amplifies a region on Chr5.

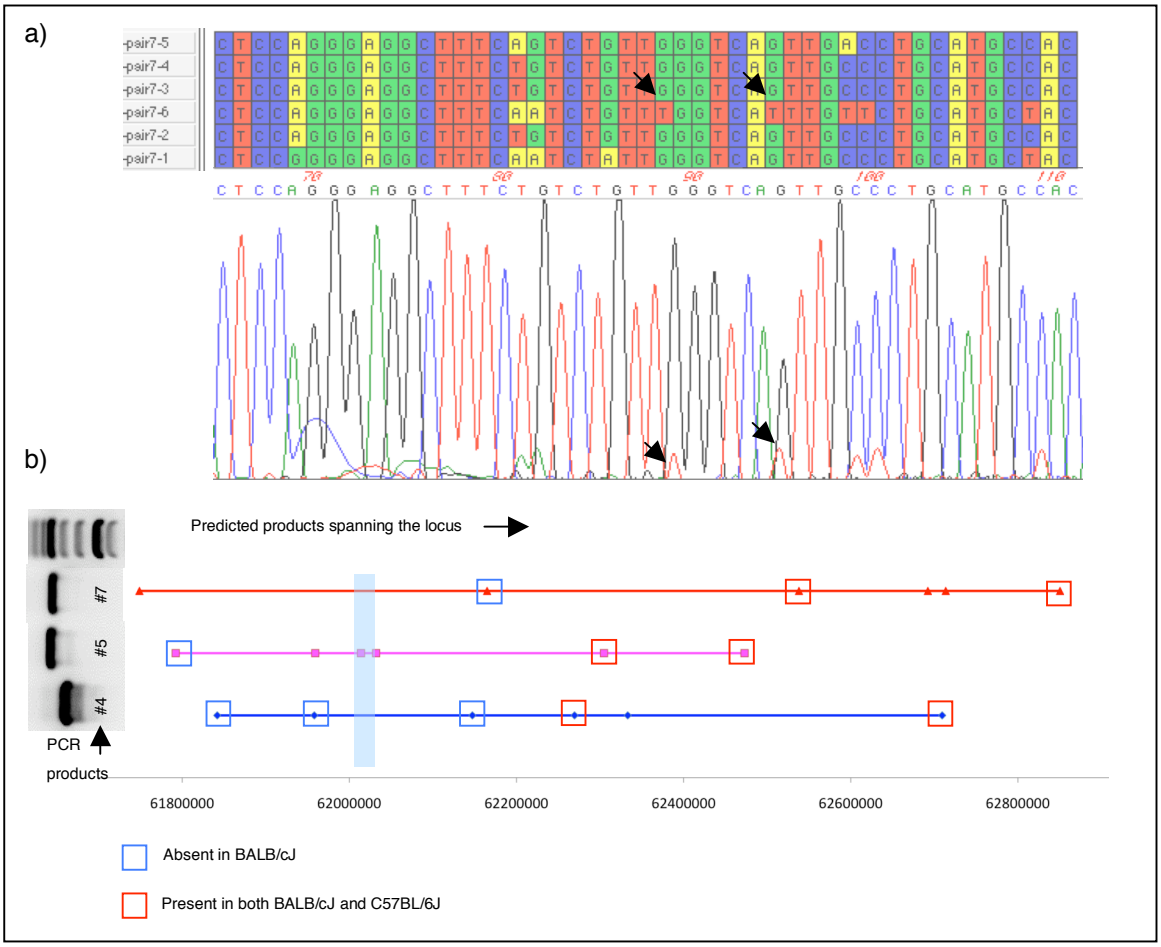


Figure 29. Assessment of missing region in BALB/cJ genome

a) The expected products with polymorphic positions were compared to PCR product sequences. b) Three different pools of PCR products from BALB/cJ and C57BL/6J were sequenced. Missing products in BALB/cJ were compared to C57BL/6J present products. Light blue rectangle represents the start point (in Figure 29).

5.2.4 Identification of *Mdac* within the *KRAB-ZFP* genes

The deleted interval was containing part of a previously annotated KRAB Zinc Finger Proteins (KRAB-ZFPs) cluster. Most of these KRAB-ZFPs correspond to pseudo-genes, some being transcriptionally active based on ESTs. After in depth scrutiny, I have found two distinct clusters of KRAB domain sequences with fourteen of them on the plus strand and twelve on the negative strand of DNA, by looking at the *Mdac* interval for the KRAB-ZFPs using BLAT for KRAB domain and tBLASTn for C2H2 Zinc finger protein sequence. The separation point of two clusters was very close to predicted deleted region in BALB/cJ (Figure 30). Some of these KRAB sequences were associated with few Zinc finger domains with no annotation in the

genome browser other than the eight hypothetical KRAB-ZFPs that were previously annotated (shown in Figure 30 with grey and blue rectangles). I have checked for the potentially coding KRAB-ZFPs sequences and I found two more genes in the structurally variant *Mdac* interval, which were from the transcriptome database (MTR064334.13.1214-2 and 6), with EST coverage in addition to one previously annotated (EG630579) (shown with blue rectangles in Figure 30).

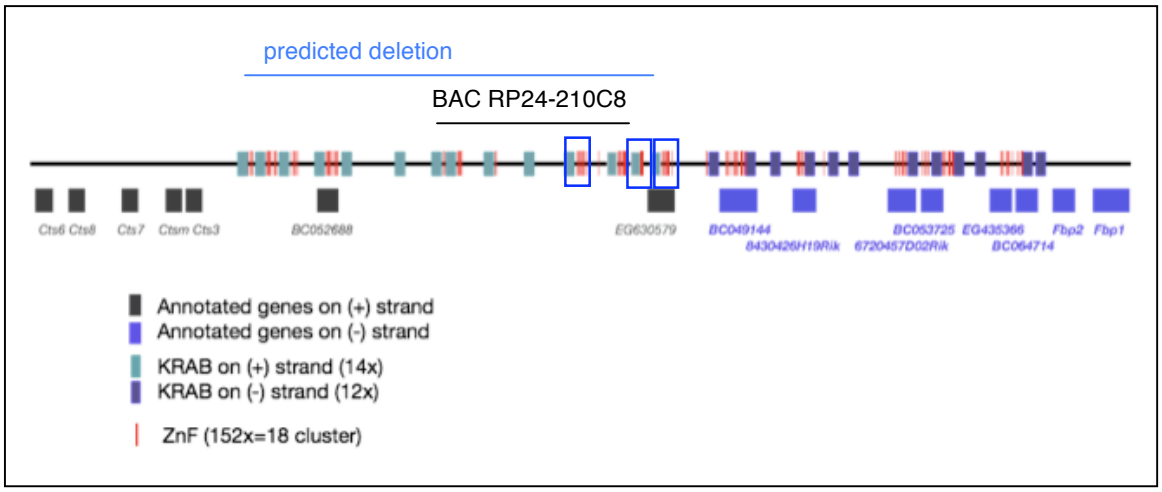


Figure 30. KRAB-ZFPs in *Mdac* interval
Potentially coding KRAB-ZFPs in the deleted region are marked with blue rectangles.

I have cloned the coding sequence for all three genes into lentiviral expression vectors with N-terminal HA and FLAG tags. Together with Katja Langenfeld, we have prepared lentivirus for each construct and she has performed injections into BALB/cJ zygotes. The success rate of transgenesis was not very high with these transgenes, compared to other provirus injected during the same period. So far, I have obtained two transgenic animals for one of the genes (MTR064334.13.1214-2). However, and quite unexpectedly, in both animals, the transgene was containing distinct 3' truncated versions of the gene leading to production of a protein missing some of its zinc fingers (11 left for the first one, 10 for the second). Interruption of the reverse transcription process of virus delivered transgene or recombination between the repeat could lead to a truncation of the highly repeated C2H2 zinc finger domains. However, observing this phenomenon twice is rather surprising and, given the low rate of transgenesis observed with this transgene, it suggests that eventually there is a counter-selection against full-length ones. Indeed, when I tested the transfection of ES cells stably with two of the candidate genes under the control of CAGGS promoter, I observed a high rate of dying cells and the surviving few clones did not express the protein to a detectable

level by Western Blot, even though I was able to detect protein in HEK293T cells stably transfected with the same constructs. This raised the possibility that high-level (or non-properly regulated) expression of KRAB-ZFP could have had toxic effects on cells and embryos. A similar problem has been reported before by Wolf D. et al who also managed only to express a 3'truncated *Zfp809* in 293A cells (Wolf and Goff 2009).

To circumvent the need of establishing stable clones and by-pass embryonic lethality, we tried to see if expression of these genes could have an effect on primary mouse embryonic fibroblast (MEF) prepared from *Dac^{2j}-MusD^{het};Olig2-1-MusD^{het};mdac/mdac* embryos. After transduction with lentivirus expressing candidate KRAB-ZFPs, these cells were checked for methylation changes of 5'LTR-MusD. We did not find any significant change in DNA methylation. The same approach was used on ES cells but when analyzed for transcript levels of the ERVs (MusD/ETns) we observed no change in expression of these. One explanation for these observations could be that transient expression (cells were harvested after 48-72 hours) is not sufficient for the required amount of expression of the genes or repression of MusD elements.

In order to provide a better system that could reproduce the endogenous expression levels of KRAB-ZFPs, I complemented this missing locus by the stable transfection of a BAC (RP24-210C8, Chr13:61996697-62189261) from C57BL/6J (resistant) strain. After retrofitting the BAC with a neomycin selection marker, we chose permissive E14 ES cells (129/Ola strain) following several unsuccessful attempts to use and grow ES cells from the BALB/cJ background (clone obtained from Jackson Laboratories and originated from (Noben-Trauth et al. 1996)). I found a similar deletion in the *Mdac* interval of E14 ES cells, making them a suitable alternative to BALB/cJ. Moreover, the expression levels of MusD/ETns were much higher in E14 ES cells in comparison to the levels in C57BL/6J ES cells, showing a correlation with the *Mdac* status (Figure 31).

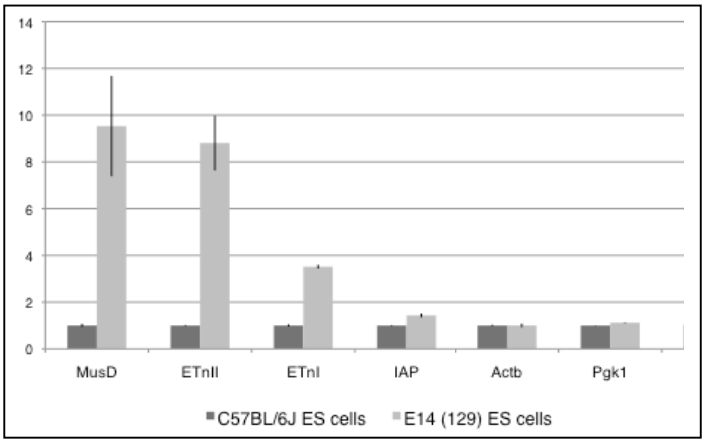


Figure 31. ERV expression in E14 and C57BL/6J ES cells
ERV expression was normalized to Gusb and compared to C57BL/6J cells (Actb and Pgk1 are the other housekeeping control genes)
(Error bars show technical standard deviation)

In total sixteen ES cell colonies were picked and I could amplify fifteen of them. I have tried measuring the expression of the KRAB-ZFP gene included in this BAC in these picked ES cell clones. However, due to sequence similarity to another KRAB-ZFP gene I was not able to detect the specific expression of this gene. Therefore, I used a genomic approach to examine the presence and the integrity of the BAC clone by PCR with six different primer pairs that are amplifying unique products spanning the full length BAC (shown in Figure 32 with blue rectangles), as there might be a variation in the expression in different clones due to position effects. I selected six clones out of fifteen for the preliminary analysis. Clones #3, 6 and #16 were positive for all products, clones #1,8 and 11 lacked product 4 and in addition to this product clone #8 lacked products 10 and 11 as well. I measured MusD/ETn expression levels in all six clones by quantitative real time PCR and normalized it to multiple housekeeping gene controls (Actb, Pgk1 and Gusb). MusD/ETn transcript levels were significantly reduced in clones #1, #3 and #16, whereas other clones showed no significant difference in MusD transcript levels compared to wild type ES cells (Figure 33).

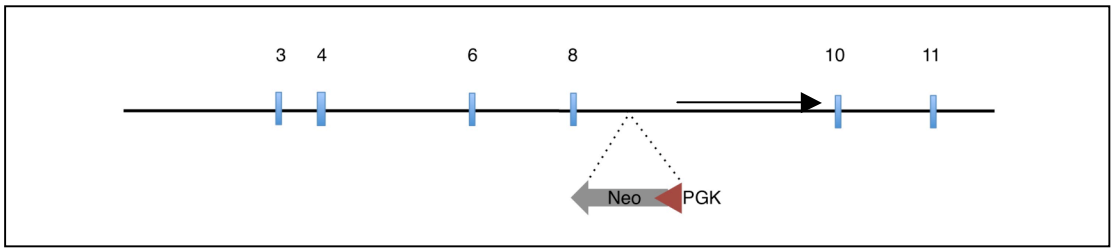


Figure 32. Overview of BAC RP24-210C8

PCR products from the screening are shown with blue rectangles and the PGK driven neomycine cassette is located on the BAC between products 8 and 10. The KRAB-ZFP is shown with an arrow.

Expression levels of various repeat elements are known to go down as ES cells are differentiating. I further analyzed clone #3 by measuring the activity of another ERV, IAP and the expression level of stem cell markers (*Oct4* and *Nanog*). *Oct4* and *Nanog* were expressed at similar levels to those in the parental clone, indicating that this clone is not differentiating. For clone #3, ETnII and ETnI transcript levels were also reduced, maybe as a secondary effect of MusD repression. Importantly, IAP transcripts levels were not changed compared to the parental control (Figure 34), indicating a specific effect of Mdac on MusDs.

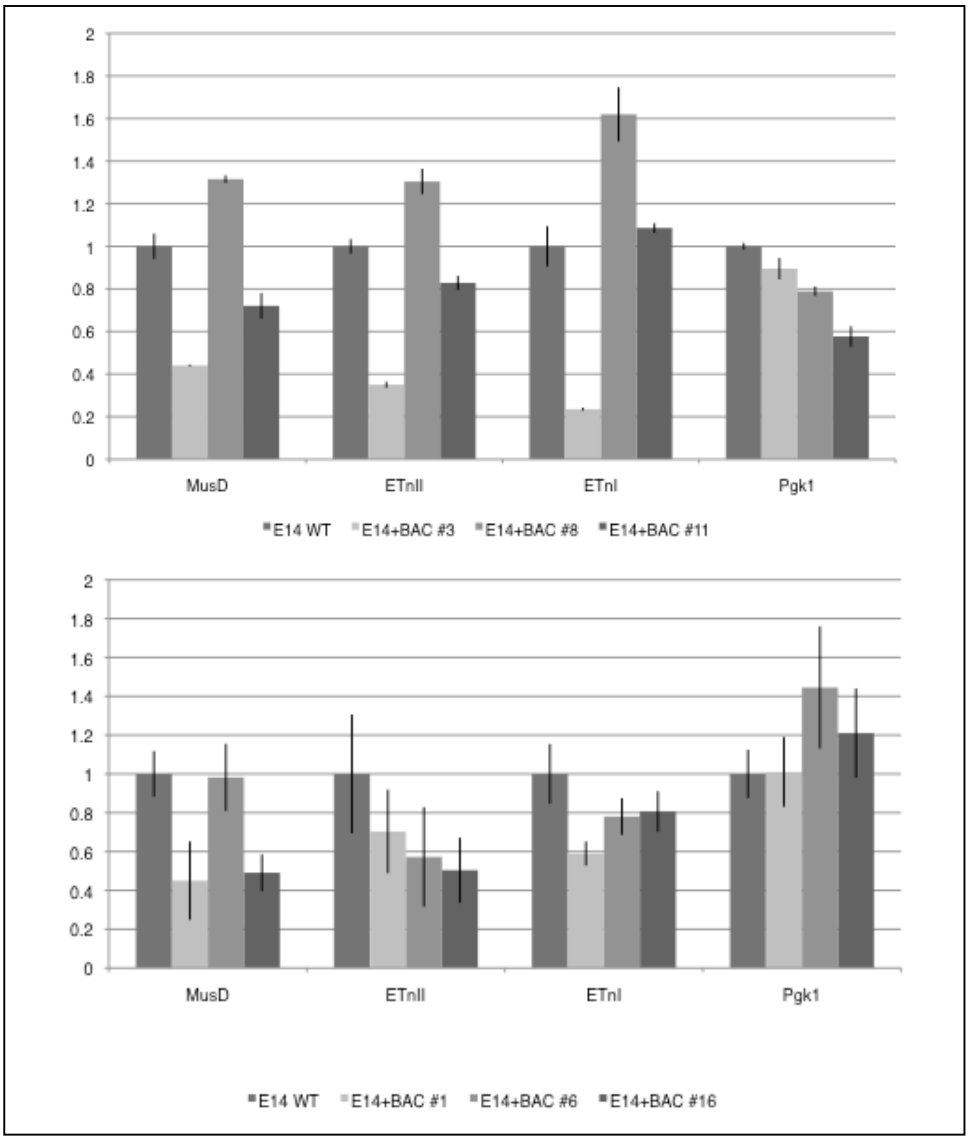


Figure 33. ERV expression levels of BAC complemented ES cells

In two rounds of experiments in total six clones were investigated for MusD, ETnII and ETnI expression levels normalized to *Pgk1* and compared to wild type ES cell clone (E14 WT).

(Error bars show technical standard deviation)

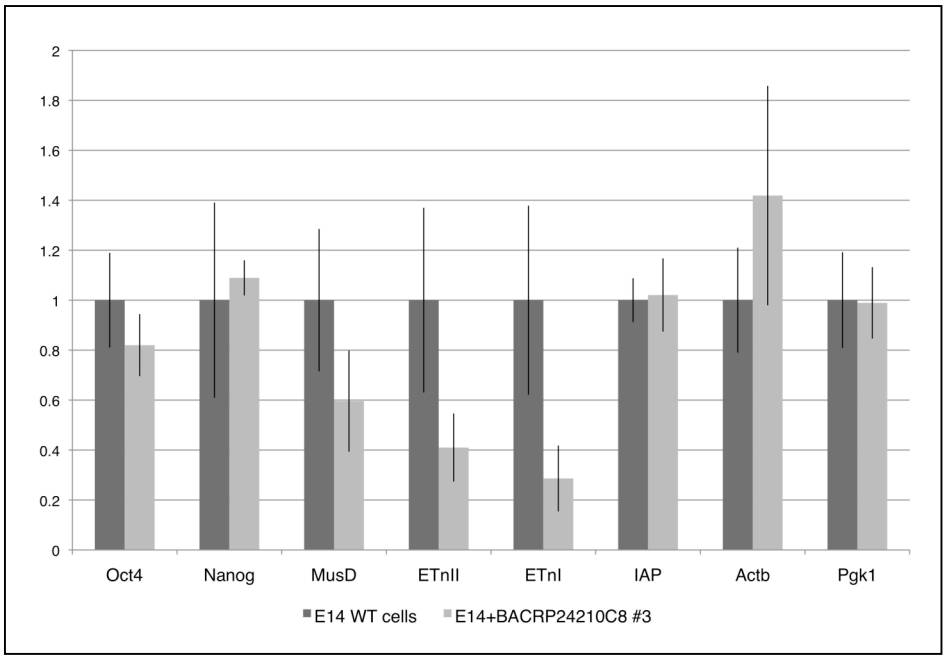


Figure 34. Characterization of BAC complemented clone#3

ERV expression was normalized to Pgk1 and compared to levels in wild type E14 ES cells.

(Error bars represent standard error of the mean between three biological replicates)

In this study, I narrowed down the *Mdac* interval to a repeated region that is deleted in permissive strain. This region contains a cluster of KRAB-ZFP genes. KAP1 functions as a corepressor with KRAB-ZFPs (Friedman et al. 1996), by interacting with HP1 (Ryan et al. 1999) and ESET (Schultz et al. 2002). Furthermore, KAP1 was shown to target the primer binding site (PBS) of Murine Leukemia Virus (MLV) to silence these retroviruses in Embryonic Carcinoma (EC cells) (Wolf and Goff 2007) and ES cells (Wolf et al. 2008) through its interaction with a KRAB-ZFP (*Zfp809*) (Wolf and Goff 2009). In collaboration with Helen Mary Rowe (Trono Group, EPFL), we have showed that induced knock-out of *KAP1* gene in ES cells from C57BL/6J (resistant) strain resulted in an up-regulation of MusDs to seven folds of wild type ES cell levels (Figure 35), suggesting KAP1 is needed for the silencing of MusDs. (Rowe, Jakobsson et al. 2010). These results coming from a cell line derived from a resistant strain are in accord with the lack of the KRAB-ZFP in the permissive stains and these observations collectively point out to a role of KAP1/KRAB-ZFP assisted ERV silencing mechanism. IAPs were more up-regulated in comparison to MusDs in these KAP1 deficient cells. The expression peak and the silencing that follows is normally achieved at an earlier stage for IAPs than MusDs (E3.5 *vs.* E6.5), making the observation of KAP1 dependent control of IAPs easier in an ES cell

based system. Nevertheless, we showed that KAP1 was needed for the silencing of these ERVs.

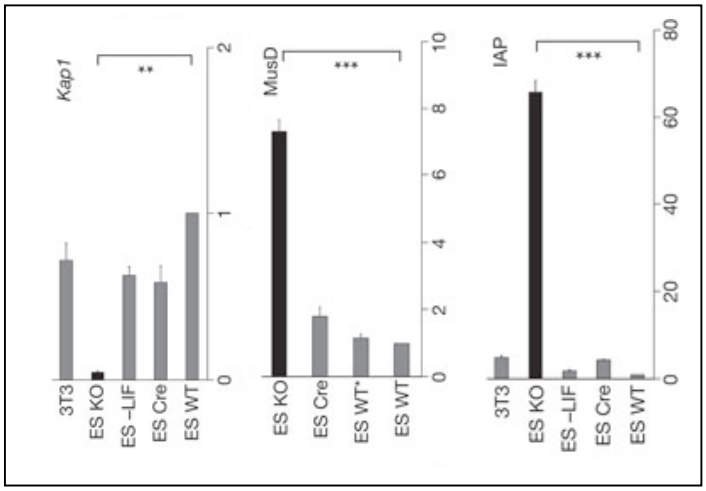


Figure 35. KAP1 controls ERVs, IAP and MusD in ES cells

Induced KAP1 knock-out of ES cells had a 7-fold change for MusD expression.

The repeated nature of KRAB-ZFPs and the difficulty of expressing full-length forms (probably due to toxicity) made it difficult to evaluate the candidate genes for MusD silencing, but complementation of the deleted locus by the BAC provided a strong evidence for the KRAB-ZFP mediated MusD repression. Knocking-down the candidate KRAB-ZFP gene included in the BAC during complementation experiments could be further helpful to confirm its role in MusD silencing. Similarly, KAP1 dependent silencing could be demonstrated by a knock-down of KAP1 in these BAC-complemented cells. Since, MusD expression peaks around sixth to seventh day of embryological development, the ES cell based system may not be a accurate representation of the KAP1/KRAB-ZFP assisted silencing of MusDs. Therefore, BAC complemented cells can be used to generate transgenic animals to measure the MusD silencing at later stages as well as to rescue the *Dactylaplasia* phenotype.

Several host restriction factors that block the retroviral replication have been characterized for different stages of the retroviral life cycle (see the Introduction for details). In this study, I have investigated the polymorphic *Mdac* interval that is responsible for the epigenetic silencing of MusD elements. Mdac acts specifically on MusD elements and restricts their transcription and is very likely to be a KRAB-ZFP, due to the strong correlation of this gene family with the recruitment of silencing complexes to their target loci. Our findings also imply that a group of sequence-specific KRAB-ZFPs act in early embryogenesis and mediate the association of KAP1 with ERV sequences. It is still a mystery which KRAB-ZFP genes are responsible for the silencing of each ERV family. Investigation of the genes that are

expressed around the stages when silencing occurs might help us identify them. The data presented here are very promising for the identification of a transcriptional restriction gene dedicated to a family of endogenous retroviruses and its mechanism of action.

6. Conclusions and Outlook

Our starting point in this study was the ectrodactyly phenotype observed in *SHFM* patients and its phenotypic similarity to the *Dactylaplasia* mice. In fact, *Dactylaplasia* mice have been accepted widely as a model for human *SHFM3* not only because of the similarity between the two phenotypes and but importantly due to the synteny of the loci which the underlying mutations were mapped. However, the mutations that result in abnormal limb formation were radically different between human and mice and it was difficult to predict how a transposon insertion of a relatively small size in mice could lead to the same malformation caused by a duplication of a large genomic region in humans. Since MusD insertions into the *Fgf8* locus are so far the only reported mutagenic insertions of this family (Maksakova, Zhang et al. 2009), our understanding of these elements and their action mechanisms to cause disease has been very limited. Even though *SHFM3* patients were carefully explored for different chromosomal re-arrangements, it has been very difficult to infer a molecular mechanism from the genomic sequence since it is a developmental malformation. Therefore, the basis of the *Dactylaplasia/SHFM3* phenotypes was not sufficiently explained.

6.1 SHFM3: multi-genic and multi-enhancer explanation to disease phenotype

In this thesis, I demonstrated that in an engineered chromosomal duplication between the *Fgf8* gene and *Lbx1::GFP* reporter, transgenic embryos displayed the reporter gene expression in *Fgf8* expression domains. Thus, we showed that chromosomal re-arrangements could cause ectopic gene expression in place of endogenous gene expression domains due to the re-allocation of existing regulatory modules. We proposed that the re-direction of regulatory sequences to other gene(s) in the locus as our model for the ectrodactyly phenotype in both *SHFM3* patients and *Dactylaplasia* mutants. We have tried expressing two genes (*Lbx1* and *βTrc)* from the locus in the AER in an effort to phenocopy the limb malformation observed in *Dactylaplasia* mice. *Lbx1* expression in the AER caused a duplicated or enlarged thumb that is also observed in several *SHFM3* patients, but we did not observe the cleft limb phenotype. On the other hand, *βTrc* expression in the AER did not cause a limb phenotype in transgenic animals. The onset of mutagenic gene expression leading to the disease phenotype might be crucial and the expression stage of

candidate genes in the AER might not have been the same as the disease situation in our experimental set-up. Alternatively, the cleft limb phenotype might simply be caused by the ectopic expression of another gene or by multiple genes acting cooperatively.

So far different diseases were associated with multiple structural variations that disrupt multiple genes (such as in schizophrenia (Walsh et al. 2008)), however the spectrum of the SHFM3 phenotype is caused by structural variations in a single locus. The extent of the duplicated region has been related to the variability of the limb phenotype in *SHFM3* (such as preaxial polydactyly) and to additional syndromes (such as hearing disorders and renal hypoplasia). For instance, if the duplication break point is found between the *LBX1* and *βTRC*, patients have proximally placed thumbs and/or triphalangeal thumbs (TPT) or preaxial polydactyly (Elliott, Reed et al. 2005), (Everman, Morgan et al. 2006). Based on the results of our forced *Lbx1* expression in the AER, it now seems likely that this duplication might be altering the regulatory architecture of the locus in such a way that *LBX1* is ectopically expressed in the AER of affected patients. Consistently, patients with the duplication break point closer to the *LBX1* gene have ectrodactyly but not the polydactylous phenotype, indicating the difference in regulatory re-allocations for different break points. Furthermore, patients that have a similar size of duplication to previously reported cases with a different break point closer to *FGF8* gene have additional syndromes to the limb defects (Dimitrov, de Ravel et al. 2010). These reports indicate a multi-genic and multi-regulatory contribution to the disease phenotype that is caused by structural variation in a single locus. When our chromosomal engineering and transgenic approaches are taken into consideration, we have provided an explanation for the regulatory re-arrangements and their consequences on gene expression. Therefore, the disease phenotype is better understood than just having the information on the break points of the duplications in the mapped locus. As a result, *SHFM3* has been a case example that warns us about the way of interpreting structural variations and their impact on human phenotypes.

6.2 MusD's impact on gene expression

In *Dactylaplasia* mutants, MusD insertions bring a regulatory reshuffling activity to the locus that is similar to the duplications in *SHFM3* patients. As also shown for the *Olig2-1* locus, MusD's capacity to act on gene regulation seems to be a general feature of this element. However, MusD's specific impact on gene

expression depends on the locus of insertion and to the regulatory potential therein. This regulatory re-shuffling induced by MusDs seems to correlate with their insulator activity. Insulators may function as transcriptional decoys for enhancer sequences, interrupting chromatin loops that bring enhancers and promoters together. In concordance with this, many insulators include promoter-proximal elements (listed in (Geyer and Clark 2002)) that may prevent enhancers from interacting with promoters (Ohtsuki and Levine 1998). Furthermore, in some circumstances, insulator sequences are not transcriptionally inert, such as the gypsy insulator that can act as a transcriptional activator (Wei and Brennan 2001). In fact, insulator sequences were also considered to have evolved from promoter sequences, supporting this transcriptional decoy model (Geyer and Clark 2002). Evidence to this model comes from the α-globin gene cluster, in which a gain-of-function regulatory SNP can potentially create a new binding site for the GATA-1 transcription factor, leading to the preferential interaction of this heterologous promoter with the upstream elements that results in a competition between the endogenous promoters of α-globin genes (De Gobbi et al. 2006). Thus, this SNP was insulating the regulatory elements from their cognate promoters. Another functional insulator in relation to promoter sequences was shown for stalled *Hox* promoters in *Drosophila* (Chopra et al. 2009). In this study, stalled RNA polymerase II (PolII) and its interaction with insulator sequences through NELF and DSIF protein complexes was proposed to result in an enhancer blocking activity via the formation of high-order loops. These loop domains would prevent aberrant interactions and favor proper gene regulation. A different example of transcription-triggered insulation was shown in the *Growth Hormone* (*GH*) gene locus (Lunyak, Prefontaine et al. 2007). In *GH* locus the transcriptional regulation of a SINE B2 repeat determined by the state of histone modifications controls the regulatory potential in the locus. These transcriptional activities driven by PolII or by PolIII in opposite orientations correlated with the epigenetic switch as well as the *GH* expression. These findings indicated an insulator function determined by the changes in SINE B2 expression at embryological stages when the gene expression was not needed. Similar to the above listed examples, MusD insertions in the *Fgf8* locus seem to be functioning as insulator elements and introduce a competition for the endogenous enhancer-promoter interactions. This competition manifests itselfs as the expression of MusD in *Fgf8* domains. However, the enhancer blocking function may be explained by additional properties of MusD elements. In *ex-vivo* enhancer blocking assays, the insulator activity seemed to be through a few

regions in the retroviral gene sequences, and did not support the idea that transcriptional initiation from the LTR is the cause of MusD's insulator activity. Nevertheless, as a consequence of MusD-5'LTR accessibility, the transposon is transcribed and thus insulator sequences are opened to the possibility of interacting with new nuclear factors. Bisulfite sequencing showed that there is an extension of methylated cytosine positions toward the inside of the MusD elements. This could be masking the insulator fragments and preventing the binding of a factor that can be necessary for the insulator function. Furthermore, *ex-vivo* experiments also showed that the enhancer-blocking activity of MusD sequences potentially re-allocates the enhancers within the locus. Thus, MusD fragments that displayed insulator function in *ex-vivo* assays should be further examined for the binding of novel proteins that could explain this redirection of the enhancers to other gene(s) in the locus.

The first model of insulator function is mainly based on the transcriptional activity of the insulator sequence as described above. There is another model which proposes that insulators influence the physical organization of chromosomes in order to protect gene expression in a locus from the effect of surrounding chromatin. This model suggests that insulators primarily participate in the formation of higher-order chromatin structures and transcription is subsequently affected by these structures (Gerasimova and Corces 1998), (Scott et al. 1999). Our data supports the first model since the insulator activity of MusD element is related to its transcriptional status. However, we do not know if and how the insertion of a MusD element can affect the architecture of the surrounding chromatin and further experiments are required to determine if MusD's insulator function depends on structural alteration of the inserted locus.

6.3 Novel ERV silencing mechanisms

In the case of *Dactylaplasia* mice, manifestation of the limb phenotype upon MusD transposon insertions depends on an unlinked modifier locus. Previously, transposons' role in phenotypic variation has been suggested to be due to the stochasticity in their epigenetic silencing (Reiss and Mager 2007) and hence the presence of a modifier locus in the case of *Dactylaplasia* provided a striking example for the modulating effect of the genetic background. In contrast to the IAP insertions in *Axin*^{Fu} and *Agouti* loci, modifier of *Dactylaplasia* (*Mdac*) does not act on MusDs to make them function as cryptic promoters to cause differential expression of the nearby genes or cause a stochastic silencing of the MusD elements that would lead

to the spreading of repressive epigenetic marks over the locus of insertion. The presence of the dominant allele of *Mdac* suppresses the *Dactylaplasia* phenotype and leads to the epigenetic silencing of MusD elements. However, the majority of MusD elements remain silenced in the presence of the the Mdac allele from the permissive strain, suggesting this mechanism could be used as a back up to ensure the silencing of specific elements that are outside heterochromatin regions or nearby genes. Therefore, the major defense mechanism seems to be the prurifying selection due to deleterious effects of MusD's impact on endogenous gene regulation. When I investigated the extent of cytosine methylation of ETnII elements that are degenerate derivatives of MusD transposons I observed no differences in cytosine methylation among different strains, even though these elements share a high sequence similarity of their LTRs and a short region upstream of viral genes. Therefore, they are presumably controlled by different silencing mechanisms and Mdac-dependent silencing is specific to MusD elements. The specificity of repression to the MusD elements suggests an epigenetic spreading of silencing over the 5' LTR that follows the recognition of MusD specific sequences encoded by *Mdac*. The silencing of the 5'LTRs, which serve as promoters for retrotransposons, seems to be enough for the repression of MusD activity, as this spreading does not reach to the 3' of the MusD.

KAP1 seems to function as a general repressor of retroviral activities, as shown for MLV (Wolf and Goff 2007), IAP and MusD (Rowe, Jakobsson et al. 2010). The specificity of this repression to a particular family of transposons could be achieved via the KRAB-ZFPs. According to the phylogenetic analyses KRAB ZFP genes appeared first in the tetrapods, and have since been under strong positive selection (Emerson and Thomas 2009) and subject to rapid expansion (Consortium et al. 2002). The reason for the expansion of this family in tetrapods is currently unknown. The KRAB-ZFP cluster in the *Mdac* locus seems to be important for the epigenetic control of MusD elements. Therefore, it is also likely to find different clusters of KRAB-ZFP that evolved by duplications in time, each specialized to silence a specific family of transposable element. Indeed, a recent study proposed a correlation across vertebrate genomes between the number of LTR-transposons and the number of host's tandem zinc finger genes (Thomas and Schneider 2011). MusD sequences are believed to have entered in the germ line of a *Muridae*, as these elements are found only in the *Mus* genus (i.e., 5 Myr) (Mager and Freeman 2000), (Ribet, Harper et al. 2007). The KRAB-ZFP cluster in the *Mdac* interval seems to be unique to the mouse genome. When the locus was explored for conservation, none

was found between the mouse and the rat genomes except for the surrounding metabolic genes and the *Cathepsin* cluster. This raises the possibility of a co-evolution of MusDs and as a host response, KRAB zinc-finger proteins, which can repress MusD transcription and prevent detrimental mutations that can be caused by these retrotransposons. There seems to be an expansion of zinc-finger domains that specifies the DNA binding, whereas KRAB domains remain conserved (Thomas and Schneider 2011). This indicates a modular zinc-finger part of the protein, which can provide opportunities for targeting specific families of retrotransposons through a global repressor, KAP1, via interactions with the more conserved KRAB domain. This modularity in the substrate recognition module together with an invariable effector domain is reminiscent of the well-known antibody-antigen phenomena and the possibility of creating new binding modules through deletion/duplication of the zinc-finger domains suggests an analogous mechanism to generating immunoglobulin diversity. This idea supports the rapid expansion model for a key role in immune system, as KRAB-ZFPs seem to prevent viral element invasion of the genomes.

This arms race between the host genome and transposon activity is an on-going process, thus MusD elements that manage to escape the KRAB-ZFP interactions would be preserved and amplified in the genome. In this respect, it is possible to propose that ETnII elements sacrificed their genes that code for viral proteins required for transposition and became MusD-dependent, but in turn managed to escape from silencing mechanisms when compared to MusD elements.

MusD and its silencing elements were noticeable owing to the presence of a modifier locus that is variable in different mouse strains. In order to investigate such KRAB-ZFP genes dedicated to one family of transposons, one could investigate the appearance of a transposon family and its relation to the divergence of a KRAB-ZFP cluster in an evolutionary time scale. Subsequently, the mutagenesis of such candidates could be used to test this proposition of co-evolution and help us understand why this family of genes is expanding in mammalian genomes.

6.4 The working model

In this thesis, I have presented a multigenic model of *Dactylaplasia* phenotype in relation to *SHFM3* and its modifier locus. The data presented here suggest a model with a gene that is specifically responsible for the epigenetic silencing of MusD elements. In the absence of this repressor protein, MusD acts as an insulator sequence by hijacking nearby enhancers and introducing new enhancer-promoter

associations. Binding of a novel insulator protein to MusD sequences could drive this re-direction of regulatory elements. Similarly, the duplication found in *SHFM3* patients changes the position of enhancers relative to their target gene, possibly enabling them act on other gene(s). Ultimately, when the enhancers activate other gene(s) from the locus, the AER terminates prematurely and leads to ectrodactyly (see Figure 36 for the model).

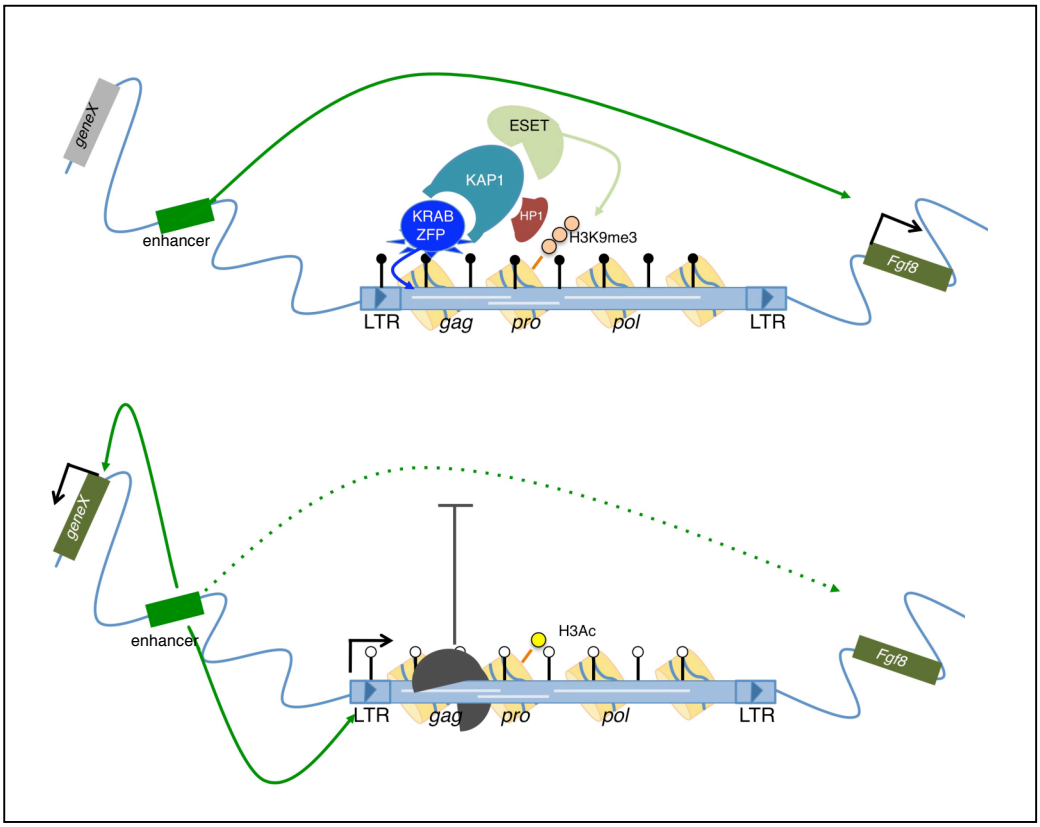


Figure 36. The working model for the *Dactylaplasia* mutation
MusD as an epigenetically controlled mobile insulator element.
KRAB-ZFP specific binding to the transposon sequences recruits repressor proteins and induces the DNA methylation. In the absence of this protein, transcriptional activity from the transposon blocks or titrates endogenous enhancers, re-shuffling the regulatory interactions.

7. Literature

Agrawal, A., Q. M. Eastman, et al. (1998). "Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system." Nature **394**(6695): 744-51.

Akiyama, H., M. C. Chaboissier, et al. (2002). "The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6." Genes & Development **16**(21): 2813-28.

Allen, E., S. Horvath, et al. (2003). "High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes." Proc Natl Acad Sci USA **100**(17): 9940-5.

Aravin, A. A., G. J. Hannon, et al. (2007). "The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race." Science **318**(5851): 761-4.

Aravin, A. A., R. Sachidanandam, et al. (2007). "Developmentally regulated piRNA clusters implicate MILI in transposon control." Science **316**(5825): 744-7.

Ayala, F. J. and M. Coluzzi (2005). "Chromosome speciation: humans, Drosophila, and mosquitoes." Proc Natl Acad Sci USA **102 Suppl 1**: 6535-42.

Bailey, J. A., D. M. Church, et al. (2004). "Analysis of Segmental Duplications and Genome Assembly in the Mouse." Genome Res.

Bailey, J. A., G. Liu, et al. (2003). "An Alu transposition model for the origin and expansion of human segmental duplications." Am J Hum Genet **73**(4): 823-34.

Barrow, J. R., K. R. Thomas, et al. (2003). "Ectodermal Wnt3/beta-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge." Genes & Development **17**(3): 394-409.

Baust, C., L. Gagnier, et al. (2003). "Structure and expression of mobile ETnII retroelements and their coding-competent MusD relatives in the mouse." J Virol **77**(21): 11448-58.

Beisel, C. and R. Paro (2011). "Silencing chromatin: comparing modes and mechanisms." Nat Rev Genet **12**(2): 123-35.

Bejerano, G., C. B. Lowe, et al. (2006). "A distal enhancer and an ultraconserved exon are derived from a novel retroposon." Nature **441**(7089): 87-90.

Bejerano, G., M. Pheasant, et al. (2004). "Ultraconserved elements in the human genome." Science **304**(5675): 1321-5.

Bell, A. C., A. G. West, et al. (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." Cell **98**(3): 387-96.

Bestor, T. H. (1999). "Sex brings transposons and genomes into conflict." Genetica **107**(1-3): 289-95.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 60-3.

Böhne, A., F. Brunet, et al. (2008). "Transposable elements as drivers of genomic and biological diversity in vertebrates." Chromosome Res **16**(1): 203-15.

Bourque, G., B. Leong, et al. (2008). "Evolution of the mammalian transcription factor binding repertoire via transposable elements." Genome Res **18**(11): 1752-62.

Britten, R. J. and E. H. Davidson (1969). "Gene regulation for higher cells: a theory." Science **165**(891): 349-57.

Britten, R. J. and E. H. Davidson (1971). "Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty." Q Rev Biol **46**(2): 111-38.

Brohmann, H., K. Jagla, et al. (2000). "The role of Lbx1 in migration of muscle precursor cells." Development **127**(2): 437-45.

Brower-Toland, B., S. D. Findley, et al. (2007). "Drosophila PIWI associates with chromatin and interacts directly with HP1a." Genes & Development **21**(18): 2300-11.

Butler, J. E. and J. T. Kadonaga (2001). "Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs." Genes & Development **15**(19): 2515-9.

Calhoun, V. C., A. Stathopoulos, et al. (2002). "Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex." Proc Natl Acad Sci USA **99**(14): 9243-7.

Carninci, P., A. Sandelin, et al. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." Nat Genet **38**(6): 626-35.

Carroll, S. B. (2008). "Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution." Cell **134**(1): 25-36.

Celli, J., P. Duijf, et al. (1999). "Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome." Cell **99**(2): 143-53.

Chai, C. K. (1981). "Dactylaplasia in mice a two-locus model for development anomalies." J Hered **72**(4): 234-7.

Charlesworth, B., P. Sniegowski, et al. (1994). "The evolutionary dynamics of repetitive DNA in eukaryotes." Nature **371**(6494): 215-20.

Chen, X., H. Xu, et al. (2008). "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells." Cell **133**(6): 1106-17.

Chopra, V. S., J. Cande, et al. (2009). "Stalled Hox promoters as chromosomal boundaries." Genes & Development.

Chung, J. H., M. Whiteley, et al. (1993). "A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila." Cell **74**(3): 505-14.

Cohen, D., L. Davidow, et al. (2007). "The DXPas34 repeat regulates random and imprinted X inactivation." Dev Cell **12**(1): 57-71.

Conley, A. B., W. J. Miller, et al. (2008). "Human cis natural antisense transcripts initiated by transposable elements." Trends Genet **24**(2): 53-6.

Consortium, G. P. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-73.

Consortium, M. G. S., R. H. Waterston, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Corces, V. G. and P. K. Geyer (1991). "Interactions of retrotransposons with the host genome: the case of the gypsy element of Drosophila." Trends Genet **7**(3): 86-90.

Crackower, M. A., J. Motoyama, et al. (1998). "Defect in the maintenance of the apical ectodermal ridge in the Dactylaplasia mouse." Dev Biol **201**(1): 78-89.

Dathe, K., K. Kjaer, et al. (2009). "Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2." Am J Hum Genet **84**(4): 483-92.

De Gobbi, M., V. Viprakasit, et al. (2006). "A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter." Science **312**(5777): 1215-7.

De La Fuente, R., C. Baumann, et al. (2006). "Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells." Nat Cell Biol **8**(12): 1448-54.

De Moerlooze, L., B. Spencer-Dene, et al. (2000). "An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal-

epithelial signalling during mouse organogenesis." Development **127**(3): 483-92.
de Mollerat, X. J., F. Gurrieri, et al. (2003). "A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24." Human Molecular Genetics **12**(16): 1959-71.
Dekker, J., K. Rippe, et al. (2002). "Capturing chromosome conformation." Science **295**(5558): 1306-11.
Dennis, K., T. Fan, et al. (2001). "Lsh, a member of the SNF2 family, is required for genome-wide methylation." Genes & Development **15**(22): 2940-4.
Dewannieux, M. and T. Heidmann (2005). "L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells." J Mol Biol **349**(2): 241-7.
Di-Poï, N., J. I. Montoya-Burgos, et al. (2009). "Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard." Genome Res **19**(4): 602-10.
Di-Poï, N., J. I. Montoya-Burgos, et al. (2010). "Changes in Hox genes' structure and function during the evolution of the squamate body plan." Nature **464**(7285): 99-103.
Dimitrov, B. I., T. de Ravel, et al. (2010). "Distal limb deficiencies, micrognathia syndrome, and syndromic forms of split hand foot malformation (SHFM) are caused by chromosome 10q genomic rearrangements." Journal of Medical Genetics **47**(2): 103-11.
Dong, K. B., I. A. Maksakova, et al. (2008). "DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity." EMBO J.
Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**(5757): 601-3.
Duijf, P. (2003). "Pathogenesis of split-hand/split-foot malformation." Human Molecular Genetics **12**(90001): 51R-60.
Elliott, A. M., M. H. Reed, et al. (2005). "Discrepancies in upper and lower limb patterning in split hand foot malformation." Clinical Genetics **68**(5): 408-23.
Emerson, R. O. and J. H. Thomas (2009). "Adaptive evolution in zinc finger transcription factors." PLoS Genet **5**(1): e1000325.
Esnault, C., O. Heidmann, et al. (2005). "APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses." Nature **433**(7024): 430-3.
Everman, D. B., C. Morgan, et al. (2006). "Frequency of genomic rearrangements involving the SHFM3 locus at chromosome 10q24 in syndromic and non-syndromic split-hand/foot malformation." Am J Med Genet A **140**(13): 1375-83.
Fernandez-Teran, M. and M. A. Ros (2008). "The Apical Ectodermal Ridge: morphological aspects and signaling pathways." Int J Dev Biol **52**(7): 857-71.
Feschotte, C. (2008). "Transposable elements and the evolution of regulatory networks." Nat Rev Genet **9**(5): 397-405.
Ficz, G., M. Branco, et al. (2011). "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." Nature.
Frankel, N., G. Davis, et al. (2010). "Phenotypic robustness conferred by apparently redundant transcriptional enhancers." Nature.
Friedli, M., I. Barde, et al. (2010). "A systematic enhancer screen using lentivector transgenesis identifies conserved and non-conserved functional elements at the Olig1 and Olig2 locus." PLoS ONE **5**(12): e15741.
Friedli, M., S. Nikolaev, et al. (2008). "Characterization of mouse Dactylaplasia mutations: a model for human ectrodactyly SHFM3." Mamm Genome **19**(4): 272-8.

Friedman, J., W. Fredericks, et al. (1996). "KAP-1, a novel corepressor for the highly conserved KRAB repression domain." Genes & Development **10**(16): 2067-78.

Gaszner, M. and G. Felsenfeld (2006). "Insulators: exploiting transcriptional and epigenetic mechanisms." Nat Rev Genet **7**(9): 703-13.

Gentles, A. J., M. J. Wakefield, et al. (2007). "Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica." Genome Res **17**(7): 992-1004.

Gerasimova, T. I. and V. G. Corces (1998). "Polycomb and trithorax group proteins mediate the function of a chromatin insulator." Cell **92**(4): 511-21.

Geyer, P. K. and I. Clark (2002). "Protecting against promiscuity: the regulatory role of insulators." Cell Mol Life Sci **59**(12): 2112-27.

Gilthorpe, J., M. Vandromme, et al. (2002). "Spatially specific expression of Hoxb4 is dependent on the ubiquitous transcription factor NFY." Development **129**(16): 3887-99.

Girard, A. and G. J. Hannon (2008). "Conserved themes in small-RNA-mediated transposon control." Trends Cell Biol **18**(3): 136-48.

Gong, Q. and A. Dean (1993). "Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NF-E2." Mol Cell Biol **13**(2): 911-7.

Goodier, J. and H. Kazazianjr (2008). "Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites." Cell **135**(1): 23-35.

Goodier, J. L., E. M. Ostertag, et al. (2000). "Transduction of 3'-flanking sequences is common in L1 retrotransposition." Human Molecular Genetics **9**(4): 653-7.

Guo, Q., C. Loomis, et al. (2003). "Fate map of mouse ventral limb ectoderm and the apical ectodermal ridge." Developmental Biology **264**(1): 166-78.

Hart, M., J. P. Concordet, et al. (1999). "The F-box protein beta-TrCP associates with phosphorylated beta-catenin and regulates its activity in the cell." Curr Biol **9**(4): 207-10.

Hérault, Y., M. Rassoulzadegan, et al. (1998). "Engineering chromosomes in mice through targeted meiotic recombination (TAMERE)." Nat Genet **20**(4): 381-4.

Hickey, D. A. (1982). "Selfish DNA: a sexually-transmitted nuclear parasite." Genetics **101**(3-4): 519-31.

Hong, J. W., D. A. Hendrix, et al. (2008). "Shadow enhancers as a source of evolutionary novelty." Science **321**(5894): 1314.

Huang, J., T. Fan, et al. (2004). "Lsh, an epigenetic guardian of repetitive elements." Nucleic Acids Res **32**(17): 5019-28.

Hurst, G. D. and J. H. Werren (2001). "The role of selfish genetic elements in eukaryotic evolution." Nat Rev Genet **2**(8): 597-606.

Ianakiev, P., M. W. Kilpatrick, et al. (2000). "Split-hand/split-foot malformation is caused by mutations in the p63 gene on 3q27." Am J Hum Genet **67**(1): 59-66.

Ikeda, T., K. H. Abd El Galil, et al. (2011). "Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons." Nucleic Acids Res.

Inoue, F., M. S. Parvin, et al. (2008). "Transcription of fgf8 is regulated by activating and repressive cis-elements at the midbrain-hindbrain boundary in zebrafish embryos." Developmental Biology **316**(2): 471-86.

Ito, S., A. D'alessio, et al. (2010). "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification." Nature **466**(7310): 1129-1133.

J. Courey, A. l. b. e. r. t. (2008). "Mechanisms in transcriptional regulation." 231.

Jack, J., D. Dorsett, et al. (1991). "Expression of the cut locus in the Drosophila wing margin is required for cell type specification and is regulated by a distant enhancer." Development **113**(3): 735-47.

Jagla, K., P. Dollé, et al. (1995). "Mouse Lbx1 and human LBX1 define a novel mammalian homeobox gene family related to the Drosophila lady bird genes." Mech Dev **53**(3): 345-56.

Jhunjhunwala, S., M. C. van Zelm, et al. (2008). "The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions." Cell **133**(2): 265-79.

Johnson, K. R., P. W. Lane, et al. (1995). "Mapping the mouse dactylaplasia mutation, Dac, and a gene that controls its expression, mdac." Genomics **29**(2): 457-64.

Jones, K. A., J. T. Kadonaga, et al. (1987). "A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication." Cell **48**(1): 79-89.

Jones, P. A. and G. Liang (2009). "Rethinking how DNA methylation patterns are maintained." Nat Rev Genet **10**(11): 805-11.

Jordan, I. K., I. B. Rogozin, et al. (2003). "Origin of a substantial fraction of human regulatory sequences from transposable elements." Trends Genet **19**(2): 68-72.

Jurka, J., V. V. Kapitonov, et al. (2007). "Repetitive sequences in complex genomes: structure and evolution." Annu Rev Genomics Hum Genet **8**: 241-59.

Juven-Gershon, T., J. Y. Hsu, et al. (2008). "Caudal, a key developmental regulator, is a DPE-specific transcriptional factor." Genes & Development **22**(20): 2823-30.

Juven-Gershon, T., J. Y. Hsu, et al. (2008). "The RNA polymerase II core promoter - the gateway to transcription." Curr Opin Cell Biol **20**(3): 253-9.

Kadonaga, J. T. and R. Tjian (1986). "Affinity purification of sequence-specific DNA binding proteins." Proc Natl Acad Sci USA **83**(16): 5889-93.

Kano, H., H. Kurahashi, et al. (2007). "Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype." Proc Natl Acad Sci USA **104**(48): 19034-9.

Kapitonov, V. V. and J. Jurka (2008). "A universal classification of eukaryotic transposable elements implemented in Repbase." Nat Rev Genet **9**(5): 411-2; author reply 414.

Kato, Y., M. Kaneda, et al. (2007). "Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse." Human Molecular Genetics **16**(19): 2272-80.

Kawakami, Y., J. Capdevila, et al. (2001). "WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo." Cell **104**(6): 891-900.

Keller, S. A., J. M. Jones, et al. (1994). "Kidney and retinal defects (Krd), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the Pax2 locus." Genomics **23**(2): 309-20.

Kellum, R. and P. Schedl (1992). "A group of scs elements function as domain boundaries in an enhancer-blocking assay." Mol Cell Biol **12**(5): 2424-31.

Keshet, E., R. Schiff, et al. (1991). "Mouse retrotransposons: a cellular reservoir of long terminal repeat (LTR) elements with diverse transcriptional specificities." Adv Cancer Res **56**: 215-51.

Ketting, R. F., T. H. Haverkamp, et al. (1999). "Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD." Cell **99**(2): 133-41.

Kidwell, M. G. and D. R. Lisch (2000). "Transposable elements and host genome evolution." Trends Ecol Evol (Amst) **15**(3): 95-99.

Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." Cell **128**(6): 1231-45.

King, M. C. and A. C. Wilson (1975). "Evolution at two levels in humans and chimpanzees." Science **188**(4184): 107-16.

Kleinjan, D. A. and L. A. Lettice (2008). "Long-range gene control and genetic disease." Adv Genet **61**: 339-88.

Kleinjan, D. A. and V. van Heyningen (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." Am J Hum Genet **76**(1): 8-32.

Klopocki, E., C. Ott, et al. (2008). "A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome." Journal of Medical Genetics **45**(6): 370-5.

Kmita, M., T. Kondo, et al. (2000). "Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing." Nat Genet **26**(4): 451-4.

Kubo, S., M. C. Seleme, et al. (2006). "L1 retrotransposition in nondividing and primary human somatic cells." Proc Natl Acad Sci USA **103**(21): 8036-41.

Kunarso, G., N. Chia, et al. (2010). "Transposable elements have rewired the core regulatory network of human embryonic stem cells." Nat Genet **42**(7): 631-4.

Kurth, I., E. Klopocki, et al. (2009). "Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia." Nat Genet **41**(8): 862-3.

Lallemand, Y., V. Bensoussan, et al. (2009). "Msx genes are important apoptosis effectors downstream of the Shh/Gli3 pathway in the limb." Developmental Biology **331**(2): 189-98.

Lallemand, Y., M. A. Nicola, et al. (2005). "Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb development." Development **132**(13): 3003-14.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Leeb, M., D. Pasini, et al. (2010). "Polycomb complexes act redundantly to repress genomic repeats and genes." Genes & Development **24**(3): 265-76.

Lettice, L. A., S. J. Heaney, et al. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Human Molecular Genetics **12**(14): 1725-35.

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-51.

Lewandoski, M., X. Sun, et al. (2000). "Fgf8 signalling from the AER is essential for normal limb development." Nat Genet **26**(4): 460-3.

Lieberman-Aiden, E., N. Van Berkum, et al. (2009). "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." Science **326**(5950): 289-293.

Lippman, Z., A. V. Gendrel, et al. (2004). "Role of transposable elements in heterochromatin and epigenetic control." Nature **430**(6998): 471-6.

Liu, K., Y. F. Wang, et al. (2003). "Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells in vivo." Mol Cell Biol **23**(8): 2709-19.

Liu, Y. H., L. Ma, et al. (1994). "Regulation of the Msx2 homeobox gene during mouse embryogenesis: a transgene with 439 bp of 5' flanking sequence is expressed exclusively in the apical ectodermal ridge of the developing limb." Mech Dev **48**(3): 187-97.

Lo Iacono, N., S. Mantero, et al. (2008). "Regulation of Dlx5 and Dlx6 gene expression by p63 is involved in EEC and SHFM congenital limb defects." Development **135**(7): 1377-88.

Loebel, D. A., B. Tsoi, et al. (2004). "Restricted expression of ETn-related sequences during post-implantation mouse development." Gene Expr Patterns **4**(4): 467-71.

Lois, C., E. J. Hong, et al. (2002). "Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors." Science **295**(5556): 868-72.

Loomis, C. A., E. Harris, et al. (1996). "The mouse Engrailed-1 gene and ventral limb patterning." Nature **382**(6589): 360-3.

Lowe, C. B., G. Bejerano, et al. (2007). "Thousands of human mobile element fragments undergo strong purifying selection near developmental genes." Proc Natl Acad Sci USA **104**(19): 8005-10.

Lu, Q. R., T. Sun, et al. (2002). "Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection." Cell **109**(1): 75-86.

Lu, Q. R., D. Yuk, et al. (2000). "Sonic hedgehog--regulated oligodendrocyte lineage genes encoding bHLH proteins in the mammalian central nervous system." Neuron **25**(2): 317-29.

Lunyak, V. V., R. Burgess, et al. (2002). "Corepressor-dependent silencing of chromosomal regions encoding neuronal genes." Science **298**(5599): 1747-52.

Lunyak, V. V., G. G. Prefontaine, et al. (2007). "Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis." Science **317**(5835): 248-51.

Lyle, R., U. Radhakrishna, et al. (2006). "Split-hand/split-foot malformation 3 (SHFM3) at 10q24, development of rapid diagnostic methods and gene expression from the region." Am J Med Genet A **140**(13): 1384-95.

Lyon, M. F. (2000). "LINE-1 elements and X chromosome inactivation: a function for "junk" DNA?" Proc Natl Acad Sci USA **97**(12): 6248-9.

Mager, D. L. and J. D. Freeman (2000). "Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences." J Virol **74**(16): 7221-9.

Mahajan, M. C., S. Karmakar, et al. (2007). "Control of beta globin genes." J Cell Biochem **102**(4): 801-10.

Maksakova, I. A. and D. L. Mager (2005). "Transcriptional regulation of early transposon elements, an active family of mouse long terminal repeat retrotransposons." J Virol **79**(22): 13865-74.

Maksakova, I. A., D. L. Mager, et al. (2008). "Endogenous retroviruses : Keeping active endogenous retroviral-like elements in check: the epigenetic perspective." Cell Mol Life Sci **65**(21): 3329-47.

Maksakova, I. A., M. T. Romanish, et al. (2006). "Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line." PLoS Genet **2**(1): e2.

Maksakova, I. A., Y. Zhang, et al. (2009). "Preferential epigenetic suppression of the autonomous MusD over the non-autonomous ETn mouse retrotransposons." Molecular and Cellular Biology.

Mariani, F., C. Ahn, et al. (2008). "Genetic evidence that FGFs have an instructive role in limb proximal–distal patterning." Nature **453**(7193): 401-405.

Mariño-Ramírez, L. and I. K. Jordan (2006). "Transposable element derived DNaseI-hypersensitive sites in the human genome." Biol Direct **1**: 20.

Martens, J. H., R. J. O'Sullivan, et al. (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." EMBO J **24**(4): 800-12.

Mastrangelo, I. A., A. J. Courey, et al. (1991). "DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement." Proc Natl Acad Sci USA **88**(13): 5670-4.

Matsui, T., D. Leung, et al. (2010). "Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET." Nature.

Mazumder, B., V. Seshadri, et al. (2003). "Translational control by the 3'-UTR: the ends specify the means." Trends in Biochemical Sciences **28**(2): 91-8.

MCCLINTOCK, B. (1956). "Controlling elements and the gene." Cold Spring Harb Symp Quant Biol **21**: 197-216.

Medstrand, P., L. N. van de Lagemaat, et al. (2005). "Impact of transposable elements on the evolution of mammalian gene regulation." Cytogenet Genome Res **110**(1-4): 342-52.

Meng, X., R. Poon, et al. (2001). "Suppressor of fused negatively regulates beta-catenin signaling." J Biol Chem **276**(43): 40113-9.

Merlo, G. R., L. Paleari, et al. (2002). "Mouse model of split hand/foot malformation type I." Genesis **33**(2): 97-101.

Meyers, E. N., M. Lewandoski, et al. (1998). "An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination." Nat Genet **18**(2): 136-41.

Mi, S., X. Lee, et al. (2000). "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis." Nature **403**(6771): 785-9.

Michaud, E. J., M. J. van Vugt, et al. (1994). "Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage." Genes & Development **8**(12): 1463-72.

Miki, Y., I. Nishisho, et al. (1992). "Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer." Cancer Res **52**(3): 643-5.

Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-60.

Mills, A. A., B. Zheng, et al. (1999). "p63 is a p53 homologue required for limb and epidermal morphogenesis." Nature **398**(6729): 708-13.

Mills, R., E. A. Bennett, et al. (2006). "Recently mobilized transposons in the human and chimpanzee genomes." Am J Hum Genet **78**(4): 671-9.

Mills, R., K. Walter, et al. (2011). "Mapping copy number variation by population-scale genome sequencing." Nature **470**(7332): 59-65.

Min, H., D. M. Danilenko, et al. (1998). "Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless." Genes & Development **12**(20): 3156-61.

Mola, G., E. Vela, et al. (2007). "Exonization of Alu-generated splice variants in the survivin gene of human and non-human primates." J Mol Biol **366**(4): 1055-63.

Moon, A. M. and M. R. Capecchi (2000). "Fgf8 is required for outgrowth and patterning of the limbs." Nat Genet **26**(4): 455-9.

Morgan, H. D., H. G. Sutherland, et al. (1999). "Epigenetic inheritance at the agouti locus in the mouse." Nat Genet **23**(3): 314-8.

Muotri, A., V. Chu, et al. (2005). "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition." Nature **435**(7044): 903-10.

Muotri, A., M. C. Marchetto, et al. (2010). "L1 retrotransposition in neurons is modulated by MeCP2." Nature **468**(7322): 443-6.

Muramatsu, M., K. Kinoshita, et al. (2000). "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme." Cell **102**(5): 553-63.

Murchison, E. P., P. Stein, et al. (2007). "Critical roles for Dicer in the female germline." Genes & Development **21**(6): 682-93.

Nekrutenko, A. and W. H. Li (2001). "Transposable elements are found in a large number of human protein-coding genes." Trends Genet **17**(11): 619-21.

Ni, T., D. Corcoran, et al. (2010). "A paired-end sequencing strategy to map the complex landscape of transcription initiation." Nat Methods **7**(7): 521-7.

Niswander, L., S. Jeffrey, et al. (1994). "A positive feedback loop coordinates growth and patterning in the vertebrate limb." Nature **371**(6498): 609-12.

Noben-Trauth, N., G. Köhler, et al. (1996). "Efficient targeting of the IL-4 gene in a BALB/c embryonic stem cell line." Transgenic Res **5**(6): 487-91.

Nobrega, M. A., I. Ovcharenko, et al. (2003). "Scanning human gene deserts for long-range enhancers." Science **302**(5644): 413.

Nunes, M. E., G. Schutt, et al. (1995). "A second autosomal split hand/split foot locus maps to chromosome 10q24-q25." Human Molecular Genetics **4**(11): 2165-70.

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

Ohlsson, R., R. Renkawitz, et al. (2001). "CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease." Trends Genet **17**(9): 520-7.

Ohno (1972). "An argument for the genetic simplicity of man and other mammals." Journal of Human Evolution **1**(6): 651-662.

Ohtsuki, S. and M. Levine (1998). "GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo." Genes & Development **12**(21): 3325-30.

Ohtsuki, S., M. Levine, et al. (1998). "Different core promoters possess distinct regulatory activities in the Drosophila embryo." Genes & Development **12**(4): 547-56.

Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.

Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-7.

Osborne, C. S., L. Chakalova, et al. (2004). "Active genes dynamically colocalize to shared sites of ongoing transcription." Nat Genet **36**(10): 1065-71.

Ovcharenko, I., G. G. Loots, et al. (2005). "Evolution and functional classification of vertebrate gene deserts." Genome Res **15**(1): 137-45.

Ozen, R. S., B. E. Baysal, et al. (1999). "Fine mapping of the split-hand/split-foot locus (SHFM3) at 10q24: evidence for anticipation and segregation distortion." Am J Hum Genet **64**(6): 1646-54.

Peaston, A., A. Evsikov, et al. (2004). "Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos." Dev Cell **7**(4): 597-606.

Pennacchio, L. A., N. Ahituv, et al. (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Peters, A. H., D. O'Carroll, et al. (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell **107**(3): 323-37.

Pi, W., X. Zhu, et al. (2010). "Long-range function of an intergenic retrotransposon." Proc Natl Acad Sci USA **107**(29): 12992-7.

Pickeral, O. K., W. Makałowski, et al. (2000). "Frequent human genomic DNA transduction driven by LINE-1 retrotransposition." Genome Res **10**(4): 411-5.
Prak, E. T. and H. H. Kazazian (2000). "Mobile elements and the human genome." Nat Rev Genet **1**(2): 134-44.
Rakyan, V. K., M. E. Blewitt, et al. (2002). "Metastable epialleles in mammals." Trends Genet **18**(7): 348-51.
Rakyan, V. K., S. Chong, et al. (2003). "Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission." Proc Natl Acad Sci USA **100**(5): 2538-43.
Reiss, D. and D. L. Mager (2007). "Stochastic epigenetic silencing of retrotransposons: does stability come with age?" Gene **390**(1-2): 130-5.
Ribet, D., M. Dewannieux, et al. (2004). "An active murine transposon family pair: retrotransposition of "master" MusD copies and ETn trans-mobilization." Genome Res **14**(11): 2261-7.
Ribet, D., F. Harper, et al. (2007). "Murine MusD retrotransposon: structure and molecular evolution of an "intracellularized" retrovirus." J Virol **81**(4): 1888-98.
Riddle, R. D., R. L. Johnson, et al. (1993). "Sonic hedgehog mediates the polarizing activity of the ZPA." Cell **75**(7): 1401-16.
Riethoven, J. (2010). "Regulatory regions in DNA: promoters, enhancers, silencers, and insulators." Methods Mol Biol **674**: 33-42.
Rinne, T., H. G. Brunner, et al. (2007). "p63-associated disorders." cc **6**(3): 262-8.
Robledo, R. F., L. Rajan, et al. (2002). "The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development." Genes & Development **16**(9): 1089-101.
Rollins, R. A., F. Haghighi, et al. (2006). "Large-scale structure of genomic methylation patterns." Genome Res **16**(2): 157-63.
Rowe, H., J. Jakobsson, et al. (2010). "KAP1 controls endogenous retroviruses in embryonic stem cells." Nature **463**(7278): 237-40.
Rubin, C. M., C. A. VandeVoort, et al. (1994). "Alu repeated DNAs are differentially methylated in primate germ cells." Nucleic Acids Res **22**(23): 5121-7.
Ryan, R. F., D. C. Schultz, et al. (1999). "KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing." Mol Cell Biol **19**(6): 4366-78.
Safe, S. and K. Kim (2004). "Nuclear receptor-mediated transactivation through interaction with Sp proteins." Prog Nucleic Acid Res Mol Biol **77**: 1-36.
Sambrook, J. o. s. e. p. h. and D. a. v. i. d. William Russell (2006). "The condensed protocols from Molecular cloning : a laboratory manual." 800.
Sandelin, A., P. Carninci, et al. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nat Rev Genet **8**(6): 424-36.
Sasaki, T., H. Nishihara, et al. (2008). "Possible involvement of SINEs in mammalian-specific brain formation." Proc Natl Acad Sci USA **105**(11): 4220-5.
SAUNDERS, J. W. (1948). "The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm." J Exp Zool **108**(3): 363-403.
Schäfer, K. and T. Braun (1999). "Early specification of limb muscle precursor cells by the homeobox gene Lbx1h." Nat Genet **23**(2): 213-6.
Scherz, P. J., B. D. Harfe, et al. (2004). "The limb bud Shh-Fgf feedback loop is terminated by expansion of former ZPA cells." Science **305**(5682): 396-9.
Schultz, D. C., K. Ayyanathan, et al. (2002). "SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins." Genes Dev **16**(8): 919-32.

Schwartz, Y. B. and V. Pirrotta (2007). "Polycomb silencing mechanisms and the management of genomic programmes." Nat Rev Genet **8**(1): 9-22.

Scott, K. C., A. D. Taubman, et al. (1999). "Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength." Genetics **153**(2): 787-98.

Sekine, K., H. Ohuchi, et al. (1999). "Fgf10 is essential for limb and lung formation." Nat Genet **21**(1): 138-41.

Seo, E., H. Kim, et al. (2009). "Multiple isoforms of beta-TrCP display differential activities in the regulation of Wnt signaling." Cellular Signalling **21**(1): 43-51.

Seto, M. L., M. E. Nunes, et al. (1997). "Pathogenesis of ectrodactyly in the Dactylaplasia mouse: aberrant cell death of the apical ectodermal ridge." Teratology **56**(4): 262-70.

Shaw, C. J. and J. R. Lupski (2004). "Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease." Human Molecular Genetics **13 Spec No 1**: R57-64.

She, X., Z. Cheng, et al. (2008). "Mouse segmental duplication and copy number variation." Nat Genet **40**(7): 909-14.

Sheen, F. M., S. T. Sherry, et al. (2000). "Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition." Genome Res **10**(10): 1496-508.

Sidow, A., M. S. Bulotsky, et al. (1999). "A novel member of the F-box/WD40 gene family, encoding dactylin, is disrupted in the mouse dactylaplasia mutant." Nat Genet **23**(1): 104-7.

Sijen, T. and R. H. Plasterk (2003). "Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi." Nature **426**(6964): 310-4.

Simonis, M., P. Klous, et al. (2006). "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)." Nat Genet **38**(11): 1348-54.

Simons, C., M. Pheasant, et al. (2006). "Transposon-free regions in mammalian genomes." Genome Res **16**(2): 164-72.

Siomi, M., K. Sato, et al. (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-58.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.

Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." Annu Rev Biochem **72**: 449-79.

Smith, A. M., M. J. Sanchez, et al. (2008). "A novel mode of enhancer evolution: the Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity." Genome Res **18**(9): 1422-32.

Summerbell, D. (1974). "A quantitative analysis of the effect of excision of the AER from the chick limb-bud." J Embryol Exp Morphol **32**(3): 651-60.

Sun, F. L., K. Haynes, et al. (2004). "cis-Acting determinants of heterochromatin formation on Drosophila melanogaster chromosome four." Mol Cell Biol **24**(18): 8210-20.

Sun, T., Y. Echelard, et al. (2001). "Olig bHLH proteins interact with homeodomain proteins to regulate cell fate acquisition in progenitors of the ventral neural tube." Curr Biol **11**(18): 1413-20.

Sun, T., B. P. Hafler, et al. (2006). "Evidence for motoneuron lineage-specific regulation of Olig2 in the vertebrate neural tube." Dev Biol **292**(1): 152-64.

Svoboda, P., P. Stein, et al. (2004). "RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos." Developmental Biology **269**(1): 276-85.

Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." Science **324**(5929): 930-5.

Takasu, M., R. Hayashi, et al. (2007). "Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon." Tissue Antigens **70**(2): 144-50.

Tarlinton, R. E., J. Meers, et al. (2006). "Retroviral invasion of the koala genome." Nature **442**(7098): 79-81.

Thien, H. and U. Rüther (1999). "The mouse mutation Pdn (Polydactyly Nagoya) is caused by the integration of a retrotransposon into the Gli3 gene." Mamm Genome **10**(3): 205-9.

Thomas, J. H. and S. E. Schneider (2011). "Coevolution of retroelements and tandem zinc finger genes." Genome Res.

Tickle, C. (1981). "The number of polarizing region cells required to specify additional digits in the developing chick wing." Nature **289**(5795): 295-8.

Tsumura, A., T. Hayakawa, et al. (2006). "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." Genes Cells **11**(7): 805-14.

Tümpel, S., F. Cambronero, et al. (2008). "A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2." Proc Natl Acad Sci USA **105**(51): 20077-82.

Ugur, S. A. and A. Tolun (2008). "Homozygous WNT10b Mutation and Complex Inheritance in Split Hand Foot Malformation." Human Molecular Genetics.

Urrutia, R. (2003). "KRAB-containing zinc-finger repressor proteins." Genome Biol **4**(10): 231.

Valinluck, V. and L. Sowers (2007). "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1." Cancer Res **67**(3): 946-50.

van Bokhoven, H., B. C. Hamel, et al. (2001). "p63 Gene mutations in eec syndrome, limb-mammary syndrome, and isolated split hand-split foot malformation suggest a genotype-phenotype correlation." Am J Hum Genet **69**(3): 481-92.

Vasyutina, E., J. Stebler, et al. (2005). "CXCR4 and Gab1 cooperate to control the development of migrating muscle progenitor cells." Genes & Development **19**(18): 2187-98.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Volff, J. N. (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes." Bioessays **28**(9): 913-22.

Volff, J. N., C. Körting, et al. (2001). "Evolution and discontinuous distribution of Rex3 retrotransposons in fish." Molecular biology and evolution **18**(3): 427-31.

Volff, J. N., C. Körting, et al. (2000). "Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes." Molecular biology and evolution **17**(11): 1673-84.

Walsh, T., J. M. McClellan, et al. (2008). "Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia." Science **320**(5875): 539-43.

Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-43.

Wei, W. and M. D. Brennan (2001). "The gypsy insulator can act as a promoter-specific transcriptional stimulator." Molecular and Cellular Biology **21**(22): 7714-20.

West, A. G., S. Huang, et al. (2004). "Recruitment of histone modifications by USF proteins at a vertebrate barrier element." Mol Cell **16**(3): 453-63.

Williams, K., J. Christensen, et al. (2011). "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity." Nature **473**(7347): 343-348.

Wolf, D. and S. P. Goff (2007). "TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells." Cell **131**(1): 46-57.

Wolf, D. and S. P. Goff (2009). "Embryonic stem cells use ZFP809 to silence retroviral DNAs." Nature.

Wolf, D., K. Hug, et al. (2008). "TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells." Proc Natl Acad Sci USA **105**(34): 12521-6.

Yang, A., R. Schweitzer, et al. (1999). "p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development." Nature **398**(6729): 714-8.

Ymer, S., W. Q. Tucker, et al. (1985). "Constitutive synthesis of interleukin-3 by leukaemia cell line WEHI-3B is due to retroviral insertion near the gene." Nature **317**(6034): 255-8.

Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-40.

Zamudio, N. and D. Bourc'his (2010). "Transposable elements in the mammalian germline: a comfortable niche or a deadly trap?" Heredity **105**(1): 92-104.

Zeller, R., J. López-Ríos, et al. (2009). "Vertebrate limb bud development: moving towards integrative analysis of organogenesis." Nat Rev Genet **10**(12): 845-58.

Zhang, Y., I. A. Maksakova, et al. (2008). "Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements." PLoS Genet **4**(2): e1000007.

Zhou, Q. and D. J. Anderson (2002). "The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification." Cell **109**(1): 61-73.

Zhou, V., A. Goren, et al. (2011). "Charting histone modifications and the functional organization of mammalian genomes." Nat Rev Genet **12**(1): 7-18.

Zúñiga, A., A. P. Haramis, et al. (1999). "Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds." Nature **401**(6753): 598-602.

8. Indices

8.1 List of Figures

8.1.1 Introduction

8.1.2 Results and Discussion

8.2 List of Tables

8.2.1 Introduction

8.2.2 Materials and Methods

8.2.3 Results and Discussion

8.3 Abbreviations

5hmC 5'hydroxymethyl Cytosine
5mC 5'methyl Cytosine
AER Apical ectodermal ridge
BCIP 5-Bromo-4-chloro-3-indolyl-phosphate 4-toluidine salt
BMP Bone morphogenic protein
C.elegans *Caenorhabditis elegans*
cis-NAT cis-natural antisense transcript
CNV copy number variation
DIG Digoxigenin
DMSO Dimethyl sulfoxide
DNA deoxyribonucleic acid
dsRNA double stranded RNA
E embryonic stage
e.g. exempli gratia (for example)
EBA Enhancer blocking assay
EDTA Ethylenediaminetetraacetic acid
ERV Endogeneous Retrovirus
ES Embryonic Stem
et al. et alii (and others)
ETn Early Transposon
FGF Fibroblast growth factor
g gram
H3K4/9/2027 histone3 lysine4/9/20/27
i.e. id est (that is)
IAP Intracisternal A-particle
IPTG Isopropyl β-D-1-thiogalactopyranoside
kb kilobase
KoRV Koala retrovirus
L1 (LINE1) Long Interspersed Nuclear Element-1
LCR Low copy repeats
LINE Long Interspersed Nuclear Element
LTR Long Terminal Repeat
m mili (prefix)
M molar
Mb megabase
Mdac Modifier of *Dactylaplasia*
MLV Murine Leukemia Virus
MOPS 3-(N-morpholino) propanesulfonic acid
MT Mouse Transposon
MuERV-L Murine endogenousretrovirus-L
MusD Mouse type-D virus
n nano (prefix)
N normal
NAHR Non-allelic homologous recombination
NBT 4-Nitro blue tetrazolium chloride
NHEJ Non-homologous end joining
NMD Nonsense mediated decay
ORF Open reading frame
p pico (prefix)
PBS Phosphate buffered saline
PBS Primer binding site
PcG Polycomb Group
PFA Paraformaldehyde
piRNA PIWI-interacting RNA
PPT Poly purine tract

PRC Polycomb Repressive Complex

RNA ribonucleic acid

RNAi RNA interference

SDS Sodium dodecyl sulfate

SHFM Split hand-foot malformation

SINE Short Interspersed Nuclear Element

siRNA small interfering RNA

SNP single nucleotide polymorphism

SV Structural Variation

TE Transposable Element

TFR Transposon free region

Tris 2-Amino-2-hydroxymethyl-propane-1,3-diol

UTR untranslated region

WT Wild Type

XCI X Chromosome inactivation

X-Gal 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Xic X Chromosome inactivation center

ZPA Zone of polarizing activity

ZRS ZPA regulatory sequence

μ micro (prefix)

9. Appendix

9.1 Acknowledgements

This period of approximately 4 years has been challenging and sometimes bitter due to many different reasons, which do not matter anymore. So many people helped me get through these difficulties and perhaps I should start with acknowledging them.

I would like to thank François Spitz for his good supervision of my PhD project, and his support to my opinions and confidence in my scientific vision, which helped me mature scientifically enormously. Then I would also like to thank my TAC members: Prof. Jochen Wittbrodt, Dr. Eileen Furlong and Dr. Asifa Akhtar for their valuable input and suggestions.

I would also like to thank all the members of Spitz Group for the great support they provided and for standing to hours and hours long of scientific discussion during my group meeting presentations. I especially thank to Veli Uslu for helping me to stay as a sane human being, to Katja Langenfeld for her neat (seraphic) personality and gracious cooperativity, and to Wibke Schwarzer for understanding the technical difficulties I have encountered and helping me to cope with them, as we shared the same destiny. Finally, as an alumnus of Spitz Group I would like to thank Dasha Shlyueva for her lovely fellowship.

I am grateful to my dear predoc friends; Boryana Petrova and Ilaria Piazza (crazy corner) for their continuous reminder on disorderliness, Fargol Mazaheri and Proteeti Bhattacharjee for their sincere friendship as well as their cooking and feeding efforts. As previous EMBL predocs I appreciate the precious friendship of Sevil Yavuz and all the time we spent together watching movies, going out and having numerous coffee breaks, the intellectual feed-back from Erinç Hallaçlı and Oğuz Kanca.

For all the coffee breaks and Neckarwiese gatherings that helped me get away from work from some time to time, I would like to thank the Turkish community of EMBL: Yuva Öz, Murat İskar, Sevi Durdu and Bora Uyar.

I am thankful to my family, as my mother Öznur (Şenel) Aktaş supported me with no exception for every decision I make, my father Raşit Aktaş, as he never gives up believing in I will succeed anything as long as I want to, and my brother Tuğcan Aktaş for not only teaching me everything I know about mathematics, physics and computer science but also being the greatest brother on earth.

Finally, yet importantly, I want to thank my husband İbrahim Avşar Ilık for shaping me and turning me into what I am now, I am so happy he never gave up and made me a better person. Nothing would be possible without your eternal love and assistance, thank you.

I would like to end with a quote from my favorite anime:
“A lesson without pain is meaningless. That is because you cannot gain something without sacrificing something else in return. But, once you've withstood the pain and overcome it, you will gain a heart that is stronger than everything else.
Yes. A Fullmetal heart**.”**
Edward Elric, Fullmetal Alchemist Brotherhood, Episode64

Tuğçe Aktaş Ilık, 2011 Heidelberg Germany

9.2 Publications

Helen M. Rowe, Johan Jakobsson, Daniel Mesnard, Jacques Rougemont, Séverine Reynard, **Tugce Aktas**, Pierre V. Maillard, Hillary Layard-Liesching, Sonia Verp, Julien Marquis, François Spitz, Daniel B. Constam& Didier Trono, **KAP1 controls endogenous retroviruses in embryonic stem cells**. Nature (2010) vol. 463 (7278) pp. 237-40

LETTERS

KAP1 controls endogenous retroviruses in embryonic stem cells

Helen M. Rowe[1], Johan Jakobsson[1]†, Daniel Mesnard[1], Jacques Rougemont[1], Séverine Reynard[1], Tugce Aktas[2], Pierre V. Maillard[1], Hillary Layard-Liesching[1], Sonia Verp[1], Julien Marquis[1], François Spitz[2], Daniel B. Constam[1] & Didier Trono[1]

More than forty per cent of the mammalian genome is derived from retroelements, of which about one-quarter are endogenous retroviruses (ERVs)[1]. Some are still active, notably in mice the highly polymorphic early transposon (ETn)/MusD and intracisternal A-type particles (IAP)[2,3]. ERVs are transcriptionally silenced during early embryogenesis by histone and DNA methylation[4–6] (and reviewed in ref. 7), although the initiators of this process, which is essential to protect genome integrity[8], remain largely unknown. KAP1 (KRAB-associated protein 1, also known as tripartite motif-containing protein 28, TRIM28) represses genes by recruiting the histone methyltransferase SETDB1, heterochromatin protein 1 (HP1) and the NuRD histone deacetylase complex[9], but few of its physiological targets are known. Two lines of evidence suggest that KAP1-mediated repression could contribute to the control of ERVs: first, KAP1 can trigger permanent gene silencing during early embryogenesis[10], and second, a KAP1 complex silences the retrovirus murine leukaemia virus in embryonic cells[11–13]. Consistent with this hypothesis, here we show that KAP1 deletion leads to a marked upregulation of a range of ERVs, in particular IAP elements, in mouse embryonic stem (ES) cells and in early embryos. We further demonstrate that KAP1 acts synergistically with DNA methylation to silence IAP elements, and that it is enriched at the 5′ untranslated region (5′UTR) of IAP genomes, where KAP1 deletion leads to the loss of histone 3 lysine 9 trimethylation (H3K9me3), a hallmark of KAP1-mediated repression. Correspondingly, IAP 5′UTR sequences can impose *in cis* KAP1-dependent repression on a heterologous promoter in ES cells. Our results establish that KAP1 controls endogenous retroelements during early embryonic development.

KAP1, a member of the RBCC (ring, B-box, coiled–coiled) or TRIM (tripartite motif) family of proteins, is recruited to genes by the tetrapod-specific, DNA sequence-specific KRAB-ZFPs (Krüppel-associated box domain-zinc finger proteins)[14], which constitute the largest family of transcriptional regulators encoded by higher vertebrates. However, until now, few KAP1 target genes and their KRAB zinc finger intermediates have been identified[15]. To assess the potential role of KAP1 in ERV regulation we generated two conditional KAP1-knockout ES cell lines, in which the *Kap1* gene could be inactivated by a 4-hydroxytamoxifen (4-OHT)-inducible system. Both lines expressed pluripotency markers, and one was used to demonstrate contribution to mouse chimaeras after injection into blastocysts (Fig. 1, Supplementary Fig. 1 and data not shown). Control and KAP1-deleted ES cells (see Fig. 1a) were subjected to a combination of large-scale RNA sequencing and specific PCR with reverse transcription (RT–PCR) measurements. The results showed a modest increase in LINE1 (long interspersed nuclear elements 1) transcripts after KAP1 removal, but a marked upregulation of a range of ERVs, in particular IAP elements, which exhibited 15- and 66-fold overexpression in the two ES cell lines, respectively (Fig. 1b–d and Supplementary Figs 1 and 2). Because ES cells progressively lose self-renewal ability after KAP1 depletion (refs 16, 17 and data not shown), as a control we examined ES cells cultured in the absence of leukaemia inhibitory factor (LIF). Under these conditions we also observed a decrease in the stem cell markers NANOG and to a lesser extent OCT4 (also known as POU5F1), but IAP elements were not upregulated (Fig. 1a, c). Furthermore, partially restoring KAP1 levels by transduction with a KAP1-expressing lentiviral vector proportionately reduced the upregulation of IAP elements in KAP1-deleted ES cells (Supplementary Fig. 3). Notably, the stimulation of IAP transcription was accompanied by an increase in the IAP DNA load of KAP1-deleted compared to control ES cells, demonstrating that IAP genes were not only overexpressed but could also reverse transcribe and probably integrate into the genome (Fig. 1e). In contrast, IAP expression was not increased after KAP1 deletion in mouse embryonic fibroblasts (MEFs) (Supplementary Fig. 4), which supports a model in which ERV control in differentiated cells relies on more stable silencing mechanisms[4,5].

To investigate more thoroughly the mechanisms of IAP control during early embryogenesis, we treated ES cells with the DNA methyltransferase inhibitor 5-azacytidine (5-aza) and observed a similar upregulation of IAP elements. Notably, combining this drug with KAP1-knockout induced a synergistic effect on IAP overexpression (Fig. 1f). This suggests that DNA methylation and KAP1 repression act cooperatively to silence these elements. Moreover, validating the results of our ES cell-based experiments, we found that IAP elements were markedly increased in KAP1-depleted blastocysts cultured *ex vivo* (Supplementary Fig. 5), and, most importantly, were upregulated more than five-hundred times in KAP1-knockout embryos (Fig. 2).

These data indicate that KAP1 has a crucial role in controlling ERVs during early embryonic development. To determine whether this effect is direct, we performed chromatin immunoprecipitation (ChIP) studies. First, we found that KAP1 associates with the IAP genome in ES cells (Fig. 3a, b). Notably, KAP1 was significantly enriched over the IAP 5′UTR compared to the U3 ($P = 0.034$) or *gag* ($P = 0.026$) regions in control cells. Furthermore, KAP1 knockout correlated with a decrease in H3K9 trimethylation and an increase in H4 acetylation on the IAP genome (Fig. 3c, d). This combination of chromatin modifications is consistent with a loss of KAP1-mediated repression[9]. We also assessed the global level of IAP DNA methylation by Southern blot and more specifically by bisulphite sequencing, but did not detect a significant difference with or without KAP1 (not shown).

To evaluate the genetic diversity of IAP elements controlled by KAP1, we sequenced the 5′UTR of IAP complementary DNAs isolated from KAP1-depleted cells. The 33 sequences thereby analysed were highly diverse (Fig. 4a and Supplementary Fig. 7). Of interest

[1]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [2]EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany. †Present address: Wallenberg Neuroscience Center, Department of Experimental Medical Sciences, Lund University, 22184 Lund, Sweden.

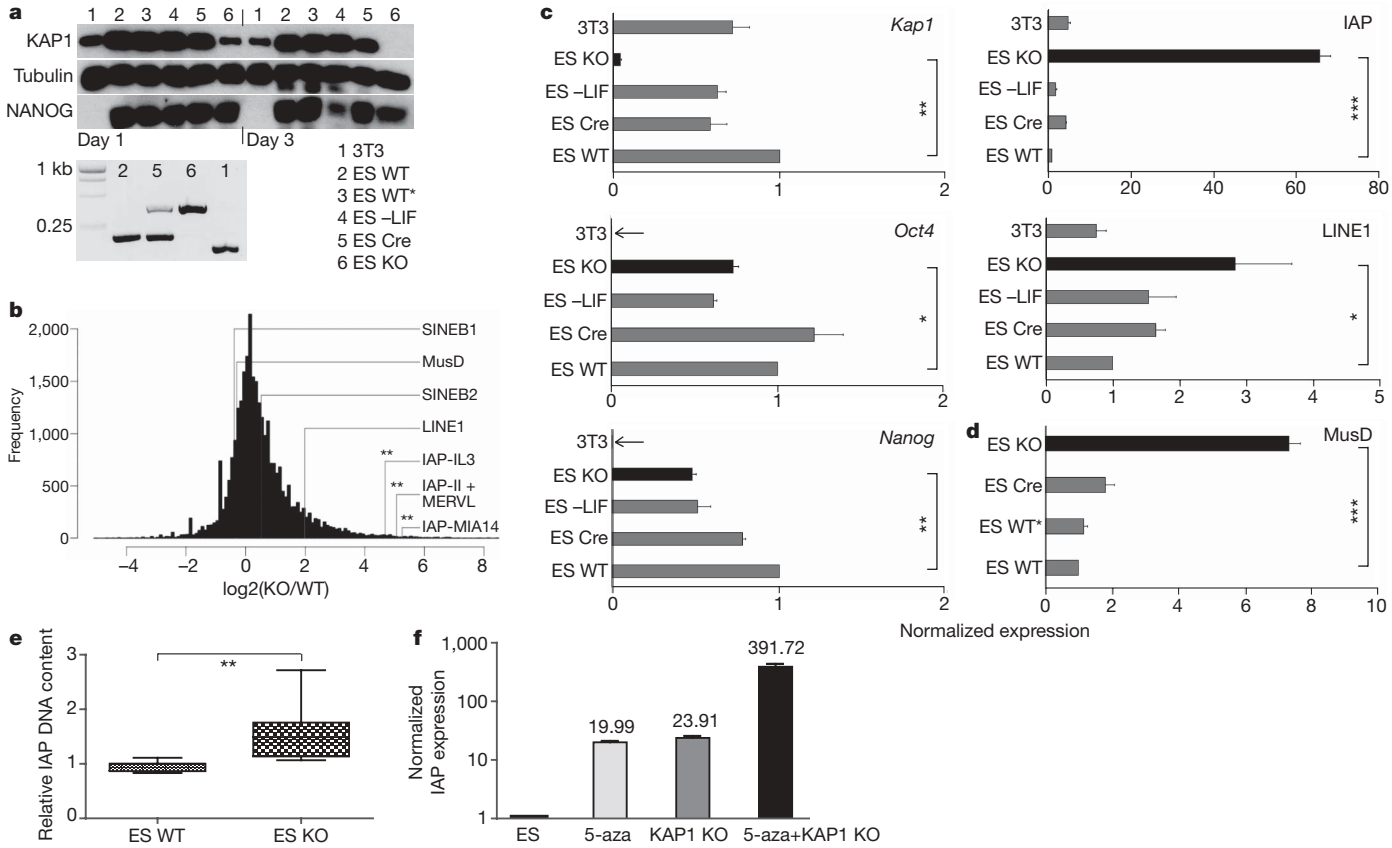


Figure 1 | IAP elements are upregulated in KAP1-depleted ES cells. **a**, Western blot and KAP1 PCR[27] after 4-OHT addition (times indicated). WT, wild-type; WT*, treated with 4-OHT; −LIF, cultured without LIF; Cre, cells transduced with a 4-OHT-inducible Cre vector; KO, KAP1 knockout Cre cells treated with 4-OHT. KAP1 is 95 kDa, tubulin 52 kDa, and NANOG 42 kDa. PCR: 152-, 171- and 390-bp products correspond to wild-type, *loxP*-flanked or excised *Kap1*, respectively. **b**, RNA-sequencing. Histogram shows the difference in gene expression between wild-type and knockout samples after *Gapdh* normalization. Double asterisks, ≤1% of genes are more highly upregulated than IAP and MERVL elements. NLM GenBank accession numbers: X04120 (IAP-IL3), M17551 (IAP-MIA14), X03521 (106–494 bp, insertion sequence specific to IAP-II class), AF030883 (SINEB1 and 2), AB305072/3 (MusD) and M13002 (LINE1). MERVL was from Repbase (see Supplementary Fig. 2). **c**, Quantitative RT–PCR analysis on stated treatment groups ($n = 3$) normalized to ES wild-type cells showing the mean and s.d. IAP primers were specific for the 5′UTR region. One representative experiment of three is shown. Unpaired *t*-tests were used to compare wild-type and knockout ES cells. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. $P = 0.002$ (*Kap1*), $P = 0.039$ (*Oct4*), $P = 0.004$ (*Nanog*), $P \leq 0.001$ (IAP) and $P = 0.023$ (LINE1). **d**, Results for the second ES line, analysed as described in **a**. $P \leq 0.001$ for MusD. See also Supplementary Fig. 1. **e**, Relative IAP DNA content was measured by quantitative PCR (qPCR) on genomic DNA. A summary of three experiments is shown ($n = 7$) as a box plot with error bars indicating the lowest and highest values. $P = 0.0012$, two-tailed Mann–Whitney test. **f**, IAP upregulation was compared after treatment with 5-aza ($n = 3$) (at 7 μM, added for 24 h), after KAP1 deletion, or after a combination of both (5-aza was added for 24 h 3 days after 4-OHT addition). Results are mean and s.d. One representative experiment of two is shown.

was their heterogeneity in the region coding for the primer-binding site (PBS)—the sequence complementary to the cellular transfer RNA that acts as primer for retroviral minus-strand DNA synthesis. Murine leukaemia virus (MLV) is indeed silenced in embryonic cells by the ZFP809-mediated recruitment of a KAP1-containing complex to its proline tRNA PBS (PBS Pro)[11–13]. The PBS of IAP elements expressed in KAP1-depleted cells covered a range closely related to PBS Phe (Supplementary Fig. 8). To address whether these PBS variants were sufficient to confer KAP1-sensitivity to a heterologous promoter in ES cells, we cloned them into a lentiviral vector expressing a green fluorescent protein (GFP) transgene from the MND (myeloproliferative sarcoma virus enhancer, negative control region deleted[18]) promoter. We then transduced KAP1-excisable ES cells with the resulting vectors and examined GFP expression. Although the MLV Pro sequence induced potent KAP1-dependent silencing compared to its functionally inactive B2 point mutant, none of the IAP PBS variants induced significant repression (Supplementary Fig. 8). This concurs with the results of a previous study in which the silencing activity of two

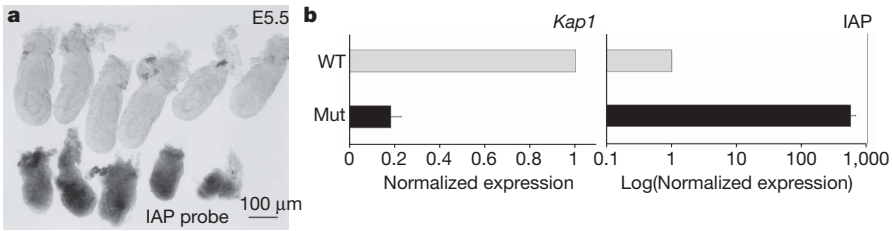


Figure 2 | KAP1 controls IAP elements in embryogenesis. **a**, Embryos from $Kap1^{+/-}$ intercrosses were dissected and *in situ* hybridization performed with a 500-bp IAP probe (specific for a cDNA region spanning the IAP 5′UTR as shown in Fig. 4a). One litter is shown; in total, 23 out of 84 embryos (27%) stained positive for IAP elements. See also Supplementary Fig. 6. E5.5, embryonic day 5.5. **b**, Mice were crossed as above and embryos dissected at E6.5, and two mutant versus two wild-type embryos selected and pooled based on morphology. Samples were analysed by qRT-PCR at three dilutions and results show the mean and s.d. One of two experiments is shown.

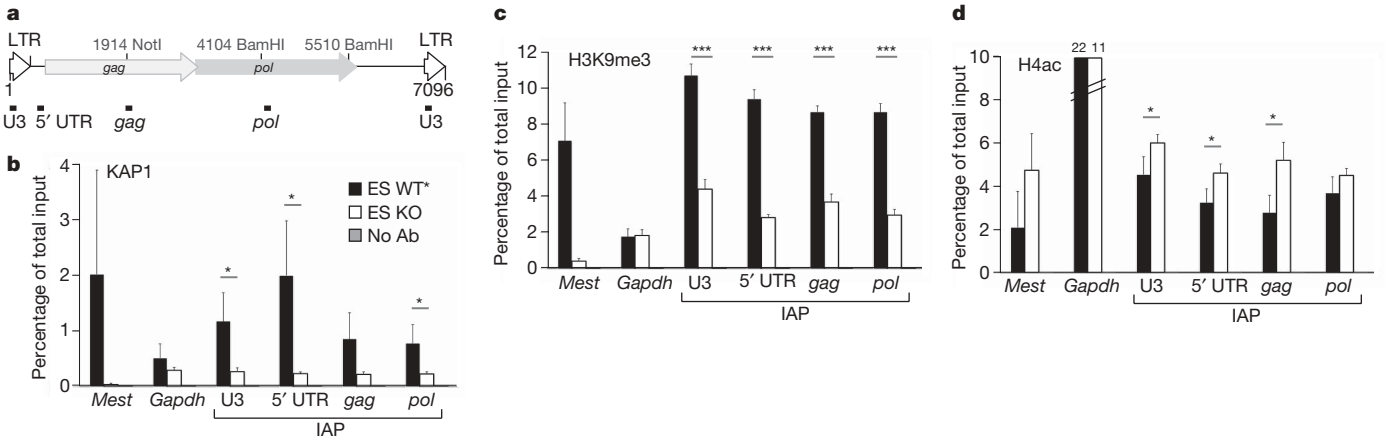


Figure 3 | KAP1 is enriched at the 5′UTR of IAP genomes and loss of KAP1 leads to loss of H3K9me3 and an increase in H4 acetylation. **a**, IAP map (based on IAP-MIA14; ref. 28) with positions of primers used for ChIPs. LTR, long terminal repeat. **b**, KAP1 ChIP results 4 days after 4-OHT addition to control (ES WT*) or Cre-expressing (ES KO) ES cells. Graphs show the mean enrichment in the immunoprecipitations ($n = 3$) relative to the total input samples and error bars show the s.d. A control with no antibody (Ab) gave background enrichment (mean 0.008%). All significant differences between wild-type and knockout for IAP primers are marked. *Mest* (mesoderm-specific transcript), a direct target of KAP1 in embryonic carcinoma (EC) cells[29], was a positive control and *Gapdh* a negative control. **c**, **d**, ChIPs as above but with an anti-H3K9me3 antibody (**c**), and with an antibody specific for acetylated H4 (**d**). Data are representative of 2–3 experiments. *P* values (by unpaired *t* tests) are *Kap1*: $P = 0.037$ (U3); $P = 0.036$ (5′UTR); $P = 0.047$ (*pol*), but summary values for the three experiments are $P < 0.0001$ for the U3 and 5′UTR, and $P < 0.003$ for *gag* and *pol*. H3K9me3: $P \leq 0.001$ (U3, 5′UTR, *gag* and *pol*). H4Ac: $P = 0.049$ (U3); $P = 0.037$ (5′UTR); $P = 0.022$ (*gag*).

other IAP PBS sequences was evaluated[19]. However, when we cloned 500-base-pair-(bp)-long fragments overlapping the 5′UTR of IAP elements expressed in KAP1-null cells (shown in Fig. 4a) either upstream or downstream of the MND promoter, we could induce up to 53-fold KAP1-dependent silencing, which was comparable to the 50-fold repression induced by the MLV Pro sequence (Fig. 4b). Notably, the same region cloned from an IAP element expressed in control cells (IAP1, which has multiple sequence differences, see Supplementary Fig. 10) failed to repress the MND–GFP reporter, thereby serving as a negative control.

Taken together, these data demonstrate that ERVs are repressed in murine embryonic stem cells by the recruitment of a KAP1-containing chromatin remodelling complex to their 5′UTRs, corroborating the previously noted marked enrichment for H3K9me3 at and near ERV

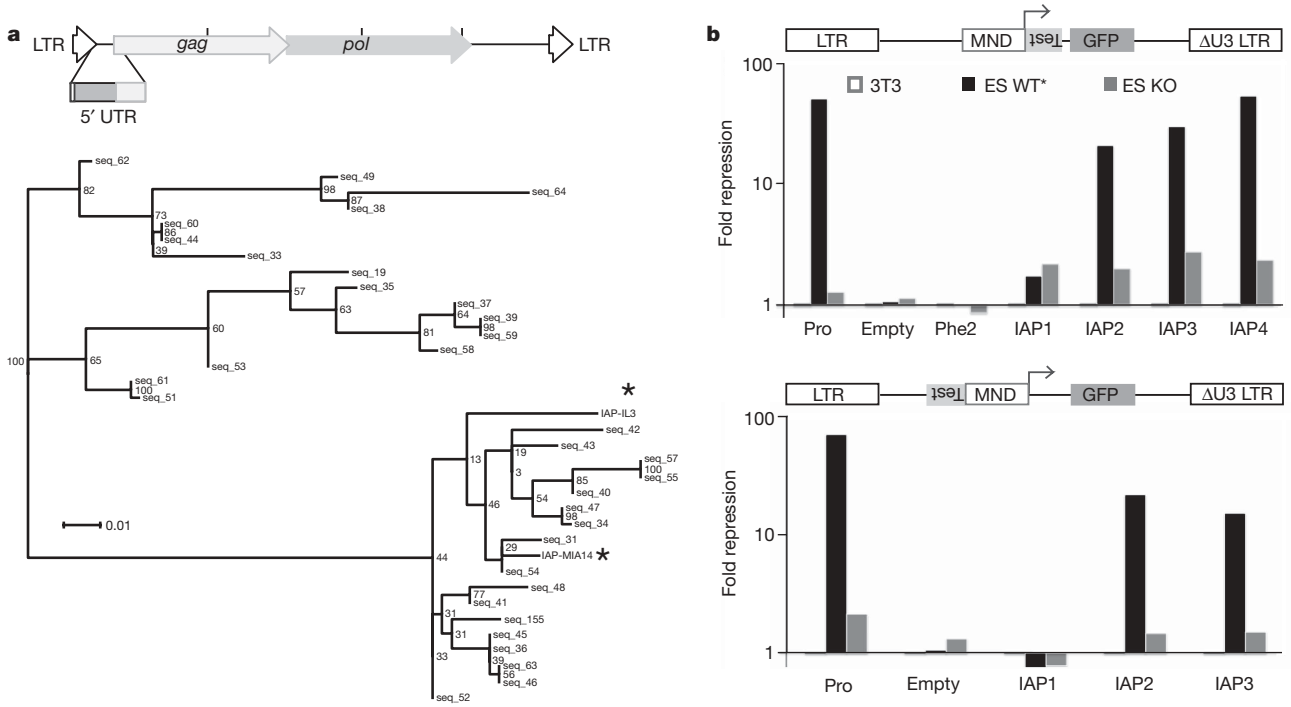


Figure 4 | IAP 5′UTR sequences expressed in KAP1-depleted ES cells are polymorphic and can repress a GFP reporter in a KAP1-dependent way. **a**, Map showing the 500-bp 5′UTR/5′*gag* region of IAP transcripts sequenced, and a phylogenetic tree of the 33 sequences obtained. Sequences for IAP-MIA14 and IAP-IL3 were included in the alignment and are marked by an asterisk. **b**, Lentiviral vector map with a GFP reporter and the test site where the following sequences were cloned antisense: IAP fragments 1–4 or PBS sequences Pro, B2 or Phe2 (see Supplementary Fig. 8). ES cells were transduced (1 day after 4-OHT treatment) with these vectors (or an empty vector) and GFP was measured 3 days later in the SSEA-1^{hi} cell fraction (see Supplementary Fig. 9). Fold repression shows the ratio of expression between these vectors and the control B2 one, normalized to 3T3 cells where the ratio equals 1. The bottom panel shows the results for IAP fragments 1–3 or the PBS sequences, Pro or B2 cloned upstream of the promoter. All sequences are in antisense orientation except IAP2, which is in sense orientation. Results were normalized as above and are representative of 2–3 experiments. Results were also reproduced in E14 ES cells (see Supplementary Fig. 11).

DNA in these cells[5]. The described mechanism thus seems to represent a tetrapod-specific complement to the small-RNA-mediated retrotransposon silencing that is at play from plants to mammals[20,21].

By analogy with the demonstrated mechanism of repression of other targets including MLV, it is likely that KAP1 is tethered to IAP genomes by KRAB-ZFPs. The observed sequence diversity of KAP1-dependent ERVs further suggests a corresponding degree of polymorphism in the zinc fingers mediating their recognition. Our large-scale RNA sequencing analysis detected more than 250 KRAB-ZFP transcripts in ES cells, including 56 at levels higher than ZFP809 (data not shown). Phylogenetic studies further show that the DNA-binding domains of KRAB-ZFPs have been under strong positive selection during evolution, pointing to their participation in genetic conflicts[22]. Our data are consistent with a model in which rapidly mutating retroelements have been protagonists of these conflicts, exerting strong selective pressures on KRAB-ZFPs responsible for their control.

Finally, considering that epigenetic silencing can spread from repetitive elements to neighbouring genes[5,23,24], the work presented here opens new perspectives to explore ERV-mediated control of cellular genes in development and in adult tissues.

METHODS SUMMARY

ES cells. Two ES cell lines were derived from *Kap1*$^{loxp/loxP}$ mice (gift from F. Cammas), karyotyped and cultured as described[25]. KAP1-knockout cells were analysed 4 days after treatment with 4-OHT (used at 1 μM overnight, from Sigma). Differentiation was monitored using an anti-SSEA-1 antibody (BD Pharmingen, MC480). Western blots were performed as described[26] using antibodies specific for KAP1 (Chemicon, MAB3662), NANOG (Abcam, 21603-100), OCT4 (Santa Cruz, sc5279) and α-tubulin (Sigma).

Lentiviral vectors. The LVCT[10] vector was modified to express CAG-4-OHT-inducible Cre (from P. Chambon) and an SV40-puro cassette. An LV PGK-GFP vector was modified by substituting GFP with Cre for MEF experiments, and was also used to construct LV-silencing vectors by replacing PGK with MND[18] and including test sequences upstream or downstream. Vectors were titrated on 3T3 cells.

RT–PCR. Total RNA was purified using a Trizol kit (Invitrogen), treated with DNase (Ambion), and 0.5 μg was reverse transcribed using random primers and SuperScript II (Invitrogen). Alternatively, for low starting material, an RNeasy micro kit was used. Primers (see Supplementary Table 1) were used for SYBR green Q-PCR (Applied Biosystems) and their specificity confirmed with dissociation curves. All data are *Gapdh* normalized, although the actin gene gave similar results. IAP DNA PCR was normalized to the titin gene and results confirmed with *Gapdh*, MusD and major satellite primers.

ChIP. Chromatin was prepared according to the Upstate protocol, and starting material was normalized between wild-type and knockout samples. Triplicate immunoprecipitations were performed using protein A agarose beads (Millipore) and the following rabbit antibodies: KAP1 (ref. 9) (from D. Schultz), H3K9me3 (Abcam, ab8898) and acetyl H4 (Upstate, 06-866). Input and immunoprecipitation samples were analysed by SYBR green Q-PCR.

Bioinformatics. Sequences were aligned with Mafft and a phylogenetic tree generated using RAxML (Randomized Axelerated Maximum Likelihood). Illumina RNA-sequencing was analysed with MAQ 0.7.1 and reads mapped to the collection of mouse transcripts from RefSeq (version 36).

Received 24 September; accepted 11 November 2009.

1. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
2. Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G. & Heidmann, T. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nature Genet.* **36,** 534–539 (2004).
3. Zhang, Y., Maksakova, I. A., Gagnier, L., van de Lagemaat, L. N. & Mager, D. L. Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. *PLoS Genet.* **4,** e1000007 (2008).
4. Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24,** 800–812 (2005).
5. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
6. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20,** 116–117 (1998).
7. Maksakova, I. A., Mager, D. L. & Reiss, D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* **65,** 3329–3347 (2008).
8. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23,** 314–318 (1999).
9. Sripathy, S. P., Stevens, J. & Schultz, D. C. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell. Biol.* **26,** 8623–8638 (2006).
10. Wiznerowicz, M. *et al.* The Kruppel-associated box repressor domain can trigger *de novo* promoter methylation during mouse early embryogenesis. *J. Biol. Chem.* **282,** 34535–34541 (2007).
11. Wolf, D., Cammas, F., Losson, R. & Goff, S. P. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *J. Virol.* **82,** 4675–4679 (2008).
12. Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131,** 46–57 (2007).
13. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458,** 1201–1204 (2009).
14. Friedman, J. R. *et al.* KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10,** 2067–2078 (1996).
15. Krebs, C. J., Larkins, L. K., Khan, S. M. & Robins, D. M. Expansion and diversification of KRAB zinc-finger genes within a cluster including Regulator of sex-limitation 1 and 2. *Genomics* **85,** 752–761 (2005).
16. Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134,** 162–174 (2008).
17. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23,** 837–848 (2009).
18. Haas, D. L. *et al.* The Moloney murine leukemia virus repressor binding site represses expression in murine and human hematopoietic stem cells. *J. Virol.* **77,** 9439–9450 (2003).
19. Modin, C., Lund, A. H., Schmitz, A., Duch, M. & Pedersen, F. S. Alleviation of murine leukemia virus repression in embryonic carcinoma cells by genetically engineered primer binding sites and artificial tRNA primers. *Virology* **278,** 368–379 (2000).
20. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453,** 534–538 (2008).
21. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453,** 539–543 (2008).
22. Emerson, R. O. & Thomas, J. H. Adaptive evolution in zinc finger transcription factors. *PLoS Genet.* **5,** e1000325 (2009).
23. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41,** 563–571 (2009).
24. Whitelaw, E. & Martin, D. I. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nature Genet.* **27,** 361–365 (2001).
25. Bryja, V., Bonilla, S. & Arenas, E. Derivation of mouse embryonic stem cells. *Nature Protocols* **1,** 2082–2087 (2006).
26. Maillard, P. V., Reynard, S., Serhan, F., Turelli, P. & Trono, D. Interfering residues narrow the spectrum of MLV restriction by human TRIM5α. *PLoS Pathog.* **3,** e200 (2007).
27. Cammas, F., Herzog, M., Lerouge, T., Chambon, P. & Losson, R. Association of the transcriptional corepressor TIF1β with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev.* **18,** 2147–2160 (2004).
28. Mietz, J. A., Grossman, Z., Lueders, K. K. & Kuff, E. L. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. *J. Virol.* **61,** 3020–3029 (1987).
29. Riclet, R. *et al.* Disruption of the interaction between transcriptional intermediary factor 1β and heterochromatin protein 1 leads to a switch from DNA hyper- to hypomethylation and H3K9 to H3K27 trimethylation on the MEST promoter correlating with gene reactivation. *Mol. Biol. Cell* **20,** 296–305 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Turelli and I. Barde for advice, F. Cammas and R. Losson for the *Kap1*$^{loxp/loxP}$ and *Kap1*$^{+/-}$ mice, D. Schultz for the KAP1-specific antibody, and K. Harshman at the genomics platform for the Illumina high throughput sequencing. This work was supported by grants from the Swiss National Science Foundation, the Infectigen Association and the Strauss Foundation to D.T., and by a post-doctoral fellowship from the Swedish Research Council to J.J.

Author Contributions H.M.R. performed experiments, with contributions from S.R., T.A., P.V.M., H.L.-L., S.V. and J.M. J.J. developed conditional KAP1-knockout ES cell lines, D.M. performed dissection and *in situ* hybridisation of embryos and J.R. did bioinformatics analyses. P.V.M., F.S., D.B.C., D.M. and J.J. contributed to experimental design. D.T. and H.M.R. designed the study, analysed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.T. (didier.trono@epfl.ch).