Enhancer Composition and Dosage Control Developmental Gene Expression at the *Ihh* Locus

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

by
Anja J. Will

2017

This thesis was completed at the Max-Planck Institute for Molecular Genetics Berlin and Charité - Universitätsmedizin Berlin from January 2012 to March 2017.

First Referee: Prof. Dr. Stefan Mundlos

Institut für Medizinische Genetik und Humangenetik

Augustenburger Platz 1

Campus Virchow-Klinikum

Charité, Universitätsmedizin Berlin

13353 Berlin

E-Mail: stefan.mundlos@charite.de

Max Planck Institut für Molekulare Genetik

Arbeitsgruppe Development & Disease

Ihnestr. 63-73

14195 Berlin

E-Mail: mundlos@molgen.mpg.de

First Referee: Prof. Dr. Günther Weindl

Freie Universität Berlin

Institut für Pharmazie

Pharmakologie und Toxikologie

Königin-Luise-Str 2 und 4

14195 Berlin

E-Mail: Guenther.Weindl@fu-berlin.de

Submitted on 30th of March 2017.

Date of Defence: 30.06.2017.

I hereby declare that this thesis entitled "Enhancer Composition and Dosage Control Developmental Gene Expression at the *Ihh* Locus" has been composed solely by myself, independently and with no other sources or aids than cited or acknowledged. I testify that this thesis has not been previously submitted, in whole or in part, for a degree or any other qualification.

Berlin, 30th of March 2017

_____

Anja J. Will

Acknowledgements

Herewith I am expressing my gratitude to Prof. Dr. Stefan Mundlos, who gave me the opportunity to perform my PhD Thesis in his research group at the Max Planck Institute for Molecular Genetics in Berlin. Furthermore, I want to thank Prof. Dr. Eva Klopocki, Dr. Malte Spielmann and Prof. Dr. Sigmar Sticker for their support and Discussion.

I am highly grateful for the strong support of Dr. Darío Lupianez in the 'hot phase' of the thesis and paper writing – I learned a lot from you and I it was a pleasure to work with you!

Thank you Giulia Cova for supporting the project and staying 24/7 with me in the lab to finish genotyping, isolations of whatsoever, prepping mice.. you're a truly reliable and and trustworthy colleague and friend!

Without the following people it would not have been possible to conduct such a great project, and I want to thank you all:

Dr. Hardy (Wing-Li) Chen for the beautiful Micro-CT scans.

Dr. Lars Wittler for managing all the mouse aggregations/animal proposals and helping with every concern a PhD student has.

The complete animal facility of MPIMG that did a great job! Especially, Niclas, Katja, Heike, Corinna, Larissa, Christin and of course Judith Fiedler and Dr. Ludger Hartmann for the great organization! Special thanks to Anne Heß – for all your patience, understanding and hilarious emails, I'll never forget them!

Dr. Bernd Timmermann of the sequencing core facility of MPIMG and Ulrike Krüger and Mohsen Karbasiyan from Charité Universitätsmedizin Berlin, who ran all the sequencing.

Dr. Verena Heinrich of the Department Computational Molecular Biology at MPIMG who provided the pipeline for 4C-Seq data analysis, as well as Dr. Stefan Haas who gave me strong support and assistance in data analysis.

Special thanks goes also to our collaboration partners Prof. Dr. Axel Visel and Dr. Marco Osterwalder for their support with the enhancer-reporter assays and discussion.

Thank you Norbert Brieske for staying calm and performing these wonderful in situs! I am also saluting Asita and Ute for managing the chaos in the lab at all times – you three are truly the best technicians a PhD can wish for!

Thank you to all members of the group (including Charité visitors ;) – it was a great time!

Iza, you always made me laugh with your weird making faces and hidden cat pictures. ^^ see you at the bitter end! ;)

Rieke many thanks for the great time we spent together on festivals, party, having breakfast and doing whatsoever.

Liene and Philine, you are just gorgeous! NEVER stop wearing black ;)

Marina, I always enjoyed our coffee breaks and spending time with you!

Bjort, it was never lame having lunch with you. I love your way of asking all kinds of interesting questions no matter how sensitive they are. You're a true feminist ;)

Martin, thank you a lot for discussion and helping me especially during my first year and for rolling a cigarette with me if I needed one. I really really appreciate you!

Mika, so sad you left before I could finish. You were a great office mate and I loved teaching you weird german words (Eichhörnchen) and when you whished me 'komm gut nach Hause' with you're cute French accent. Stay knacki!

Thank you Aru and Sophie for the fun lunch brakes and Aru for making with me the coolest Valentines Day gift ever!!

Thank you Sinje for they time we spent together at and outside of work.

Thanks to the great STA people for organizing all kinds of events – I know it's hard to make geeky scientists socialize but you made it happen! I'm happily remembering especially our last retreat at Saxony Switzerland and our common lunches. Thank you Edgar, Verena, Nilo, Laura, Bjort and Marina for the fun 90ies party as well!

Thank you Fabian for supporting me during all this time, for your understanding, for cheering me up and for trying to distract me once in a while.. <3

I am especially grateful for my family and friends who supported me during my ups and downs and who were always willing to listen.

Table of Contents

That's the end and that's the start of it
That's the hope and that's the part of it
That's the high and that's the heart of it
That's the long and that's the short of it
That's the best and that's the test in it
That's the doubt, the doubt I trusted in
That's the sight and that's the sound of it
That's the gift and that's the trick in it

Placebo.

1 Summary

Developmental genes, like Indian hedgehog (*IHH*), are commonly regulated by a cluster of enhancers that display redundancy in reporter assays and transcription factor binding sites, properties that are thought to provide developmental robustness. Copy number variations (CNVs) of the non-coding genome often comprise *cis*-regulatory elements such as enhancers and have been reported to be a frequent cause of human disease. Thus, raising the question of how redundancy and dosage respond to systematic variations affecting the number and composition of enhancers within a cluster. In this study, these questions were addressed in an exemplary manner at the *IHH* locus where variable duplications have been associated with highly localized phenotypes, i.e. craniosynostosis, syndactyly and polydactyly. The functional and structural organization of the murine *Ihh* locus was investigated by 4C-Seq analyses in combination with enhancer trap reporter assays. 4C-Seq revealed a limited spread of interaction of the *Ihh* promoter within a surrounding of 250kb, being most prominent in the intron of its upstream neighboring gene *Nhej1*. In comparison with publicly available expression data of *Ihh* neighboring genes, the enhancer-trap assay revealed *Ihh*-specific reporter activation exclusively, suggesting that *Ihh* builds an isolated regulatory sub-domain (*Ihh*-TAD) that is populated by *Ihh*-specific regulatory elements. The combined analysis of 4C-Seq data with publicly available enhancer-associated histone marks, *Ihh*-specific transcription factor binding sites and sequence conservation identified a set of nine putative enhancers. Enhancer-reporter assays confirmed the *in vivo* activity of all nine enhancers, and further revealed that these elements function in a complex manner as they displayed only partial redundancy but also individual specificity among the scored tissues.

To address the functionality of the enhancer cluster and the effects of enhancer dosage on *Ihh* expression, CRISVar was applied to induce deletions and duplications. The *in vivo* functionality of the *Ihh*-TAD was confirmed by deleting the full *Nhej1*-intron that contained eight of the nine identified enhancers. Further, consecutive deletions of the enhancer cluster revealed that the enhancers function in an additive manner, resulting in tissue-specific impairments of *Ihh* expression and skeletal development. By duplicating parts of the enhancer clusters the human malformations were successfully reconstructed in the mouse models. However, not all duplication lines recapitulated the disease features completely, suggesting divergency among the underlying pathomechanisms. Further analysis revealed that cransiosynostosis was induced by a local upregulation of *Ihh* expression in the skull cap, which correlated with the number of skull-specific *Ihh* enhancers, suggesting a dosage effect. In contrast, the limb phenotypes did not relate with increased *Ihh* expression. In syndactyly mutants, ectopic activation of the IHH pathway was observed in the fingertips and the distal interdigital mesenchyme of E13.5 limbs. The expanded IHH signaling precisely overlapped with a local loss of apoptosis in this region, but did not interfere with the signaling pathways that are induced during digit separation, suggesting that a local misexpression of *Ihh* might be sufficient to induce syndactyly. 4C-Seq analyses revealed that this misexpression was most likely caused by the rearrangement of the enhancer cluster, resulting in a unique spatial and regulatory configuration at the boarder of the duplication. This study shows that spatio-temporal

precision of gene expression is controlled by the complexity of multipartite enhancer clusters and that alterations in dosage and composition of individual enhancer elements of these clusters affect the precision of gene expression that can result in disease.

2 Zusammenfassung

Die Expression von Genen, die wie *Indian Hedgehog* (*IHH*) bei der embryonalen Entwicklung eine Schlüsselrolle spielen, werden für gewöhnlich durch mehrere Enhancer reguliert, die sich in Form von Enhancer-Clustern organisieren. Die Enhancer eines Clusters weisen eine gewisse Redundanz in Reporter-Assays und Transkriptionsfaktorbindestellen auf, um eine stabile sowie präzise Expression des Zielgens gewährleisten zu können. Veränderungen in der Kopienzahl (CNVs) nichtkodierender genomischer Sequenzen betreffen meist *cis*-regulatorische Elemente, wie beispielsweise Enhancer. CNVs können vom Genom toleriert werden aber auch zu schweren Krankheitsbildern, wie zum Beispiel Craniosynostose, Syndactylie und Polydaktylie, die durch Duplikationen am *IHH* Lokus verursacht wurden, führen. Daraus ergibt sich die Frage, inwieweit systemische Variationen innerhalb eines Enhancer-Clusters, also Veränderungen in der Anzahl und Komposition einzelner Enhancer-Elemente, durch die redundanten Eigenschaften der übrigen Komponenten ausgeglichen werden können.

Um diese Frage zu klären, wurde die Regulation der Genexpression am *Ihh* Lokus während der Knochenentwicklung mit Hilfe von genetisch veränderten Mausmodellen untersucht. Durch eine vergleichende Analyse mit Circular Chromatin Conformation Capture (4C-Seq) und Enhancer-trap Assays konnte die regulatorische Domäne von *Ihh* bestimmt werden. Diese befindet sich hauptsächlich stromaufwärts von *Ihh*, in dem dritten Intron des Nachbargens *Nhej1*. Der Vergleich, des aus dem Enhancer-trap hervorgehenden Expressionsmusters mit öffentlich zugänglichen Expressiondaten der umliegenden Gene, zeigte, dass die Aktivität dieser regulatorischen Region isoliert agiert. Um einzelne Elemente des Enhancer-Cluster bestimmen zu können, wurden die 4C-Seq Daten sowie Enhancer-assoziierte Histonmarkierungen, Sequenzkonservierung und *Ihh*-typische Transkriptionsfaktorbindestellen analysiert. Insgesamt konnten so neun Enhancer identifiziert werden, deren *in vivo* Aktivität in den folgenden Enhancer-Reporter Assays bestätigt wurde. Entgegen der bislang gängigen Annahme, dass Enhancer redundante Eigenschaften aufweisen, konnte mit Hilfe eines Scoring-Systems gezeigt werden, dass die neun Enhancer vorrangig gewebe-spezifisch agieren. Die Funktionalität des Enhancer-Clusters konnte durch die Deletion des gesamten *Nhej1*-introns, welches acht der neun Enhancer beinhaltet, nachgewiesen werden. Homozygote Tiere dieser Linie zeigten starke Einschränkungen in der Skelettentwicklung, ähnlich dem bereits publizierten *Ihh* knockout Mausmodel.

Die Auswirkungen von veränderter Enhancer-Anzahl und -komposition innerhalb des Clusters wurden mithilfe systematischer Deletionen und Duplikation, die mit der CRISVar Technologie generiert wurden, untersucht. Die systematische Verminderung der Enhancer-Anzahl zeigte eine kontinuierliche Abnahme der *Ihh*-Expression in den jeweiligen Zielgeweben und einen damit einhergehenden verstärkten gewebespezifischen Phänotyp. Durch die Rekonstruktion der humanen Duplikationen, sowie durch die Duplikation des gesamten Enhancer-Clusters, konnten die Krankheitsbilder der Patienten im Mausmodell exakt nachgestellt werden. Expressionanalysen zeigten, dass die Entwicklung von Craniosynostose vermutlich durch einen sogenannten ‚Dosage-Effekt', also eine lokal verstärke Expression von *Ihh*, hervorgerufen durch die erhöhte

Anzahl der Schädel-spezifischen Enhancer, induziert wurde. Die Entwicklung von Syndactyly hingegen wurde ausschließlich in einer der drei Duplikationslinien beobachtet und konnte nicht mit einer erhöhten *Ihh*-Expression erklärt werden. In diesem Fall, führte eine Misexpression von *Ihh*, die von den Fingerspitzen in das umliegende interdigitale Mesenchym übergreift, zur Inaktivierung der interdigitalen Apoptose in dieser Region. Interessanterweise, wurden dabei keine zusätzlichen Signalwege gestört, die während der Separierung der Finger eine Rolle spielen, was darauf hindeutet, dass *Ihh* selbst eine Rolle während der interdigitalen Apoptose spielen könnte. Eine 4C-Seq Analyse von E14.5 Handproben zeigte, dass die Misexpression höchstwahrscheinlich durch eine einzigartige Enhancer-Promoter-Konstellation am Bruchpunkt der Duplikation induziert wurde. In dieser Studie wurde am Beispiel des *Ihh*-Lokus gezeigt, dass die zeitliche und lokale Aktivierung der Genexpression durch die komplexe Zusammenwirkung eines Enhancer-Clusters kontrolliert wird. Systemische Veränderungen in der Komposition und Anzahl innerhalb eines Enhancer-Clusters, können je nach Betroffenheit, verschiedenste Auswirkungen haben und somit zu leichten Fehlentwicklungen aber auch schweren Krankheitsbildern führen.

3 Introduction

3.1 The Organization of the Genome and the Complexity of Gene Regulation

In higher developed organisms, versatile tissues with differential morphological and mechanistic qualities are established through a complex interplay of developmental processes. These processes are regulated by the activation of developmental genes with spatio-temporal precision. The human genome consists of more than 3,000 megabases (Mb). However, only a small fraction of the genome represents protein-coding genes while the remaining sequence is mainly comprised of *cis*-regulatory elements[1].

Cis-regulatory elements are non-coding sequences of the genome with transcriptional regulatory potential that interact with trans-regulators, like transcription factors (TF), activators or repressors, to induce (promoters and enhancers) or repress (silencers and insulators) gene expression[2]. Promoters are non-coding sequences that are located 5' of the gene-coding region and play a central role in transcriptional activation. Promoters are composed of two components: the core promoter and proximal promoter elements. The core promoter is located directly upstream of the transcriptional start site of the gene body and defines the direction of transcription (5' to 3'). By interacting with TFs, the core promoter stabilizes the assembly of the preinitiation complex (PIC) and the binding of RNA-polymerase II (RNA-Pol II)[2–5]. The proximal promoter region is located a few hundred basepairs (bp) upstream of the core promoter. It contains various TF binding sites that stabilize enhancer-promoter contact and PIC assembly. Enhancers are *cis*-regulatory elements that can induce the tissue-specific activation of genes. Enhancers can be located upstream, downstream or within introns of the transcription unit and act independently of their orientation and distance to their target gene[6–8]. Enhancers are highly conserved throughout species and consist of multiple clusters of TF binding sites[2]. Through the binding of tissue-specific TFs, enhancers can stabilize the PIC, and thus enhance the transcriptional activity of a gene with spatio-temporal precision.

Cis-regulatory elements can also confer repressive functions. Silencers are non-coding elements that inhibit transcription initiation through the binding or repressor elements[9–11]. Short-range silencers are located within 100bp of the target gene and might act by contacting the promoter region directly, while long-range silencers are located at a distance of a few kilobases (kb) from the promoter and act indirectly by binding repressor proteins[10,12]. Insulators control transcriptional activation by organizing the genome into defined territories of expression[2,13]. They regulate chromatin condensation through the recruitment of histone modifiers to generate a repressive heterochromatin barrier[2,13,14]. Insulator interactions are mediated by CCCTC-binding factors (CTCF) that promote the physical separation of independent transcriptional units[15–17].

3.1.1 The Regulation of Transcription Activation

Transcription activation is a highly complex and one of the most crucial processes that regulate the expression of protein-coding genes. The dynamic interaction of various *cis*- and trans-acting regulatory elements controls the processes of transcription activation: chromatin remodeling, PIC recruitment, transcription initiation and release of poised RNA-Pol II (Figure 1)[2,18,19].

In the condensed chromatin state enhancer activity is repressed. Signal induced binding of pioneer TFs to the enhancer site results in the recruitment of co-factors to initiate histone modification and to facilitate chromatin accessibility[20]. Co-factor induced histone modifications are modifications of epigenetic marks (methylation/acetylation), histone variant substitution as well as nucleosome depletion[21–25]. Upon chromatin decondensation, enhancers are accessible for the binding of additional tissue- and stimulus-specific TFs. The TF complex recruits mediators that facilitate enhancer-promoter contact through loop formation[26]. The enhancer-promoter contact is stabilized by the interaction of CTCF sites and cohesion complex formation[27,28]. Additional TFs are recruited to induce PIC assembly that guides RNA-Pol II to the TSS of the poised promoter[29,30]. RNA-Pol II is released by P-TEFb-mediated phosphorylation of the repressive factors NELF and DSIF, forming a fully functional elongation complex[31]. As elongation proceeds, a complex of TFs and mediators remains on the core promoter to allow for the re-initiation of transcription by binding of RNA-Pol II, TFIIF and TFIIB repeatedly[32].

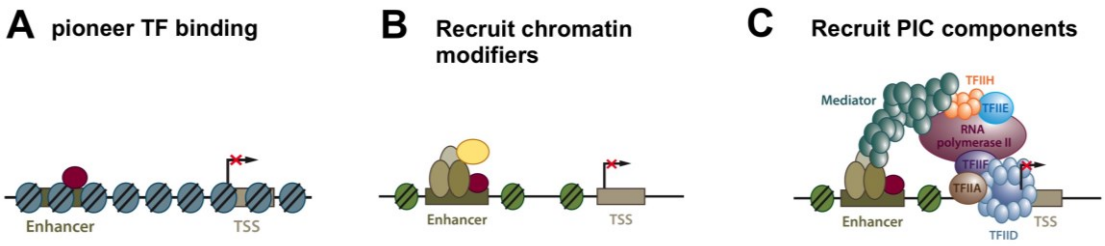

Figure 1. Schematic overview of transcription regulation and induction of PIC assembly

(A) Prior to transcription initiation the chromatin is condensed. Signal induced binding of pioneer TFs (red) to a specific enhancer initiates the activation of a target gene. (B) Chromatin modifiers (yellow) are recruited to facilitate chromatin opening. The enhancer is now accessible for the binding of additional tissue- or stimulus-specific TFs (brown). (C) The complex of TFs recruits mediators that facilitate enhancer-promoter contact and stabilize PIC assembly. Activation of the paused RNA-Pol II is achieved by pause-releasing factors (adapted from Maston *et al* 2012).

3.1.3 Developmental Gene Expression is Regulated by Redundant Enhancers

Gene expression is tightly controlled during development by a class of *cis*-regulatory elements: Enhancers. Genome wide studies conducted in *Drosophila* revealed that 65% of the genes involved in mesoderm development are not regulated by only a single but by multiple redundant enhancers. The majority of these genes are regulated by three to five redundant enhancers, of which one or more can be deleted without significant phenotypic effects[33]. Redundant enhancers, also termed shadow enhancers, were first described for the *Drosophila* genes *brinker* and *sog*[34]. One of the most prominent features of redundant enhancers is their tendency to arrange in clusters, being located proximal or in far distance to their target gene (Figure 2). Further characteristics are their overlapping expression patterns in reporter assays and their similarities in TF binding sites[34–36].

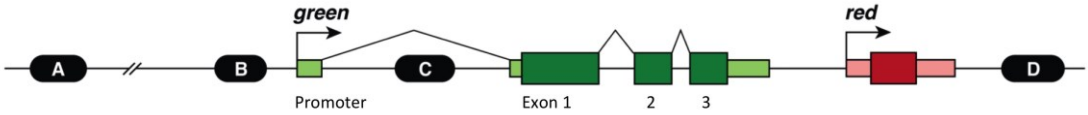

Figure 2. Schematic of a regulatory landscape comprising a cluster of redundant enhancers

The regulatory landscape of gene green is shown as an example of a hypothetical gene locus. It comprises the green gene, which consists of a promoter region and three exons, the red gene (promoter region and one exon) as well as four redundant enhancer elements that are located closely (B and C) or distantly (A and D) to the green gene. The enhancers A and B are upstream of the green gene, while enhancer C is located in the first intron of the green gene. Enhancer D and the green gene are separated from one another by the red gene. All enhancers A-D facilitate normal expression of the green gene, independent of their relative orientation and distance to the promoter of the green gene (adapted from Barolo *et al* 2012).

To date, two classes of redundant enhancers are defined: absolute and partially redundant enhancers[37]. Absolute redundant enhancers are thought to develop through naturally occurring duplications, being either degraded - due to their fully repetitive function - or having evolved to become partial redundant enhancers with overlapping activities. Partial redundant enhancers are thought to show minor differences in their spatio-temporal activity, occur much more frequently and are highly conserved throughout evolution[37–39]. In *Drosophila*, the redundancy of enhancers that control the expression of the developmental genes *shavenbaby* (embryonic epidermis) and *snail* (mesoderm) was addressed through functional deletion studies. These enhancer deletions revealed that the loss of at least one of the enhancer elements did not result in a loss of function phenotype under standardized laboratory conditions[35,40]. However, the importance of these elements in facilitating normal gene expression levels was shown when the genetically modified embryos were treated with high-temperature stress[35,40]. These observations suggest that partial redundant enhancers play a central role in regulating developmental genes by buffering genetic and environmental alterations and by facilitating spatio-temporal precision of gene expression[36,39,41,42].

3.1.4 The Importance of Spatial Folding

Enhancers have been shown to act independently of their distance and orientation to their cognate promoter[43–45]. In order to facilitate the proper functionality of enhancers spatial folding is required to establish enhancer-promoter contact and to induce precise patterns of gene expression[46]. In the folding process, the intervening Deoxyribunucleic acid (DNA) is 'looped out' by the synergistic interaction of various TFs and mediators to build enhancer-promoter contacts. Enhancer-promoter contacts are stabilized by the formation of cohesion complexes that specifically interact with CTCF binding sites, which are located near *cis*- regulatory elements[47].

Enhancer-promoter interactions can be studied on a molecular level using chromosome conformation capture (3C)-based methods[48,49]. With 3C-Seq, interaction frequencies between two known genomic loci (one with one) can be detected through fixing the three-dimensional chromatin structure in living cells, followed by restriction and proximity ligation. Circular chromosome conformation capture (4C-Seq) is frequently applied to study the interactions of one selected locus with the rest of the genome (one with all) by combining the principles of 3C with high-throughout sequencing[50]. For this, the chromatin interactions of intact nuclei are fixed using formaldehyde, which cross-links DNA and proteins, and individual interaction compartments are isolated through a series of digestion and ligation steps. The sequences of the ligated interaction partners are amplified from the viewpoint using inverse PCR primers. The genomic position of the interacting regions is determined by high-throughput sequencing and mapped to the genome. Using 4C-Seq, regions that show high interaction frequencies with a gene promoter can be identified as putative enhancer elements (in combination with other enhancer marks) and selected for functional testing (Figure 3).

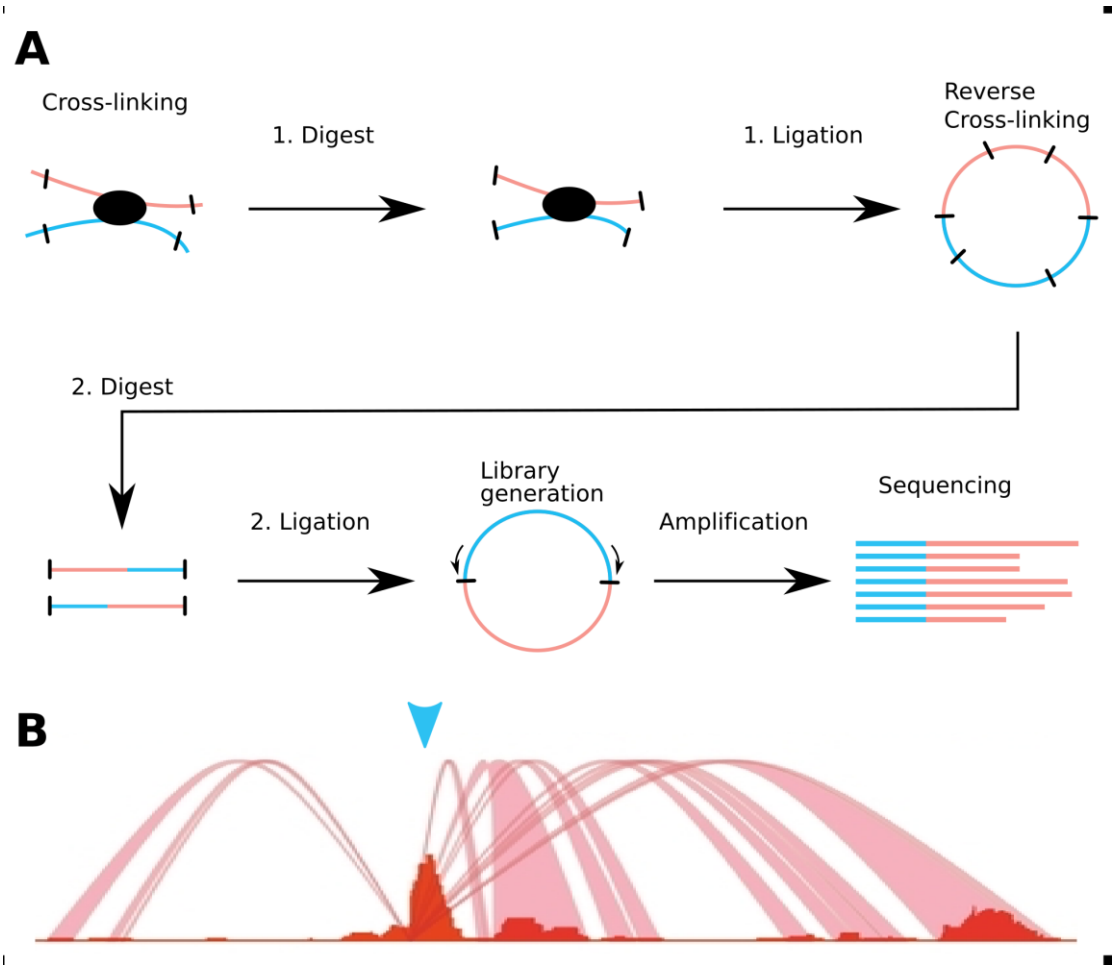

Figure 3. Principles of 4C-Seq

(A) Schematic overview of 4C-Seq: through looping, promoter and enhancer sequences are in close proximity and can be cross-linked using Formaldehyde. Two enzymatic digestion steps using different enzymes are applied. Each digestion is followed by a ligation step. The generated library contains the circularized DNA-fragments of interacting sequences. Inverse PCR from a defined viewpoint (promoter) reveals specific interaction partners. (B) Example of a 4C-Seq interaction profile: PCR fragments are sequenced and mapped to the genome. The resulting profile is a quantitative measure of the relative interaction frequencies of one defined viewpoint (arrowhead) with parts of the genome (enhancers), (adapted from https://www.devcom-itn.eu/page/research/enhancers).

3.1.4.2 Organization of the Genome in Topological Associated Domains

Further derivatives of 3C, like Hi-C, have been developed to conduct genome-wide interaction studies in an unbiased manner (all with all)[50]. The application of Hi-C in mouse and human revealed that the mammalian genome is organized in topological associated domains (TADs). TADs are large domains, which comprise several Mb of DNA that tend to frequently interact with each other but not with the rest of the genome[51–54]. TADs are commonly identified through a combined approach of Hi-C with a hidden Markov model analysis and can be seen as 'triangles' on the heat map (Figure 4)[55]. Neighboring TADs can be distinguished from another by a sudden drop of chromatin interaction. These regions are designated as boundaries. Boundaries are occupied by housekeeping genes and binding sites for various architectural proteins, including Cohesin, Condensin II and CTCF, to stabilize chromosome conformation[56,57]. TADs are highly conserved among different cell types and species indicating their importance in higher order chromatin architecture[55,57–59]. Disruptions of TAD boundaries induced by structural variations have been shown to be associated with human disease. One recent example is a large inter-TAD deletion spanning the boundary between the *EphA4* and *Pax3* locus. This deletion resulted in the fusion of the neighboring TADs and was accompanied by ectopic enhancer-promoter interactions, modified gene expression and pathogenic phenotypes[60]. Inside TADs, smaller compartments of high interaction frequencies (sub-TADs) have been detected using single-cell Hi-C. Sub-TADs are mainly present at gene dense regions and range from several kb to a few Mb[61]. Interestingly, fewer architectural protein binding sites have been detected at the boarders of sub-TADs, which might provide them with higher structural flexibility as compared to the rather rigid TADs. This structural flexibility might be crucial to confer conformational changes that induce tissue-specific enhancer-promoter interactions during developmental processes[47,62,63].

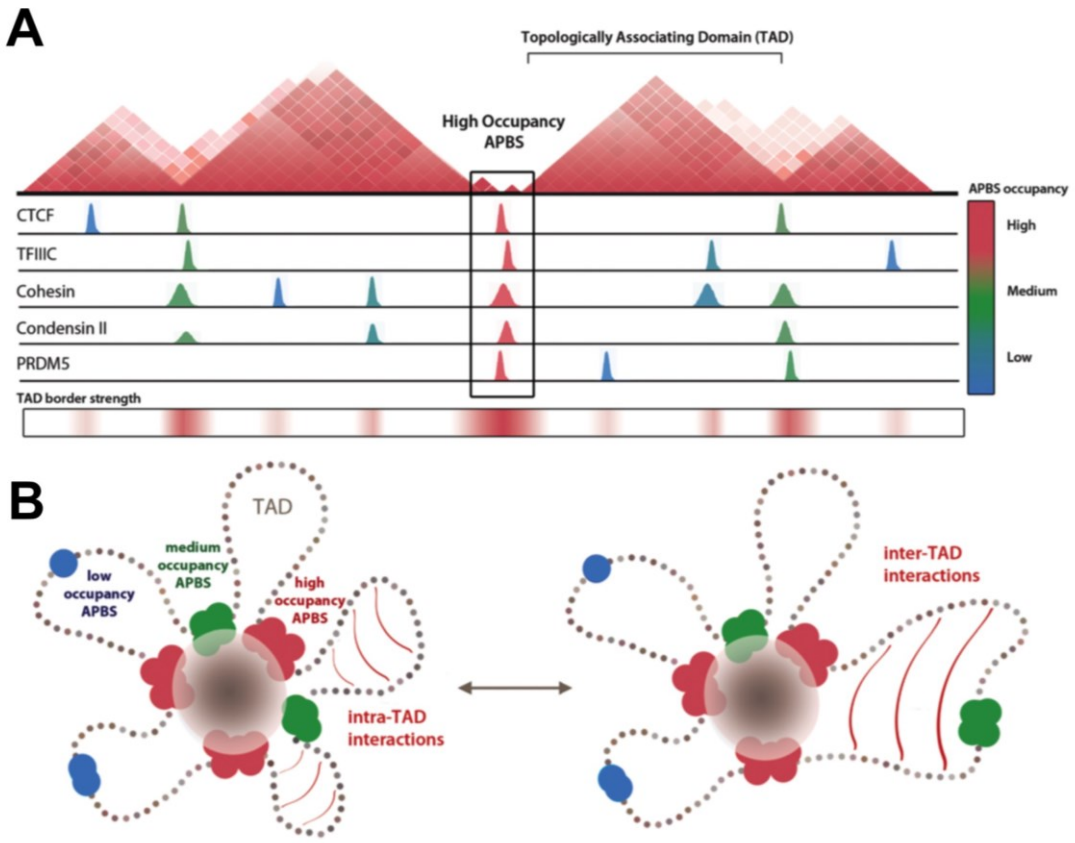

Figure 4. Higher order chromatin structures are stabilized by architectural proteins

(A) Schematic of a Hi-C heat-map: genomic regions with high interaction frequencies (dark red) form TADs. At TAD boarders, the interaction signals drop suddenly, insulating neighboring TADs from each other. At TAD boarders, the formation of clusters of architectural protein binding sites (APBSs) has been observed, which correlate with the interaction frequency: high (red), medium (green), low (blue). (B) Strong boarders (TAD boundaries), (red) with high APBS occupancy are highly conserved throughout different cell types and tissues to stabilize higher order chromatin organization. Weaker boarders (blue and green), which are often localized inside TADs, are more flexible and can facilitate tissue-specific enhance-promoter interactions in between smaller TAD compartments (sub-TADs), (adapted from van Bortle *et al* 2014).

3.1.5 Characterization of Enhancer Activity and Function *in vivo*

In the past decades, a catalogue of methods has been established that is used to identify genomic sequences with enhancer potential. Two of the most common methods for enhancer identification are Chip-Seq and the previously discussed 3C-based technologies. Chip-Seq combines Chromatin-Immunoprecipitation and high throughput sequencing to detect DNA-protein interactions. Using a set of antibodies against RNA-Pol II, co-activators, TFs or enhancer associated histone marks (H3K4me1 and H3K27ac), Chip-Seq reveals genome wide profiles of binding sites that can be used to identify potential enhancer regions in any tissue or cell type[64,65]. The recent development of 3C-based

technologies has advanced our understanding of enhancer-promoter interaction and chromosome folding. In combination with other methods such as Chip-Seq, 3C-based techniques are applied frequently to identify putative enhancers with high precision. To interpret the biological relevance of the predicted regulatory regions, functional transgenic assays have been developed.

Reporter assays are one of the most frequent methods used to characterize the potential of a genomic sequence to drive gene expression. Transgenic reporter constructs commonly contain a reporter gene with a minimal promoter and a site to insert the potential enhancer sequence[6,66]. Different methods have been established to detect reporter transcription directly (RNA *in situ* hybridization) or indirectly (LacZ and fluorescence assays)[67–69]. In mouse embryonic development, enhancer activity is commonly investigated using LacZ reporter assays that can be designed to test the activity of an individual element or a regulatory unit. The enhancer-induced *LacZ* expression enables the processing of the substrate bromochloroindoxyl-Galactosidase (X-Gal) and results in a local color reaction[70].

Enhancer-trap and enhancer-reporter are two of the various transgenic strategies that have been established to evaluate enhancer activity *in vivo* (Figure 5). The enhancer-trap strategy applies transposon-based reporter constructs containing a reporter gene with a minimal promoter to sense the activity of multiple enhancers. The integration of the reporter construct into the genome can be targeted, by homologous recombination, or random, by remobilization of the reporter that is often flanked with transposon sites[71]. By capturing the activity of all enhancers that surround the integration site, the reporter reveals an expression pattern that recapitulates the expression of the target gene. The reporter signal typically reflects the activity of the regulatory domain that often coincides with the TADs, representing an optimal tool to study the organization of complex long-range regulatory domains[55,72]. However, the signal might be biased by the position of the integration site. Therefore, the reporter may have to be integrated at various positions to provide a complete picture of the regulatory potential of a TAD[45,73]. As the enhancer trap system does not provide information of activity or position of individual enhancers it needs to be combined with complementary transgenic strategies.

Enhancer-reporter assays are commonly used to evaluate the activity of an individual element. For this, the enhancer sequence is cloned in a LacZ reporter vector that is enabled to integrate in the genome[68,74,75]. The reporter signal reflects the endogenous enhancer activity and recapitulates a part of the expression pattern of the target gene[76,77]. A great advantage of this system is that it allows the testing of one enhancer at a time and further to compare expression patterns of different enhancers of a cluster. However, the interpretation is hampered by the fact that the reporter construct integrates in the genome randomly and multiple times. Thus, the reporter signal might be influenced by the genomic context[78]. Therefore, the reproducibility of the expression signal is crucial to characterize the activity of putative enhancer elements. However, this system does not provide quantitative information of enhancer activities, which shall be addressed in a functional assay.

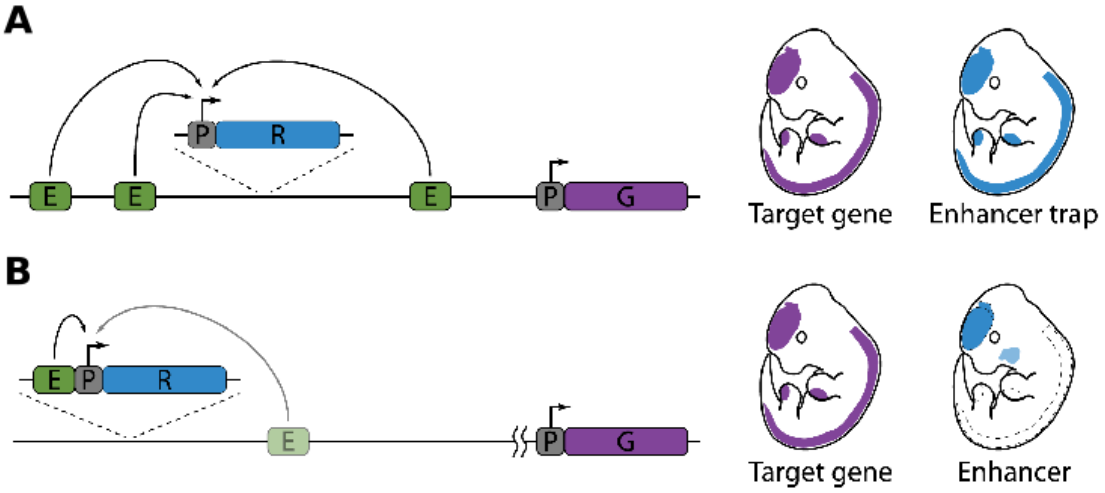

Figure 5. Schematic overview of different transgenic strategies for enhancer activity characterization

(A) Enhancer-trap is a sensor construct containing a reporter gene (blue box) with a minimal promoter that can be integrated throughout the genome by transposition. The reporter senses the activity of all enhancers (green boxes) within the regulatory domain at which the reporter construct is integrated. The reporter expression pattern (blue) recapitulates the expression pattern of the target gene (purple) that is naturally regulated by these enhancers. (B) The enhancer-reporter system contains the enhancer sequence that is inserted directly upstream of the reporter gene and a minimal promoter. The system is randomly integrated in the genome. The reporter expression is activated by the enhancer sequence and recapitulates only parts of the expression pattern of the target gene. Additional expression domains, resulting from the other regulatory elements (faded green boxes) located in the genomic context of the integration site, might be detected (faded blue domain), (adapted from Kvon 2015).

The generation of genetically modified model organisms strongly contributes to the understanding of the complex molecular mechanisms of the human body. The functional characterization of the genome is commonly conducted through knock out (k.o.) constructs, induced point mutations and overexpression of genes as well as deletion studies of *cis*-regulatory elements[79–81]. The mouse is one of the most frequently used model organisms to address regulatory functions, disease mechanisms and drug development in human disease due to its anatomical similarities and its well-annotated genome. The low maintenance costs and short generation times of mice additionally favor it besides other model organisms[82,83].

Various systems have been developed in the past decades to induce targeted genomic rearrangements in the mouse genome that make use of the cellular repair mechanisms homology directed repair (HDR) and non-homologous end joining (NHEJ). Both mechanisms are activated upon double strand breaks, which can occur naturally, e.g. by replication errors and strand exchange in meiosis, or are induced artificially, e.g. by radiation, UV exposure and genetically engineered nucleases[84,85]. HDR provides a template-based system to repair damaged DNA strands with high accuracy. In HDR, the broken DNA strand is processed by DNA nucleases to generate a 3' overhang, which invades and pairs with a homologous template by complementary base pairing. Upon DNA synthesis and ligation the repaired DNA strand is released. With homologous

recombination, the cellular HDR is applied to insert a donor sequence in the genome by providing a linear template. The template typically consists of the donor sequence that is flanked by two DNA stretches (homologous arms) each with a 3-5kb homology to the target site. Deletions can be generated using a template that contains no donor sequence but two homologous arms that align with the upstream and downstream sequence at, but not with the target region[85]. HDR provides high precision to induce targeted genomic modifications, but is limited to the G2- and S-phase of the cell cycle as well as by the availability and size of a homologous template[86,87]. In NHEJ repair, double strand breaks are repaired by ligation of open DNA fragments, without synthesis from a donor template. The broken ends are either processed to generate short homology sequences to facilitate complementary base pairing at the breakpoint or ligated directly. NHEJ repair is more error prone compared to HDR and might result in small insertion and/or deletions at the breakpoint. Thus, it is frequently used and limited to induce point mutations and frameshifts of the coding sequence genes or in TF binding sites of *cis*-regulatory elements. NHEJ repair is not restricted to a certain cell cycle and thus occurs with a higher frequency than HDR[86,88–90].

Complex rearrangements of the non-coding genome might affect the functionality of enhancer clusters by inducing alterations in enhancer number, composition and relative position to their target gene. Different systems have been developed that are frequently applied to reconstruct structural variations in model organisms.

The **TAMERE** (trans allelic targeted meiotic recombination) system applies the principles of Cre/loxP (cyclization recombination). The Cre/*loxP* system originates from the bacteriophage P1 and combines two components: the Cre recombinase protein and the *loxP* site[91]. The *loxP* site is a genomic sequence that is specifically recognized by Cre. It consists of two inverted repeats (13bp) and one central spacer sequence (8bp), which is directed and therefore facilitates site-specific recombination in the presence of two *loxP* sites depending on their orientation to each other. In *cis*, Cre mediates the inversion (head-to-head orientation) or the excision/insertion (head-to-tail) of *loxP*-flanked sequences[92–94]. The Cre/loxP system also facilitates trans-allelic recombination, a feature that is commonly applied to induce structural variations in mice (and other model organisms). For this, three transgenic alleles need to be engineered: two of which contain a *loxP* site upstream or downstream of the target region and one with a *Cre* transgene. Triple transgenic mice with two different insertion sites and a *Cre* transgene are achieved through several mating steps (Figure 6). The Cre-recombinase is designed to be active in the germline, thus recombination events occur in gametes only. By mating the triple transgenic animals to wildtype (wt) mice, new chromosomal rearrangements of the intervening region are obtained in the offspring, resulting in one deletion and one duplication allele. However, this system is limited by the time and effort that need to be invested for the generation of the genomic rearrangements[71,95].

Recently, a new technology for precise and efficient genome engineering has been discovered: **CRISPR/Cas** (clustered regulatory interspaced palindromic repeats/CRISPR associated). The CRISPR/Cas system emanates from the adaptive immune system of prokaryotes and archaea that facilitates the detection and destruction

of foreign nucleic acids. The system is composed of three components: the CRISPR RNA (crRNA), the trans-activating crRNA (tracrRNA) and the Cas endonuclease. Both RNA components are non-coding and form a complex with Cas that is guided to the target region by crRNAs-specific binding[96–99]. The Cas9 endonuclease of *Streptococcus pyogenes* has been optomized to facilitate efficient genome engineering *in vitro*[100]. It forms a ribonucleoprotein complex with a synthetic single guide RNA (sgRNA) that combines the two RNA components (crRNA and tracrRNA). The sgRNA facilitates site-specific targeting to a 20-bp genomic sequence (protospacer) and cleavage at the protospacer adjactent motif (PAM) 'NGG' that is specific for Cas9 (Figure 6). Unlike other nucleases, Cas9 can be easily re-programmed for every application by designing a new sgRNA template. The sgRNA sequence is simply ordered as an oligonucleotide and cloned in a vector system that contains the *Cas9* transgene. At the target site, Cas9 induces a double strand break that is repaired by one of the cellular repair mechanisms (HDR or NHEJ). By providing an extrachromosomal donor DNA with homology to the target region, this system can be used to generate targeted insertions of the donor DNA, as the Cas9-induced double strand break activates the cellular HDR[85,101]. Our group has shown that with the simultaneous application of two sgRNAs deletions, duplications and inversion can be induced in mESC on a kb/Mb scale in one experiment (Figure 6), referred to as CRISVar (CRISPR/Cas-induced structural variants)[102]. Including the results of this work, we have presented a 10-week protocol for the generation of structural variations in mouse lines, a process that takes more than one year with conventional systems like TAMERE, which requires the generation of multiple transgenic alleles and several mating steps. With its ease of application, precision and efficiency CRISVar represents a strong improvement in studying disease-associated structural variations and contributes substantially to gain a better understanding of human disease.

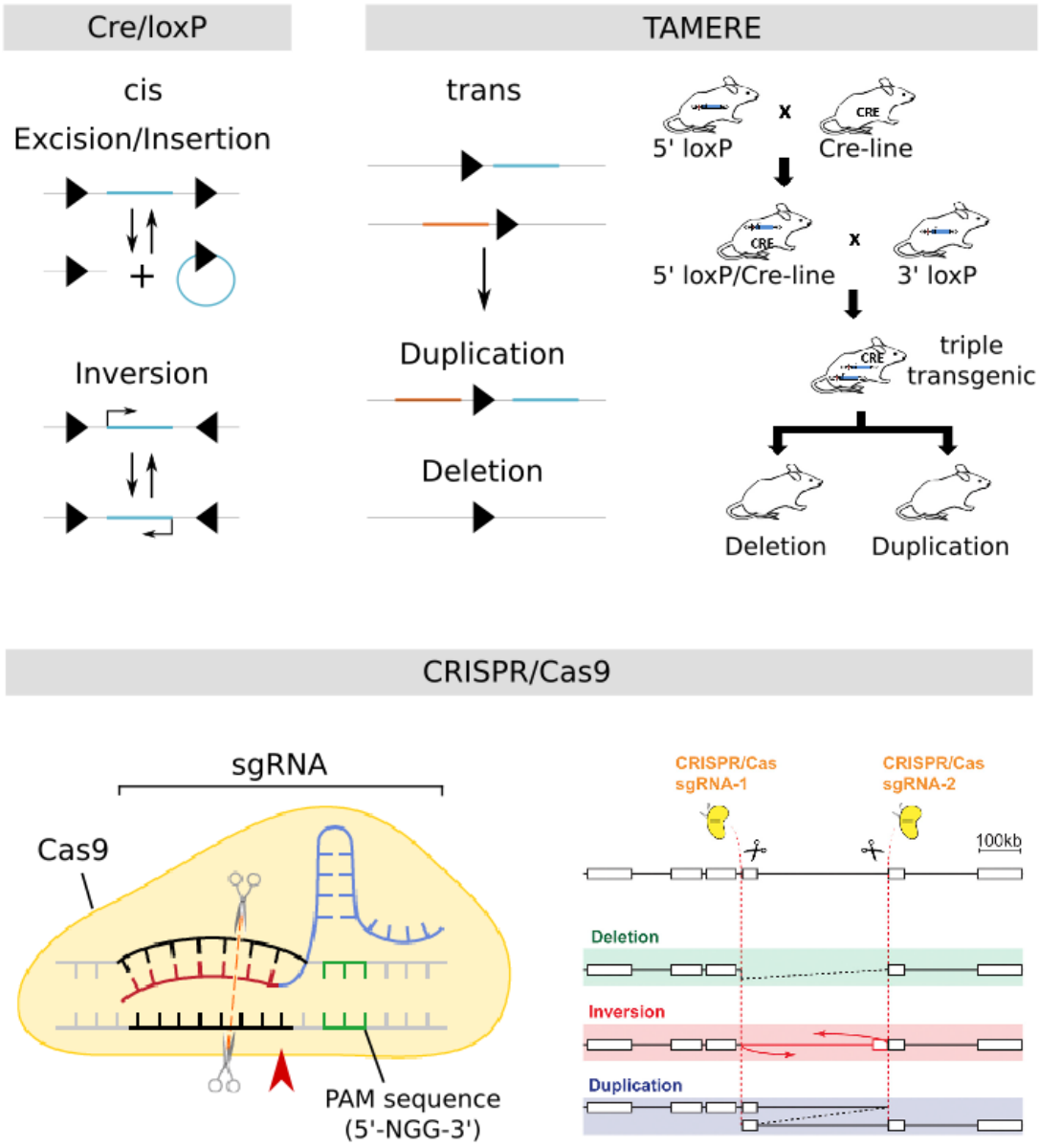

Figure 6. Genome engineering tools

Top: Cre/*loxP* recombination can act in *cis* and trans. In *cis*, Cre-mediated recombination results in the excision or insertion of the *loxP*-flanked sequence if *loxP* sites are oriented in head-to-tail, or in inversion if oriented in head-to-head. TAMERE: trans-allelic recombination mediated by Cre/loxP. Recombination of head-to-tail oriented *loxP* sites induces reciprocal deletion and duplication between two alleles. In mice, several crossing steps are conducted to achieve triple transgenic animals with a *loxP* site on both alleles (5' or 3' of the target site) and a Cre transgene. Chromosomal rearrangements of the region between the two *loxP* sites are obtained in the offspring, when triple transgenic animals are mated with wt mice. Bottom: The CRISPR/Cas9 system is composed of a chimeric sgRNA and the Cas9 endonuclease. The crRNA component of the sgRNA (red) aligns specifically with the target sequence in the genome while the tracrRNA component forms a hairpin to facilitate Cas9 complex formation. Cas9-mediated recombination induces a double strand break adjacent to the Cas9 specific PAM sequence (red arrow head). Applying two sgRNAs targeted deletions, duplications and inversions can be induced in the genome (adapted from Kraft *et al* 2015 and www.clontech.com).

3.1.6 Importance of Non-Coding CNVs in Development and Disease

The Encyclopedia of DNA Elements (ENCODE) consortium has significantly improved the knowledge of genome organization by systematically investigating its non-coding part. ENCODE has revealed that 80% of the human genome are represented by *cis*-regulatory elements and regions encoding regulatory RNAs[103–105]. Structural variations of the genome, i.e. balanced genomic rearrangements and unbalanced copy number variations (CNVs), affecting the *cis*-regulatory landscape are important to provide genetic diversity but also induce pathogenic effects that result in disease[106–108]. Structural variations can be detected through various diagnostic approaches, e.g. microarray-based comparative genomic hybridization and whole genome sequencing, and are catalogued in the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER)[109].

CNVs (deletions, duplications or insertions) of non-coding *cis*-regulatory elements have been frequently linked to human disease. Depending on the modification and the class of regulatory element that is affected, CNVs can result in versatile phenotypic effects.

Deletions of *cis*-regulatory elements have been observed to result in the elimination of the regulatory function of the affected element. Deletions of transcriptional activators, like enhancers, have been reported to result in tissue specific loss of expression. At the *PITX2* locus, enhancer deletions have been associated with Axenfeld-Rieger syndrome that has previously been linked to mutations in the *PITX2* gene[110]. At the *FOXL2* locus, a small 280kb-distant enhancer deletion could be linked to blepharophimosis-ptosis-epicanthus inversus syndrome. Using 3C-Seq, D'haene *et al* showed that the tissue-specific loss of *FOXL2* expression caused by this enhancer deletion was the result of lost enhancer-promoter contact[111]. Deletions of transcriptional repressors, however, can result in gain of function of nearby genes, an effect that has been observed at the *FOXG1* locus. Deletions at this locus resulted in upregulation of *FOXG1* expression in the fibroblasts of patients with Rett syndrome. Allou *et al* suggested that the deletions eliminated a long-range silencer[112].

Duplications and insertion can increase the copy number of *cis*-regulatory elements and might thus result in accelerated regulatory activity of the affected element. Insertions and micro-duplications often affect only individual elements, as this has been observed at the *bone morphigenic protein (BMP) 2* locus at which small duplications of less than 10kb have been linked to brachydactyly type A2 induced by *BMP2* misexpression[113]. In contrast, duplications of several hundred kb often comprise several regulatory elements (activators and repressors) and can also span boundaries of neighboring TADs, making it difficult to interpret the impact of the duplicated elements on gene expression. At the *sonic hedgehog* (*SHH)* locus, duplications of several hundred kb comprising the limb-specific *SHH* enhancer ZPA regulatory sequence (ZRS) have been linked to polysyndactyly or triphalangeal thumb-polysyndactyly syndrome[114–116]. The functional importance of the duplicated elements and how their duplication can induce a similar phenotype that is also caused by point mutations in the ZRS remains elusive. At the *SRY-box9* (*SOX9)* locus, various duplications have been reported to induce diverse human disease phenotypes, reflecting the complexity of *cis*-regulatory interactions. Intra-TAD duplications

constraining the overall TAD structure resulted in increased enhancer-promoter contacts of the duplicated regions in fibroblasts of patients with female-to-male sex reversal[117,118]. In contrast, duplications spanning the *SOX9*- and the neighboring *KCNJ2*-TAD resulted in the formation of a new TAD. The gained ectopic interactions of the duplicated *KCNJ2* gene and *SOX9* regulators resulted in consecutive misexpression of *KCNJ2* and Cooks syndrome in mice[118,119].

3.2 The Skeletal System as a Model to Investigate the Functionality of Enhancer Clusters at the *IHH* Locus

In human, skeletal development is tightly regulated by the spatio-temporal activation of various genes. Perturbations of these developmental pathways show a broad range of phenotypic effects, ranging from mild - in which only a single phalange is affected (Brachydactyly)[120] - to highly severe malformations (Split-Hand-Foot/Nievergelt Syndrome)[121]. In the past decades, human geneticists have made substantial progress in detecting variations of the human genome being associated with disease and skeletal malformation. Besides mutations affecting the gene body itself, they found that CNVs of non-coding regions relate to human disease, induced by pathomechanisms that are poorly understood. Intensive searching revealed that the non-coding part of the genome comprises various *cis*-regulatory elements such as enhancers that have been shown to control gene expression. Developmental genes, like Indian hedgehog (*IHH)*, are often regulated by a cluster of redundant enhancers that are highly conserved throughout different species. In skeletal development, IHH is a key factor of endochondral ossification as it regulates the process of chondrocyte differentiation. Variable duplications directly upstream of *IHH* have been associated with skeletal (Craniosynostosis, Polydactyly) as well as non-skeletal malformations (cutaneous Syndactyly), suggesting the presence of several redundant enhancers in this region. Therefore, the *IHH* locus represents a suitable model to investigate the *in vivo* functionality of enhancer clusters.

3.2.1 Intramembranous and Endochondral Ossification

The skeletal system develops through two distinct processes: intramembranous ossification or endochondral ossification.

Intramembranous ossification takes place in the flat bones of the skull, like the frontal and parietal cranial structures, the mandible and the maxilla[122]. In intramembranous ossification, neural crest-derived mesenchyme cells condense and give rise to osteoblasts directly, induced by the activation of the transcription factors CBFA1/Runx2 and Osterix[123–126]. Osteoblasts produce and secrete a collagen-proteoglycan matrix, which forms the osteoid and is subsequently calcified. In this process some osteoblasts are embedded within the bone matrix and evolve into osteocytes (bone cells) forming a first

layer of bone that is surrounded by compact mesenchyme cells (periosteum). At the inner side of the periosteum, osteoblasts derive and deposit osteoid matrix to form several layers of bone. The plates of the skull are formed by several ossification centers that are separated by the fontanels to allow for proper growth of the developing brain. In human, the ossification of the fontanels is carried out gradually, terminating at about two years after birth with the formation of sutures that fuse the skull plates[127–129].

Endochondral ossification takes place in the long and short bones of the body. It is initiated during embryonic development and continues up to and including early adulthood. In endochondral ossification, a cartilage model of the prospective bone is formed first and then replaced by bone tissue (Figure 7). Endochondral ossification is initiated by the expression of *Sox9* by condensed mesenchymal cells, which are located in the center of the prospective bone[130–132]. The condensed mesenchymal cells differentiate into chondrocytes, which secrete components of cartilage extracellular matrix (ECM), Collagen type II α1 (Col2a1) and Aggrecan, expanding the cartilage model of the prospective bone[133]. In the surrounding of the mid shaft the perichondrium is formed. The perichondrium mediates the shape of the prospective bone by compacting and organizing the proliferating chondrocytes along the longitudinal axis. At the center of the shaft chondrocytes stop proliferating and transition to a pre-hypertrophic state expressing *Ihh*. Further maturation into hypertrophic chondrocytes and expression of *Col10a1* facilitates mineralization by calcium carbonate. Apoptosis of hypertrophic chondrocytes allows for the intrusion of blood vessels and osteoclasts, which degrade apoptotic chondrocytes and cartilage matrix. Osteoblasts invade from the bone collar and proceed to form a primary center of ossification in the diaphysis by producing trabecular bone matrix. The primary center expands toward the ends of the cartilage model while osteoclasts remove cartilage ECM and osteoblasts deposit bone. The secondary center of ossification is formed in the epiphysis at both ends of the cartilage model. The cartilaginous growth plate persists in the diaphysis between the primary and secondary center of ossification to facilitate longitudinal growth of the bone. In the mature skeletal bone the primary center expands and fuses with the secondary centers of ossification.

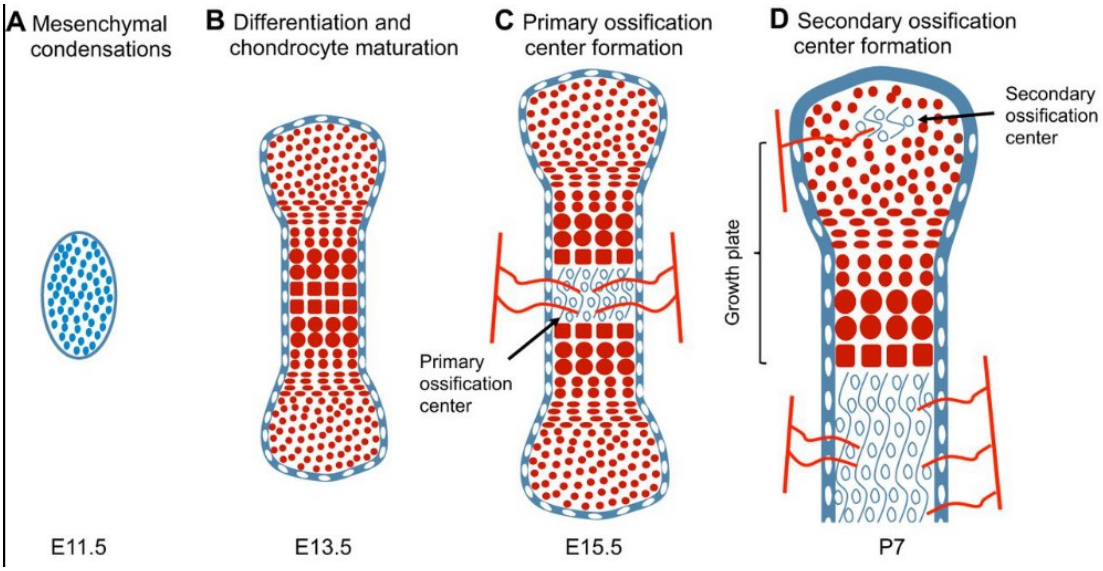

Figure 7. Schematic overview of endochondral ossification in the developing mouse limb

(A) E11.5: The cartilage of the limb is formed by condensation of mesenchymal cells (blue). (B) E13.5: Mesenchyme cells differentiate into proliferating chondrocytes (red cells) to form and expand the cartilage model of the future bone. The cartilage model is surrounded by the perichondrium (white cells). In the center, chondrocytes become hypertrophic (square cells) and mineralization is initiated. (C) E15.5: Vascularization (red lines) at the center of the cartilage anlage facilitates the invasion of osteoclasts and osteoblasts, resulting in the replacement of chondrocytes with trabecular bone (open circles) in the primary center of ossification. (D) P7: the primary center of ossification expands towards the secondary center of ossification, which forms postnatally, (adapted from Kozhemyakina *et al* 2015).

3.2.2 Indian hedgehog – a Key Regulator in Endochondral Ossification

IHH is a member of the hedgehog protein family that is conserved throughout vertebrates. It acts through specific binding to the transmembrane protein Patched1 (PTC1) facilitating the proteolytic processing of the downstream zinc-finger transcription factors Gli1, 2 and 3[134]. IHH plays a major role in endochondral ossification, by regulating chondrocyte proliferation and differentiation via a negative feedback loop (Fig. 2)[135]. IHH is initially expressed in and secreted by pre-hypertrophic chondrocytes at the center of the bone shaft and diffuses to the perichondrium where it binds to PTC1[135,136]. In the epiphyseal center, IHH induces the differentiation of reserve chondrocytes into proliferating chondrocytes and stimulates the expression of *parathyroid hormone-related peptide* (*PTHrP)* in the periarticular perichondrium (Figure 8)[137,138]. PTHrP diffuses to the growth plate and activates PTHrP receptor, which is expressed by proliferative and pre-hypertrophic chondrocytes. PTHrP receptor acts as an inhibitor of chondrocyte differentiation keeping chondrocytes in a proliferative state and reducing *Ihh* expression. This negative feedback loop controls the process of chondrocyte maturation. It is attenuated as chondrocytes become hypertrophic and do no longer express *Ihh*[135,139,140]. Furthermore, IHH coordinates FGF, BMP and RUNX2 signaling, taking a pivotal role in chondrocyte maturation and osteoblast differentiation (Figure 8)[141,142].

Alterations in IHH signaling early in embryonic development result in skeletal malformation. In human, mutations in the *IHH* gene have been shown to be associated with brachydactyly type A1 (malformation of the phalanges) and acrocapitofemoral dysplasia (cone-shaped epiphyses, short stature, and brachydactyly)[143]. In mice, *Ihh* k.o. results in strong deficiencies of skeletal structures caused by abnormal and accelerated induction of hypertrophic differentiation[120,140,144].

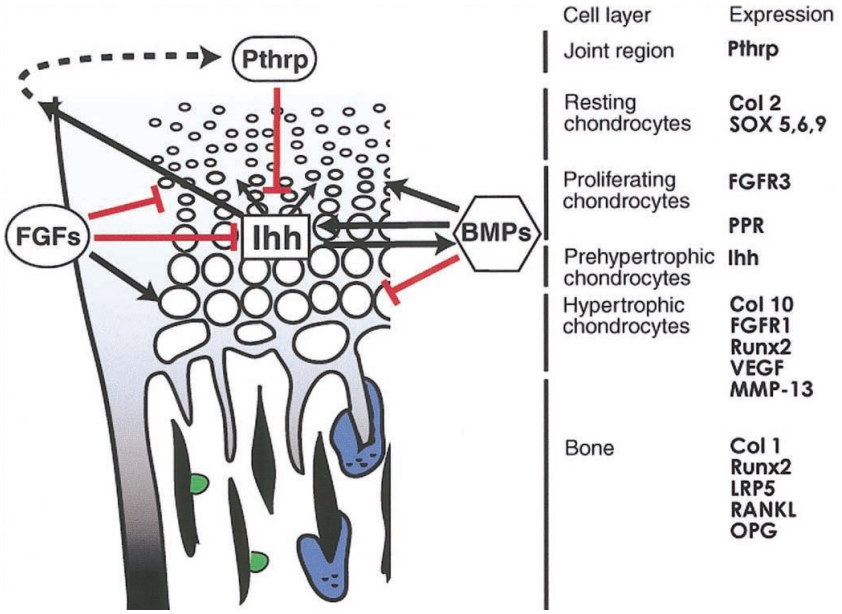

Figure 8. Chondrocyte maturation in the growth plate

The IHH/PTHrP negative feedback loop regulates the chondrocyte maturation: *Ihh* is expressed by pre-hypertrophic chondrocytes and induces *Pthrp* expression in the joint region. Diffusion of PTHrP to the pre-hypertrophic zone inhibits *Ihh* expression and thus the differentiation of proliferating chondrocytes. FGF signaling promotes hypertrophic differentiation by antagonizing IHH signaling. Chondrocyte proliferation is promoted by positive regulation of BMP and IHH. Fully mature Chondrocytes are replaced by bone through osteoblasts (green) and osteoclast (blue). The distinct layers of chondrocytes, representing the differentiation state and the characteristic genes expressed are shown at the right (Kornak and Mundlos 2003).

3.2.3 Limb Development

The development of the extremities is initiated by the formation of limb buds. The limb bud is formed of a highly condensed accumulation of proliferating mesenchyme cells originating from the lateral plate mesoderm. Mesenchymal cells of the limb bud possess the potential to give rise to skeletal structures, tendons and connective tissue, while myogenic precursor cells invading from the somites will give rise to muscle, nerves and blood vessels[145–147].

The limb bud is organized in three primary axes: the proximal-distal, the anterior-posterior and the dorsal-ventral axis[148,149]. Various genes and signaling cascades along the three axes take part in the organization of the limb bud as well as of its outgrowth and

patterning. The most important signaling centers are the apical ectodermal ridge (AER), the progress zone (PZ) and the zone of polarizing activity (ZPA)[147,149,150].

The linear outgrowth of the limb bud takes place along the **proximal-distal axis**, giving rise to the structural anlagen. The stylopod (humerus) is the most proximal element of the limb skeleton, followed by the central zeugopod (radius/ulna) and the distal autopod (carpals/metacarpals/phalanges), (Figure 9)[149–151]. The distal end of the limb bud is terminated by the AER. The formation of the AER along the dorsal-ventral junction is induced in E10.0 mouse embryos by mesenchyme FGF10 signaling and has been shown to be crucial, as limb development is inhibited if the AER fails to form[152,153]. A positive feedback loop of AER/FGF8 and mesenchyme/FGF10 signaling is established to maintain the AER and to form the PZ. In the PZ, mesenchyme cells reside in an undifferentiated strongly proliferating state to facilitate distal limb bud outgrowth[149,150,153]. A changing pattern of *Hox* gene expression along the proximal-distal axis defines the skeletal structures[154–156].

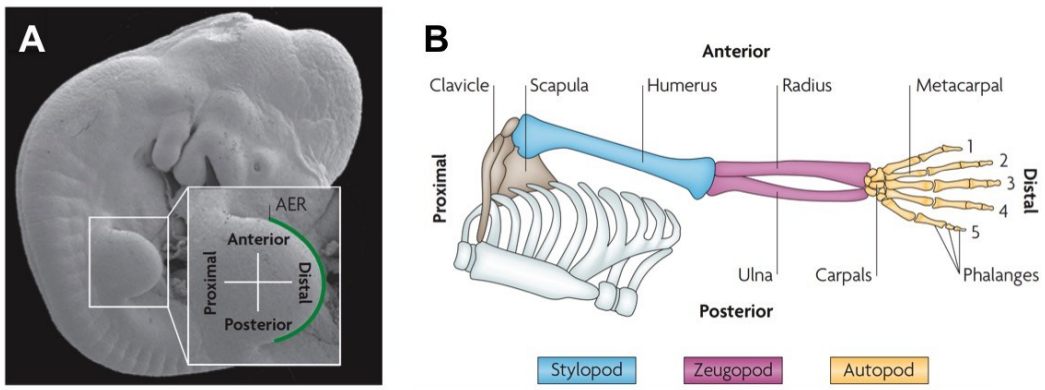

Figure 9. Schematic overview of the proximal-distal limb bud axis development

(A) The electron microscopy image of an E10.5 mouse embryo shows the forelimb bud with the two main axes (AER is indicated in green). (B) The skeletal elements of a human arm are formed along the proximal-distal axis: the stylopod (proximal), zeugopod (central) and autopod (distal); and give rise to humerus, radius/ulna (anterior/posterior) and hand, respectively. The hand is comprised of the carpals (wrist), the metacarpals (palm) and the phalanges (digits), (Zeller *et al* 2009).

The structural identity along the **anterior-posterior axis** is regulated by the ZPA. The ZPA is positioned at the posterior center of the limb bud and characterized by local expression of *Shh*. *Shh* expression is induced and maintained by various factors including FGF8, TBX, HAND2 and Retinoic acid (RA), (Figure 10)[150,157–159]. A negative feedback loop, mediated by the repressor GLI3, restricts *Shh* expression to the posterior side of the limb bud[162,163]. SHH is a diffusible morphogen that signals gradually across the limb bud and with this specifies the identity of mesenchymal cells[147,164]. Paracrine SHH signaling defines the identity of digits 4 and 5, as well as partially of digit 3, all of which have expressed *Shh* previously. However, the identity of digit 2 is entirely established by long-range SHH signaling. The identity of digit 1 is regulated independently of SHH

(Figure 10)[165,166]. Disruptions of the SHH gradient have been shown to interfere with digit number and identity. The application of SHH recombinant protein to the anterior side of the limb bud induces the formation of additional digits[164]. In contrast, *Shh* k.o. mice show loss of anterior-posterior identity and absence of distal limb structures[150,151,167,168].

The polarity of **the dorsal-ventral** axis is regulated by various factors that are specifically expressed in the dorsal or ventral ectoderm (Figure 10). In the dorsal ectoderm *Wnt7a* is expressed and induces dorsal specific gene targets by the activation of *Lmx1b* transcription factor in the underlying mesenchyme. In the ventral ectoderm, *Bmp* expression induces Engrailed1, which inhibits *Wnt7a* expression[169,170].

All three axes of the limb bud are interactive and mutually maintain their overall functionality during limb development. The AER and the ZPA stabilize each other beyond axes via a positive epithelial-mesenchymal (e-m) feedback loop[171,172]. The e-m feedback loop represents a self-regulatory system in which differential expression of *Bmp4* and its antagonist *Gremlin1* (*Grem1*) transits the three phases of limb organogenesis: initiation, propagation and termination. During initiation, high BMP4 activity induces AER function and initiates the expression of *Grem1* in close proximity of the ZPA. As limb organogenesis propagates, GREM1 decreases BMP4 activity and transmits SHH signaling towards the AER, resulting in FGF upregulation and limb bud outgrowth[173,174]. The termination of the e-m feedback loop is initiated by increasing distances between the SHH and GREM1 signaling centers, and terminated by FGF8-mediated inhibition of GREM1[175,176].

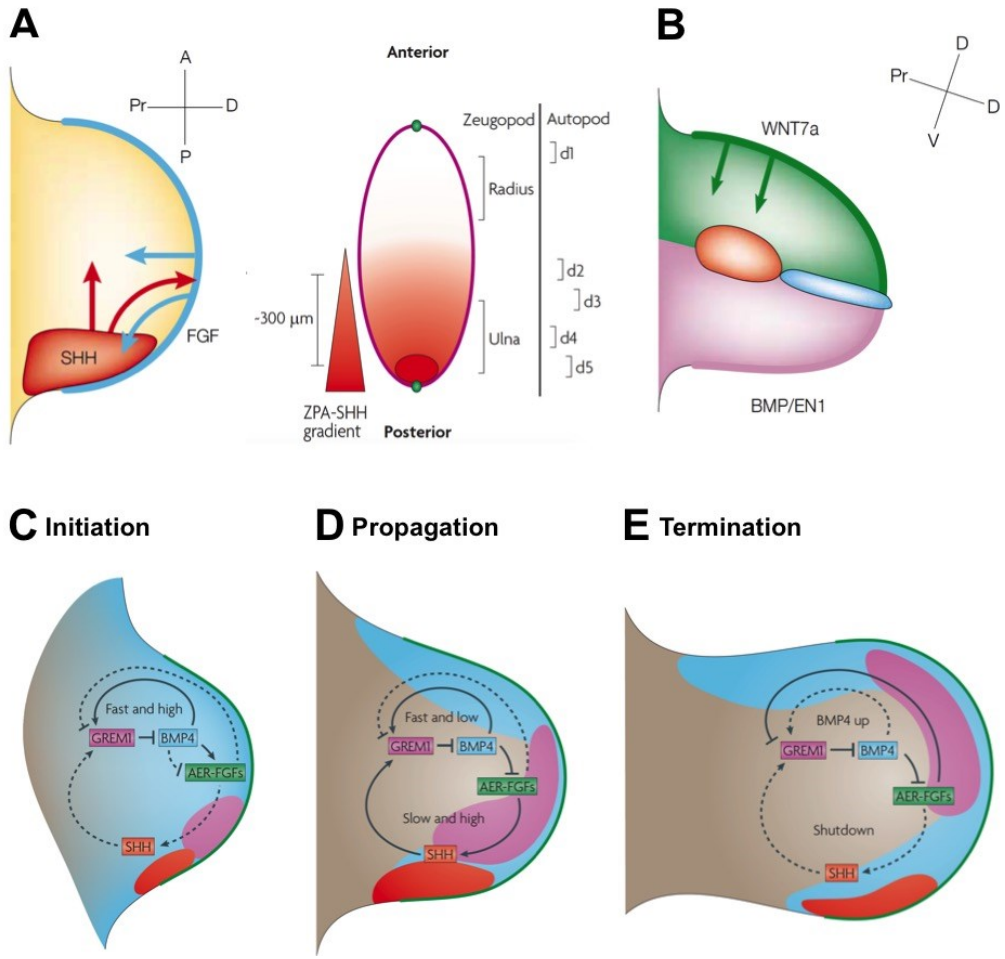

Figure 10. Interaction of signaling pathways controlling limb development

(A) Anterior-posterior axis: *Shh* (red) is spatially expressed in the ZPA at the posterior side of the limb bud and signals gradually to specify distal and posterior identities. The most anterior features (radius and digit1, d1) develop independent of SHH signaling, while the identity of the central and posterior structures (ulna and d2-5) are defined by temporal and spatial SHH signaling. The positive feedback loop of AER-FGFs (blue) and ZPA-SHH (red) mediates proliferation. (B) Dorsal-ventral axis: The patterning of the dorsal and ventral features is regulated by ectodermal signaling of BMP/EN1 (pink) and WNT7a (green), respectively. Axis separation is facilitated by the AER, which is formed along the dorsal-ventral junction. (C-E) The interlinked SHH/GREM1/FGF e-m feedback loop is required to facilitate limb bud outgrowth and early specification of limb identities. (C) Initiation: High BMP activity initiates AER function and induces the upregulation of GREM1. (D) Propagation: BMP4 activity is decreased by GREM1, which enables SHH/GREM1/FGF e-m signaling and distal limb bud outgrowth. (E) Termination: The e-m feedback loop is terminated by FGF-mediated inhibition of GREM1 and by increased distances between GREM1 and SHH signaling centers (adapted from Niswander *et al* 2003 and Zeller *et al* 2009).

3.2.5 Interdigital Cell Death

The separation of the individual digits is facilitated by interdigital cell death (ICD). ICD occurs mostly by apoptosis, which is - besides autophagy and lysosomal cell death - one type of programmed cell death[177]. In vertebrates, three zones of apoptosis are defined during limb development: The opatic patch, which separates the zeugopods structures, and the anterior and the posterior necrotic zone (ANZ, PNZ) that induce ICD to separate the digits[178,179]. ICD is coordinated by the dynamic interaction of three key regulators: RA, FGF8 and BMPs (BMP2, 4 and 7)[180] (Figure 11).

RA induces cell death in a broad extent and must be regulated tightly, as extensive RA expression results in truncated limbs in mice[181,182]. Prior to ICD induction, RA synthesis is mediated by RALDH2 in the proximal region of the interdigital mesenchyme from where it diffuses to the distal end[183]. In the distal mesenchyme, *Fgf8* is expressed extensively in the AER and antagonizes RA-mediated induction of cell death by decreasing RA levels through CYP26b1[184–187]. FGF8 further promotes cell survival in the distal mesenchyme by activating the mitogen activated protein kinase (Mapk) pathway[188]. To permit digit individualization, FGF8 signaling needs to be reduced, as persistent expression of *Fgf8* results in ICD reduction and syndactyly in mutant mice[189–191]. *Bmp2, 4* and *7* are expressed in the interdigital mesenchyme at the time of ICD induction. BMP7 induces cell death directly in the mesenchyme, while BMP4, which is expressed in the ectoderm, acts indirectly via Bmp receptor 1a[188,192–195]. As ICD begins, increased Bmp2/4 activity in the ectoderm reduces *Fgf8* expression in the AER and consequently elevates RA levels. RA disrupts the Mapk-mediated survival pathway by inhibition of *Fgfr1* expression and by lowering the levels of phosphorylated ERK1/2. In the proximal interdigital mesenchyme, apoptosis is induced by BMP7 and by RA-mediated expression of *Bax*[188,192]. The mechanisms of ICD are not fully understood. At present, two models are suggested that explain the interplay of various signaling pathways, the massive and the progressive model. In the massive ICD model, interdigital tissue is preformed during limb growth and removed massively by apoptosis between the digits to separate them from each other. In the progressive model, the growth rates of digital and interdigital tissue are similar prior to ICD induction. Then, proliferation is reduced by apoptosis in the interdigital regions while digit condensations keep growing at the same pace and separate gradually from each other (Figure 11)[188].

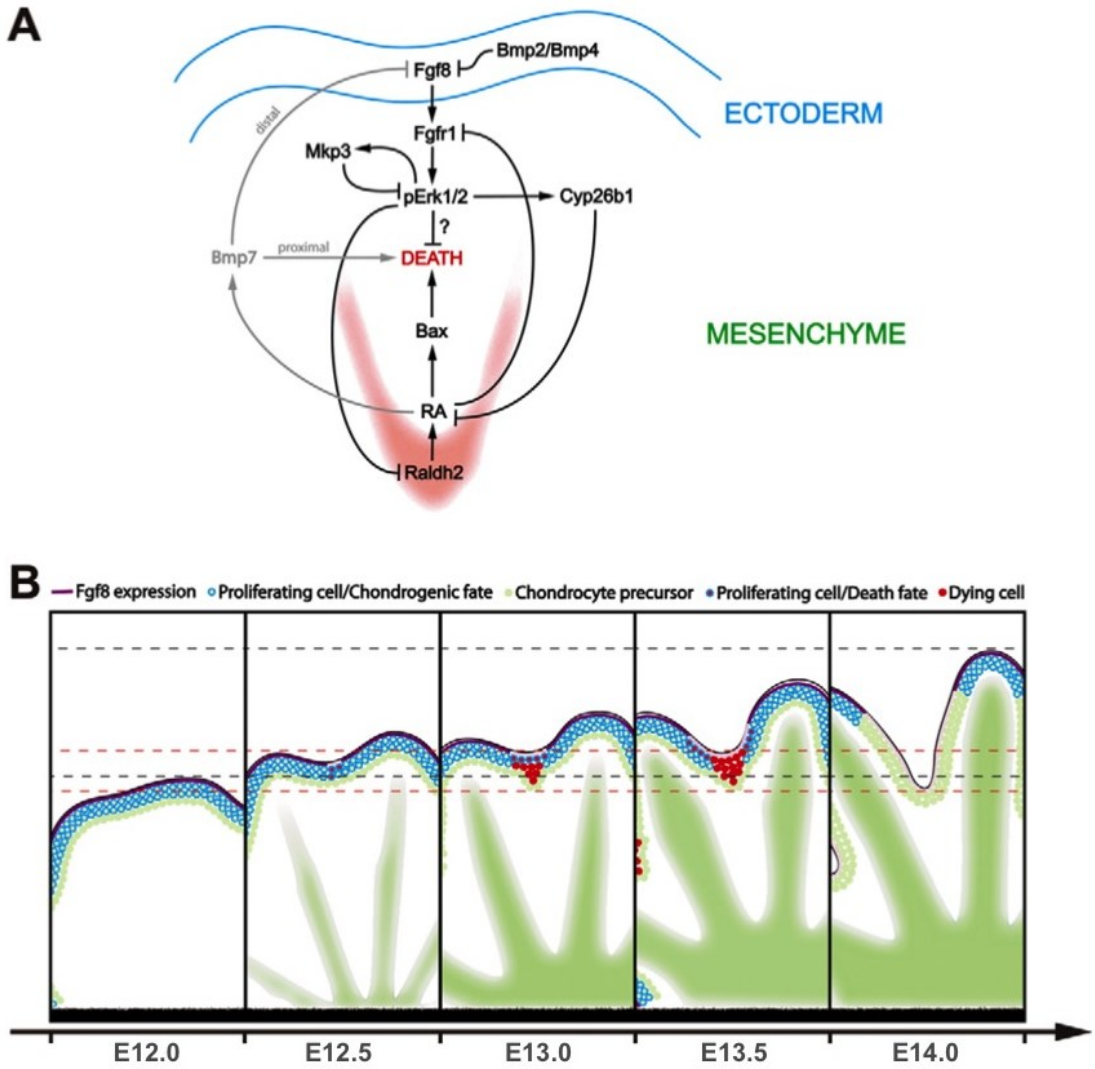

Figure 11. Schematic model of the molecular and cellular processes regulating ICD

(A) In the interdigital mesenchyme two antagonizing factors are expressed prior to ICD induction: proximal RA signals for apoptosis while distal FGF8 signals for cell survival. FGF8 restrains RA activity to support digit outgrowth via the Mapk survival pathway. ICD is induced, as FGF8 signaling is suppressed by upregulation of BMP2/4 in the ectoderm. Enhanced RA activity interferes with FGF8 signaling by inhibiting the expression of *FGF receptor 1* and disrupting the Mapk pathway. Proximally, ICD is induced by BMP7 and RA-mediated expression of *Bax*. (B) The progressive ICD model. First, the digital and the interdigital regions show similar growth rates, as indicated by the black and red dotted lines respectively. Upon FGF8 repression (light purple), apoptosis is induced in the interdigital region (red cells) lowering the rate of proliferation vigorously. Persistent growth in the digital region results in distal protrusion of the digits. Proliferating cells and chondrocyte precursors are shown in blue and green, respectively (adapted from Hernández-Martínez *et al* 2009).

3.4 CNVs at the *IHH* Locus Interfere with Skeletal Development

Four variable duplications upstream of *IHH* have been shown to be associated with craniosynostosis Philadelphia type and syndactyly type 1, phenotypes that are not directly related with *IHH* function[196,197]. In the first family, 77 affected members of eight generations showed cutaneous syndactyly in hands and feet of variable degree[198]. The other families featured cutaneous (family 4) as well as distal osseous syndactyly (family 2 and 3) together with craniosynostosis of the sagittal suture[197]. All duplications are located in close proximity of *IHH*, spanning parts of the neighboring gene *NHEJ1*, but do not expand towards other neighboring genes. The functional characterization of the common overlapping region by reporter gene expression in mouse embryos suggested the localization of *Ihh*-specific enhancers in this region[196]. However, the *in vivo* functionality of these elements in skeletal development and how the variable duplications might affect the pathogenesis in human phenotypes remains elusive. Thus, the *IHH* locus represents a suitable region to investigate the effects of non-coding CNVs on the regulation of gene expression during skeletal development.

3.4.1 Craniosynostosis

Craniosynostosis is characterized by the premature fusion of one or more cranial sutures of an infant skull[199,200]. Thus, the skull is forced to expand perpendicularly to the closed sutures to compensate for the spatial limitations during brain growth. These changes in the growth pattern of the skull can result in an abnormal head shape and facial features[201,202]. Craniosynsotosis Philadelphia type has been linked to three variable non-coding duplications at the *IHH* and is a rather mild form of craniosynostosis. It is clinically characterized by sagittal craniosynostosis with a relatively normal facial appearance, accompanied by complete cutaneous syndactyly of the fingers and toes[203]. In severe cases of Craniosynostosis, such as Apert Syndrome, the compensation does not provide enough space for proper brain growth, resulting in increased intracranial pressure and thus impairments in sleeping, hearing, vision and eating as well as mental retardation[204–207] (Figure 12). Craniosynostosis is inherited in an autosomal dominant fashion and occurs either isolated (non-syndromic) or as part of other syndromes with an incidence of 1:2000[200,208]. Most cases are caused by mutations that result in impairments of the FGF or IHH pathway[129,200,209]. Gain of function mutations in *Fgfr2* increase the affinity of the receptor for FGF binding, resulting in increased proliferation and differentiation of osteoblasts at the midline calvarial sutures[210]. MSX2 is a transcriptional activator of *Ihh* during bone development. Mutations causing an upregulation of *Msx2* expression were shown to induce Craniosynostosis[211,212]. Furthermore, Craniosynostosis is associated with mutations in *Rab23* or *Gli3* being a negative regulator or transducer of IHH signaling, respectively[213–215].

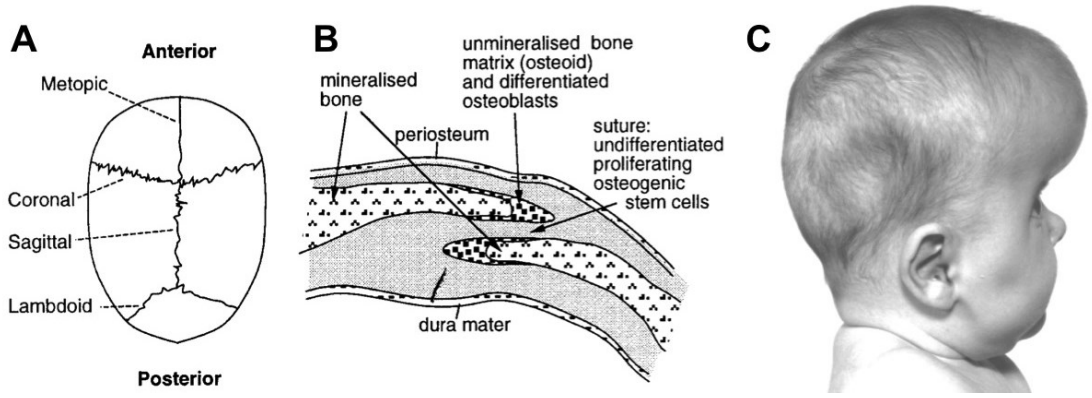

Figure 12. Cranial suture development

(A) Top view of a normally developing infant skull showing the position of the four major sutures: metopic, coronal, sagittal and lambdoid. (B) Schematic cross section of the coronal suture showing a slight overlap of the skull bones with a narrow space separating the bones. (C) Apert syndrome: premature synostosis of the coronal sutures of a six-month-old patient. The brain is pressed through the widely opened metopic suture (adapted from Wilkie *et al* 1997, Wilkie and Morris-Kay 2001).

3.4.2 Synpolydactyly

Defects in limb patterning and digit separation are correlated with congenital malformations of the limb. Synpolydactyly represents one classic disorder of skeletal patterning, comprising the two clinical phenotypes syndactyly and polydactyly[216]. Synpolydactyly is one of the most frequent congenital skeletal malformations in human, with an incidence of 3-10:10000 for syndactyly and 2:1000 for polydactyly[217,218].

Syndactyly is characterized by the fusion of digits, affecting all digits of the limb (complete syndactyly) or only some digits (incomplete)[219], (Figure 13). Thus, different types of syndactyly have been described (Table 1). The phenotypic heterogeneity of syndactyly is determined by the disruption of two fundamentally different mechanisms of limb development: defective bone growth or digit separation. In complex forms of syndactyly the bones of adjacent digits are fused due to irregular bone growth in early limb development. In simple or cutaneous syndactyly digits are joined by webbing of soft tissue as a result of the impairment of ICD and the failure of digit separation[180,217–219]. Various mechanisms have been described to interfere with ICD and all of them are linked to persistent FGF8 signaling, which promotes cell survival in the distal mesenchyme[189–191]. In ICD, FGF8 signaling is reduced upon increased BMP2/4 activity. Thus, impairment of the BMP pathway caused by mutations in *Bmp receptor 1a* or *Bmp2* and *Bmp4* or k.o. of the BMP downstream targets *Msx1* and *Msx2* result in syndactyly phenotypes[190,192,193]. It has been shown that ectopic expression of *Shh* in the interdigital mesenchyme is associated with cutaneous syndactyly in chicken[220]. Recently, non-coding CNVs at the *IHH* locus have been linked to syndactyly type 1[221,222].

Table 1. Non-syndromic syndactyly phenotypes

SD[1] Type	Features
SD I	Syndactyly of the 3rd + 4th finger web space and/or the web between the 2nd and 3rd toes
SD II	Syndactyly associated with polydactyly
SD III	Complete/bilateral, generally soft tissue syndactyly between the 4th and 5th fingers, 5th finger with absent or rudimentary middle phalanx
SD IV	Complete syndactyly, bilateral with polydactyly, generally 6 metacarpals and 6 digits
SD V	Soft tissue webbing between the 3rd and 4th fingers and 2nd and 3rd toes, associated metatarsal and metacarpal fusion
SD VI	Unilateral syndactyly of digits 2-5
SD VII	Severe shortening of the ulna and radius with fusion, fusion of the metacarpals and 'disorganization' of phalangeal development including syndactyly
SD VIII	Fusion of the 4th and 5th metacarpals
SD VIIII	Complete syndactyly and synostosis of the 3rd and 4th fingers, severe bone reduction in the proximal phalanges, hypoplasia of the thumbs and halluces, aplasia/hypoplasia of the middle phalanges of the 2nd and 5th fingers, complete or partial soft tissue syndactyly of the toes

[1] SD: Syndactyly

Polydactyly is characterized by the appearance of additional digits at the anterior or posterior side of the limb, referred to as preaxial or postaxial polydactyly, respectively (Figure 13)[217,223]. In severe polydactyly, all fingers or toes are duplicated resulting in a mirror image of the hand or foot[224]. For polydactyly various modes of inheritance have been described, however in many cases the mode of inheritance remains unclear. A multitude of genes are involved in the anterior-posterior patterning of the limb and have been described to cause polydactyly[225]. SHH is a key regulator of anterior-posterior limb patterning, which signals from the ZPA gradually across the limb bud[158]. *Shh* expression is restricted to the ZPA to ensure the correct formation of digit number and identity[150,151,167,168]. Ectopic expression of *Shh* at the anterior side of the limb bud induces the formation of additional digits and loss of digit identity by disrupting the Gli3A/Gli3R gradient[164,217,223]. Misregulation of *Shh* expression caused by disruptions of the SHH regulatory region ZRS was shown to be associated with polydactyly and malformation of individual digits [217].

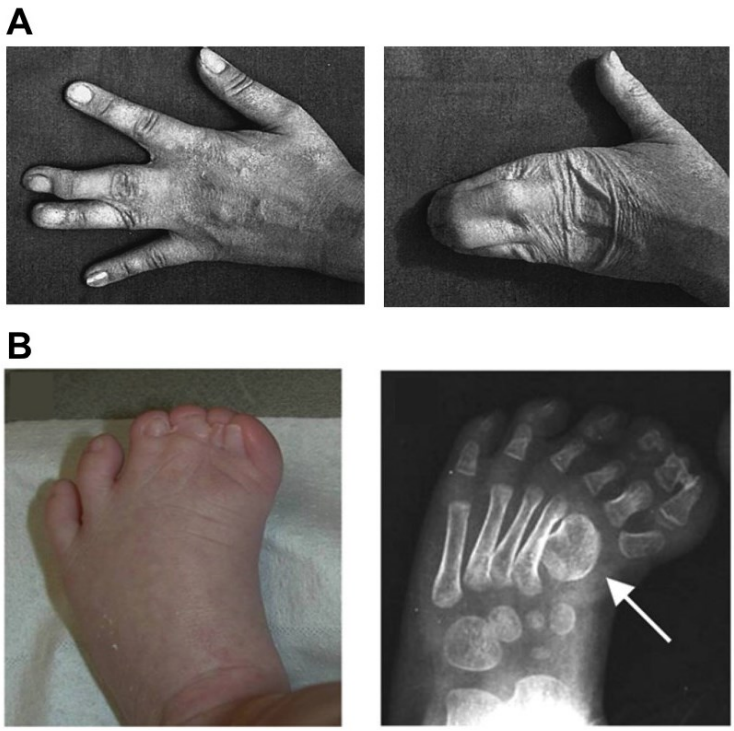

Figure 13. Phenotypic variability of syndactyly and synpolydactyly

(A) The variability of the syndactyly phenotype ranges from partial webbing of two or more digits (left) to complete fusion of the digits by webbing of the skin or fusion of the bones (right). (C) Preaxial synpolydactyly of the feet: The X-rays show an abnormal shape of first metatarsal (arrow), (adapted from Bosse *et al* 2000 and Klopocki *et al* 2011).

3.5 Aims of the Study

Developmental genes, like *IHH*, are often regulated by a cluster of redundant enhancers to ensure spatio-temporal specificity of expression. CNVs of the non-coding genome might result in alterations in enhancer dosage and composition that could induce pathomechanisms of disease. In human, CNVs affecting the upstream non-coding region of *IHH* have been shown to be associated with disease, suggesting alterations of regulatory mechanisms. This study aims to investigate the *cis*-regulation of *Ihh* and how perturbations in enhancer number and cluster composition affect the pathogenicity of the developing phenotypes, in an exemplary manner.

1. Identification of *Ihh* Exclusive Enhancers:

4C-Seq allows for the detection of chromatin interactions that are exclusive for a chosen viewpoint and thus will be applied to identify the regulatory landscape of *Ihh*. The insertion of a SB reporter at the center of the interacting region will reveal its regulatory potential in different tissues and developmental stages.

Based on the 4C-Seq interaction profile, mammalian conservation and enhancer specific chromatin marks individual elements will be selected for enhancer characterization. To elucidate the regulatory potential of these elements enhancer-reporter assays will be conducted at two developmental stages. A scoring of the potential enhancers will be achieved based on their tissue-specific activities to determine the degree of redundancy among these elements.

2. Generation of CNVs to Induce Alterations in Enhancer Dosage and Cluster Composition Using CRISVar:

To study the effects of altered enhancer dosage and cluster composition on gene expression, CRISVar will be applied to induce consecutive CNVs (deletions and duplications) of the *Ihh* enhancer cluster. The efficiency of CRISVar to generate deletions and duplications will be evaluated statistically for different target sizes.

3. Functional Characterization of the *Ihh* Enhancers:

To determine the functionality of potential enhancers *in vivo*, the enhancer cluster will be deleted using CRISVar. The skeletal and the molecular phenotype of the transgenic embryos will be analyzed and compared with wt and *Ihh* k.o. controls. In addition, consecutive deletions of defined cluster domains will be generated to determine how the enhancer cluster controls *Ihh* expression in different tissues and to answer the question if these elements are fully redundant or if they perform in a different mode of action.

4. Effects of Increased Enhancer Dosage and Alterations in Cluster Composition:

For the generation of duplications, CRISVar will be applied. Alterations of the skeletal phenotype in transgenic mice will be monitored through skeletal staining and Micro-CT analysis at pre- and postnatal stages and compared with wt controls. Effects on *Ihh* expression will be assayed using WISH and quantified with qPCR. Three duplications will be generated:

To assess the effect of increased enhancer dosage on skeletal development and gene expression, a duplication containing the *Ihh* enhancer cluster will be generated.

To understand how changes in cluster composition affect *Ihh* expression and how these effects can be linked to disease, two of the disease-associated human duplications will be reengineered in mice. In human, these two duplications resulted in the phenotypes craniosynostosis, syndactyly and polydactyly. Therefore, transgenic mice will be phenotypically examined with a focus on skull and limb development. To examine underlying regulatory mechanisms that induce these diseases, transgenic embryos will be investigated at different developmental stages. To determine the pathomechanism that results in syndactyly, the effects of the duplications on ICD-mediated digit separation will be investigated in the developing limb using WISH and apoptosis assay. In addition, 4C-Seq will be applied to elucidate the effects of CNV-induced alterations in enhancer dosage and cluster composition on gene regulation.

4 Material

4.1 Instruments

Name	Type	Suplier
BioRobot	M48 workstation	Qiagen
Centrifuges	5417R, 5804	Eppendorf
	Megafuge 1.0R	Kendro
Cleanbench	HERA Safe K109	Thermo Scientific
CO2-Incubator	HEPA Class 100	Thermo Scientific
Counting Chamber	0.0025mm	Marienfeld
Hematocytometer	SUPE-RIOR, Depth 0.1mm 0.0025mm^2	Neubauer CE
Heating Block	Ori-Block OV3	Techne
Imaging systems	Curix 60	Agfa
	LAS 4000	Fuji
Microscope	Stereo Microscope SteREO	Zeiss
	Discovery. V12	Zeiss
	Pro 2012 DMIL	Leica
pH-meter	MP220	Mettler
Rollermixer	Rollerdrum TC-7	New Brunswick Scientific
Shaker	G10 Gyrotory	New Brunswick Scientific
Skyscan	1172 X-ray microtomography system	Brucker microCT
Spectrophotometer	Nanodrop 2000	Thermo Scientific
Semi-Dry Blotting Chamber		Biorad
Thermo Cycler	Simply Amp Thermal Cycler	Life technologies
Thermo Cycler qPCR	ABIPrism HT 79000 RT cycler	Applied Biosystems
Thermomixer	Comfort 1.5ml	Eppendorf
Tissue Lyser		Qiagen/Retsch
Transilluminator		Hertenstein
UV chamber		Hormuth-Vetter
Vortex	Microspin FV-2400	Lab4you
Waterbath	D1	Haake fisions

4.3 Chemicals

Chemical	Supplier
2-Methoxyethylacetet (MEA)	Polysciences
Alcian Blue	Sigma
Alizarin Red	Sigma
Anti-Digoxygenin-AP (from sheep)	Roche #11093274910
CDPStar	Roche
DIG Ladder	Thermo Fisher SCIENTIFIC
DNA Ladder 100bp Plus	Thermo Fisher SCIENTIFIC
Formamide p.a.	Merck
Glycerol	MPI
Heparin	Sigma H 3149
Lysotracker w Red DND-99	life technologies
NP 40 Nonident P40	Fluka 74385
Proteinase K	Boehringer 1000144
QuickExtrackt DNA extraction Solution	Quiagen
RNAse type III from bakers yeast	Sigma R 6750
Tetramisole-Hydrocloride	Sigma L9756

4.4 Cell Culture Ingredients

Name	Supplier
100xglutamin	Lonza BE17-605E
1xPBS	Lonza BE17-512F
100xPenicillin/Streptomycin	Lonza DE17-603
Beta-Mercaptoethanol	Sigma M-7522
Bicarbonate free Media	Gibco 52100
DMEM	Lonza BE12-733
DMSO	Sigma Aldrich D-2650
FCS for feeder	Biochrome
FCS for ESCs	PAN Biotech P122011
Fugene	Promega TM-238
Gelantine	Sigma G-1393
KO-DMEM	Gibco 10829-018
LIF	Chemicon ESG1107
Mytomycin C	Sigma M-4287
Non-essential Amino acids	Gibco 11140-35
Nucleosides	Chemicon ES-008D
OptiMEM	Gibco
Trypsin-EDTA	Gibco 25300-054
Water	Lonza BE 17-724Q

4.5 Buffers and Solutions

Buffer	Ingrediens
10xPBS	80g NaCl, 2g KCl, 14.4g Na2HPO4, 2.4g KH2PO4 ad 1l ddH_2O
20xSSC	175.3g NaCl, 88.2g tri-Sodium Citrate Dihydrate, adjust pH 4.5 with 1M Citric Acid, 1ml DEPC-H2O, add 1l ddH_2O, incubate at 37°C ON, autoclave
20%SDS	28.84g SDS, add 500ml H_2O, filter with paper filter, do not autoclave
4%PFA	40gPFA ad 1l 1xPBS
4%PFA-DEPC	40gPFA ad 1l 1xPBS-DEPC
Alcian Blue Solution	150mg/l Alcian Blue, 80% EtOH, 20% Acetic acid
Alizarin Red Solution	50mg/l Alizarin Red in 1%KOH, or in 0.2%KOH
Cell-Lysisbuffer	10mM TrisHCl, 10mM EDTA, 0.2% SDS, 100mM NaCl,

	10µl/ml ProteinaseK
DEPC-H_2O	0.1% DEPC in ddH_2O
DNA Loading Buffer	6%Succrose, 0.7% Orange G, dissolved in ddH_2O
GC-PCR Buffer	1.7ml 1M SO4, 6ml Tris/HCl (pH8.8), 0.2ml MgCl2, 70µl Beta-Mercapto-Ethanol, 2.03 ml ddH_2O
GC-PCR Mix	250µl 5mM dNTPs, 250µl 1M DMSO
Proteinase K	20mg in 1ml Proteinase K Buffer
Proteinase K Buffer	1ml 1M Tris (pH7.0), 0.1ml 0.5M EDTA, add 50ml DEPC-H_2O
Tissue Lysis Buffer	17mM Tris (pH7.5), 17mM EDTA, 170mM NaCl, 0.85% SDS, 3µl/ml ProteinaseK

4.6 Kits

Name	Supplier
BigDye Terminator v3.1 Sequencing Kit	Applied Biosystems
Expand Long Template PCR System	Roche
NucleoSpin Plasmid	Macherey-Nagel
Nucleobond PC100	Macherey-Nagel
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel
PCR DIG Probe Synthesis Kit	Roche
RNAeasy Kit	Qiagen
SYBR Green qPCR Master Mix	Qiagen
TaqMan Reverse Transcription Kit	Roche
QIAquick PCR purification kit	Roche

4.7 Enzymes

Restriction enzymes, ligase, polymerases and other DNA modifying enzymes were purchased from Thermo Fisher SCIENTIFIC.

4.8 Oligonucleotides

All DNA-Oligonucleotides are listed in 5'prime to 3'prime orientation. Primers and CrispR guides were purchased as HPSF purified products from Eurofins MWG.

4.8.1 Oligonucleotides for Sleeping Beauty Cloning

Construct	Genomic Position	Size	Primer	Sequence[1]
centromeric homologous arm				
SB-HR-L1	Chr1:75,055,877-75,058,875	3kb	HR3a1-f-SalI	tata***gtcgac***caaagtccttgtaaggaacagcagt
			HR3a1-r-ClaI	tata***atcgat***gacatgcctctgctgtacatagttt
SB-HR-L2	Chr1:75,058,877-75,060,875	2kb	HR3a2-f-F3-ClaI	tata***atcgat***tacaagctttacgaagttcctattcttcaaatagtataggaacttcagcaactcaggaagaattcctaacac
			HR3a2-r-F3-SacII	tata***ccgcgg***gtagaagttcctatactatttgaagaataggaacttcttgcagccctcctatagaaaatgga
telomeric homologous arm				
SB-HR-R	Chr1:75,060,877-	3kb	HR3b-f-XhoI	tata***ctcgag***tctataagaacacacaacaatgtgcca

	75,063,875	HR3b-r-NotI	g tata***gcggccgc***actgttctgggtgaaccagaaatc tt

[1]Restrictions sites are shown in italic

4.8.2 Oligonucleotides for LacZ Reporter Constructs

Element	Vista ID	Primer forward, reverse	Genomic Position (mm9)	Size (bp)
i1	mm1142	ctcagtgtctcaaccacttgaa, ctctgccatgacttcttgtgta	chr1:75,008,008-75,012,847	4840
i2	mm1143	ggtgggattaatctctcgactg, ggtgatgaacagcagtatggaa	chr1:75,023,290-75,026,536	3247
i3	mm1148	tctcccagaccaaaatgcttat, aaccttgccctcatgaagttta	chr1:75,046,263-75,049,025	2763
i4	mm1144	cagactggagttcacagagtgc, actcaggcacaagtctagcaca	chr1:75,051,762-75,053,663	1902
i5		cctctgtgctcttgagttagactac, cctccttgctagttcttacctaaaga	chr1:75,053,880-75,055,928	2045
i6	mm1145	tccttgagagactccagaaagg, tcccccatatcagatgtttacc	chr1:75,059,085-75,064,020	4936
i7	mm1146	gtactgggaaaaatggcaagag, ctgaaaggggttagaaggact	chr1:75,068,299-75,072,430	4132
i8		ttgaggcagaaggattgtcata, agccagaggtcaacatttgagt	chr1:75,075,786-75,080,268	4483
i9	mm1439	gctgagatgaatgacagtgagg, gtcacacctgatgatctgcatt	chr1:75,085,302-75,089,234	3933

4.8.3 Oligonucleotides for CrispR Constructs

Name	Sequence	Genomic Position	Score	Genomic Score >1.0	Exonic Score >0.3
CR-N2-L1	gagacacgtggagaattcgc-agg	chr1:75,015,710-75,015,731	86	1	3
CR-N2-R1	gttacccacactactacgtt-agg	chr1:75,091,165-75,091,187	94	0	0
CR-N4-L2	ggacacgactttcataacac-tgg	chr1:75,050,992-75,051,014	84	0	1
CR-N4-R2	aatttcgggtagggcgttgg -agg	chr1:75,063,545-75,063,567	89	0	0
CR-N9-L9	agcgtggggcttttaaccgt-ggg	chr1:74,989,792-74,989,811	94	0	0
CR-N9-R6	ttagacacaccagtatacgg-agg	chr1:75,055,612-75,055,634	92	0	0
CR-N10-L4	ggggcaatctgatatagtgg-ggg	chr1:75,005,921-75,005,943	85	0	0
CR-N10-R5	tggcccctgacccgtaggat-tgg	chr1:75,060,408-75,060,430	93	0	0

4.8.5 Sequencing Primer

Name	Sequence	Target
CR-Seq-f1	gaaagtaataatttcttgggtagtttgcag	CrispR guide in pX459
CR-Seq-r2	gctctaaaacaaaaaagcaccgactc	CrispR guide in pX459
HR3a1-seq1r	gtaggaaaggtaatttgtcttgtatttttc	SB-HR-L1
HR3a1-seq2f	agagctggaataggaatttgttcttg	SB-HR-L1
HR3a1-seq2r	gctgctctggtgtgttgtgt	SB-HR-L1
HR3a1-seq3f	gaggagaaagagagtttgtgtatgc	SB-HR-L1
HR3a1-seq3r	gactttatagttggccaagcattct	SB-HR-L1
HR3a1-seq4r	caaacttttgactcatgagctttgg	SB-HR-L1
HR3a1-seq5r	tgtgtcaccactacctggct	SB-HR-L1
HR3a1-seq5f	caggacctagctaaaggaacg	SB-HR-L1
HR3a1-seq7f	ctccgtagcagaaagctgacta	SB-HR-L1
HR3a2-seq1r	cataagttcttggatagatatctaaggga	SB-HR-L2
HR3a2-seq2r	gtaattatggaacttttctttgcctttc	SB-HR-L2
HR3a2-seq2f	gactccagaaaggccctcag	SB-HR-L2
HR3a2-seq3f	cctaccatttctggtcaatccca	SB-HR-L2
HR3a2-seq4f	cgactgaacttcaagcgtactag	SB-HR-L2
HR3a2-seq5f	ttagaaagagaagtggcaccgct	SB-HR-L2
HR3a2-seq6f	ccaagggaagcttacactagca	SB-HR-L2
HR3b-seq1r	gatggtggaaggcttgtaaaatga	SB-HR-R
HR3b-seq2r	tcctttggcttattctgggttgg	SB-HR-R
HR3b-seq3r	acagaaccctgggggagtg	SB-HR-R
HR3b-seq4r	tcaaactccccaagtctggttc	SB-HR-R
HR3b-seq5r	ggatccccaaaaatttgagcagat	SB-HR-R
HR3b-seq5f	acaactaggctgttggaatcaatc	SB-HR-R
HR3b-seq6f	gacaggacagctactaaatagatc	SB-HR-R
HR3b-seq106.1f	tcaagtcaggtgtgacagttcc	SB-HR-R

4.8.6 PCR-Genotyping Primer

Line	Target Site	Primer Pair	Name	Sequence
Del(i2-9)	5'BP	F 456	CRN-L13-R14-f3	cacacgcctctggcaatttctac
and		R 458	CRN-L13-R14-r2	gctggctctgactagcatcacg
Dup(int)	3'BP	F 459	CRN-L13-R14-f4	ctgtgagagagtctggatatgagcg
		R 460	CRN-L13-R14-r3	gggagtagataattgtggcctgtc
Del(i4-6)	5'BP	F 470	CRN-L22-R22-f1	aaagacctatgacaggagagtgaagc
		R 471	CRN-L22-R22-r1	atggctttgctgattctgttgg
	3'BP	F 472	CRN-L22-R22-f2	aaagagaaaccagggacgcaga
		R 473	CRN-L22-R22-r2	gaggttattcaagtgctgtgtccc
Del(i2-3)	5'BP	456 + 458		
	3'BP	470 + 471		
Del(i7-9)	5'BP	472 + 473		
	3'BP	459 + 460		
Del(i4-9)	5'BP	470 + 471		
	3'BP	459 + 460		
Dup(syn)	5'BP	660	CR-N9-1f1.2	tttagccctttgccatttg
		661	CR-N9-1r1.2	ggtagcaggagtcagggaaa
	3'BP	662	CR-N9-1f2.1	gaaagcaacatcggcagaat
		663	CR-N9-1r2.1	ccccctccttgctagttctt
Dup(csp)	5'BP	664	CR-N10-f1.2	acacttcccgcaaccataaa
		665	CR-N10-r1.2	acaagagggctgatccacaa
	3'BP	666	CR-N10-f2.1	ttttcccaaagcatctcacc

		667	CR-N10-r2.1	gatagcgtaccccagagcaa
SB	5'SB	380	CE3-f1	gggaagaaacgcagtcagtc
		381	SB-L-r1	aacttccgacttcaactgtaggg
	3'SB	385	R2-M321-f	gtggtgatcctaactgacctaagac
		101	HR3b-seq1r	gatggtggaaggcttgtaaaatga
	WT	380+101		
Ihh ko	KO		Ihh-5	aggaggcagggacatggatagggtg
			Ihh-KO	taccggtggatgtggaatgtgtgcg
	WT		Ihh-ex1-F1	gcctcttgcctacaagcagt
			Ihh-ex1-R2	tgttctcctcgtccttgaaga

4.8.7 Primer for Genomic Copynumber Analysis

Line	Target Site	Primer Pair	Name	Sequence
Del(i2-9)	5out	509	qPCR-N2-5f1	ctgaacatgtcctgattctctca
and		510	qPCR-N2-5r1	atgcaggacacttatcttaccat
Dup(int)	5in	493	qPCR-N1-5f1	tgaccttacacagtcagactct
		494	qPCR-N1-5r1	agctttctcgttgcactctg
	5center	541	qPCR-N4-Cf1	cgtgcttggccaacatgat
		542	qPCR-N4-Cr1	gggaactttgacccaccaag
	3in	497	qPCR-N1-3f1	tgcattcgtctagtccttgc
		498	qPCR-N1-3r1	gggtaatccttccctgctact
	3out	949	qPCR-N2-3f6	accttatatgcaacccccaaa
		950	qPCR-N2-3r0	ttggctctgcgaatagaacc
Del(i4-6)	5out	525	qPCR-N4-5f1	agacacgataaccttctgcca
		526	qPCR-N4-5r1	ctcttgctttgagtttgttggac
	5in	533	qPCR-N4-Lf1	attacccaggtgcactctcc
		534	qPCR-N4-Lr1	atgcccactagaaccaacca
	5center	541 + 542		
	3in	537	qPCR-N4-Rf1	aaattaggggcgtctcaggg
		538	qPCR-N4-Rr1	ggctctgatcaaactcccca
	3out	529	qPCR-N4-3f1	gtttcccggcctgctatctt
		530	qPCR-N4-3r1	gtcaccaccacaagagtaactg
Del(i2-3)	5out	509 + 510		
	5in	493 + 494		
	5center	616	qPCR-N5-C-f11	gtgggcttaactgccaagtc
		617	qPCR-N5-C-r11	gccaccacaaaggaacaact
	3in	525 + 526		
	3out	533 + 534		
Del(i7-9)	5out	537 + 538		
	5in	529 + 530		
	5center	618	qPCR-N6-C-f4	ggccaggttatcaggcaata
		619	qPCR-N6-C-r4	cacgaatgaacgtgtcaagg
	3in	497 + 498		
	3out	949 + 950		
Del(i4-9)	5out	525 + 526		
	5in	533 + 534		
	5center	541 + 542		
	3in	497 + 498		
	3out	949 + 950		
Dup(syn)	5out	674	qPCR-N9-5f1	tttagccctttgccatttg
		675	qPCR-N9-5r1	cacagttcttgaggttgagtgtc
	5in	964	9L-f4	agaacccaggccctgagaag
		965	9L-r0	acagagcagaggtccatcgtca

	5center	973	qPCR-N10-Lf0	caccagttggtagggatgct
		974	qPCR-N10-Lr0	ggcacaagtttgttgggaaca
	3in	535	qPCR-N4-Lf2	gcatgctcacgtcttctcac
		536	qPCR-N4-Lr2	tgtcggtcataggaaccact
	3out	541 + 542		
Dup(csp)	5out	688	qPCR-N10-5f2	tttgaaccagcaaggaaagg
		689	qPCR-N10-5r2	tgttggtctagaggatggcttt
	5in	973	qPCR-N10-Lf0	caccagttggtagggatgct
		974	qPCR-N10-Lr0	ggcacaagtttgttgggaaca
	5center	613	qPCR-N5-C-f46	gctgagaagtgcgttgttagg
		615	qPCR-N5-C-r4	atagcattgtcgccatgtca
	3in	541 + 542		
	3out	539	qPCR-N4-Rf2	tggggaaattaggggcgtc
		540	qPCR-N4-Rr2	ctgggctctgatcaaactcc
Reference		Ref-f	qmCopy_no_A_F	agctagattaccctgagtcca
		Ref-r	qmCopy_no_A_R	ttcaagtaggctcggtcacc

4.8.8 qPCR Primer

Primername	Sequence	Target gene
qPCR-Gapdh-f1	gggaagcccatcaccatctt	Gapdh
qPCR-Gapdh-r1	cggcctcaccccatttg	
qPCR-Ihh-f3	gccgaccgcctcatgac	Ihh
qPCR-Ihh-r3	catgacagagatggccagtga	

4.8.9 4C-Seq Primer

Viewpoint	1st primer (5' – 3')	Genomic Position	2nd primer (5' – 3')	Genomic Position
Ihh promoter	acagctggggaccctatac	chr1:74,998,765-74,998,783	cccgtcaggaggacaatc	chr1:75,059,837-75,059,854

4.9 Bacteria

Alle cloning steps of plasmids were performed in Eschericia coli TOP10 cells.

4.10 Plasmids

Name	Application	Supplier
pBSII-SB	Sleeping Beauty cloning	Francois Spitz (EMBL, Heidelberg)
pX459	expression of sgRNA and Cas9	Addgene Feng Zhang Lab plasmids

4.11 Software

Name	Application
ApE – a plasmid editor	Construct design and Sequence analysis
Axio Vision	Digital photography
CTan (BruckerCT)	Micro-CT meassurements
DNA Star Seqman	Sequence analysis
Easy Win. 32	Gel documentation system
Inkscape	Image processing
Microsoft Office	Data analysis, text processing
SDS	qPCR analysis
ZEN	Digital photography

4.12 Internet Resources

Resource	Adress
BioGPS	http://biogps.org
CRISPR Design	http://crispr.mit.edu/
EMBL-EBI	http://www.ebi.ac.uk/
Ensembl Genome Browser	http://www.ensembl.org/index.html
Human Protein Atlas	http://www.proteinatlas.org/
MGI – Mouse Genome Informatics	http://www.informatics.jax.org/marker/
Mouse Atlas	http://www.emouseatlas.org/emap/home.html
NCBI	http://www.ncbi.nlm.nih.gov/
Primer3	http://primer3.ut.ee/
NetPrimer	http://www.premierbiosoft.com/netprimer/
GET Prime	http://bbcftools.epfl.ch/getprime/
MGH Primer Bank	https://pga.mgh.harvard.edu/primerbank/
UCSC	Genome Browser http://genome.ucsc.edu/
Vista Enhancer Browser	http://enhancer.lbl.gov/
YUElab	http://promoter.bx.psu.edu/hi-c/view.php
graphpad	https://www.graphpad.com/quickcalcs/ttest1/
Bing Ren Hi-C data	http://promoter.bx.psu.edu/hi-c/

5 Methods

5.1 Molecular Biology Methods

If not described differently all classical molecular biological or microbiological experiments were performed according to the handbook "Molecular Cloning: A Laboratory Manual"{Citation} (Sambrook et al., 2012).

5.1.1 DNA Isolation

Genomic DNA was isolated either from ear biopsies using QuickExtract (Quiagen) or from amnion.

For the genotyping of mice, ear biopsies were provided by the animal facility (Max Planck Institute for Molecular Genetics, Berlin). To each sample 50µl QuickExtract was added and incubated for 20min at 65°C, shaking at 400rpm. The reaction was stopped by incubation for 2min at 98°C and subsequently cooled on ice. The reaction volume was separated from the sample, both were stored separately at -20°C.

For embryo genotyping amnions were used. The tissue was lysed in 300µl Tissue-Lysation Buffer supplemented with 0.2µg/ml Proteinas K at 55°C in a Thermomixer under agitation of 800rpm. To remove insoluble components, like hair, the lysate was centrifuged for 10min at highest speed and the supernatant was transferred into a new Eppendorf tube. The DNA was precipitated with 350µl Isopropanol. After centrifugation for 10min at full speed the supernatant was removed and the DNA pellet was washed twice with 70%Ethanol. The DNA was air dried for 5 min and dissolved in 50µl ddH_2O. Sample and DNA were stored at -20°C.

To isolate **DNA from cultured cells**, cells are lysed by over night (ON) incubation in 50µl Cell-Lysisbuffer per 96-plate well at 55°C. The next day the DNA is precipitated in 1ml Isopropanol containing 100µl 8M LiCl. After 20 minute centrifugation at 10.000rpm and 4°C the DNA-Pellet was washed twice with 70%Ethanol. The DNA was air dried for 5 min and dissolved in 50µl ddH_2O.

Isolation of **Plasmid DNA** from *E.coli* was performed with the Mini-Kit (Macherey-Nagel) for 5ml bacteria culture and with the MIDI-Endotoxinfree-Kit (Macherey-Nagel) for 50ml bacteria culture according to the specification of the manufacturer.

5.1.2 RNA Isolation

Tissue preparation was performed in icecold PBS/DEPC and tissues were frozen in liquid nitrogen and stored at -80°C upon processing. Tissues were disrupted using TissueLyser at a frequency of 20/s for 2min. RNA-Isolation was performed with the RNAeasy Kit (Quiagen) according to the specification of the manufacturer.

5.1.3 Generation of cDNA

The generation of complementary DNA (cDNA) was performed using the TaqMan Reverse Transcription Kit (Roche) according to the specification of the manufacturer. Hexamer primers were used.

5.1.4 Polymerase Chain Reaction (PCR)

5.1.4.1 Standard PCR

Genotyping PCR of different mouse lines was performed with primers listed in 4.8.6. The protocols and the cycler programs for standard genotyping PCR are listed below. The settings and the expected product sizes for each genotyping are summarized.

Component	Final Concentration	Cycler Program		
10x Taq Buffer	1x			
12.5mM dNTPs	50µM			
Primer F (10pmol/µl)	500nM	94°C	3min	
Primer R (10pmol/µl)	500nM	94°C	30s	
Taq-DNA Polymerase	0.008U/µl	60°C	45s	30x
Pfu-DNA Polymerase	0.002U/µl	72°C	2min	
Template	2ng/µl	72°C	7min	
ddH_20	add 25µl	8°C		

Line	Product size [bp]			
Conventional Line	Wildtype		Mutant	
Ihh k.o. Line	600		307	
CrispR-Line	5'BP	3'BP	del	dup
Del(i2-9) and Dup(int)	563	483	515	574
Del(i4-6)	904	544	501	993
Del(i2-3)	563	904	950	553
Del(i7-9)	483	544	551	479
Del(i4-9)	904	483	485	905
Dup(syn)	626	883	649	906
Dup(csp)	794	943	939	849
Sleeping Beauty Line	1129	-	743	686

5.1.4.2 Sanger Sequencing

For Sanger Sequencing of PCR products or plasmids BigDyev3.1 (Applied Biosystems) was used according to the specification of the manufacturer. The reactions were preciptated and transferred to the capillary electrophoresis (ABI 3700) by Mohsen Karabsyan at Charité – Universitätsmedizin (Insitute for Molecular Genetics).

5.1.4.3 Quantitative Genomic Real-Time PCR

In this study, quantitative real-time PCR (qPCR) was performed 1) to determine the copy number of genomic modifications and 2) to quantify changes in gene expression levels (low-copy number protocol). For the reaction SybrGreen (Qiagen) was used.

The efficiency of the primers was determined with a standard curve as shown below. For normalization Albumin (genomic copy number) and Gapdh (expression analysis) were used.

Component	Amount [µl]	Dilution	Total ng
1) Standard qPCR protocol			
SybrGreen	6.0	1	10
Primerpairs [2.5µM]	1.0	1:5	2
Template [2ng/µl]	5.0	1:25	0.4
		1:125	0.08
		1:625	0.016
2) low copy number qPCR protocol			
SybrGreen	12.75	1	50
Primerpairs [10µM]	1.0	1:2	25
Template [10ng/µl]	5.0	1:10	5
ddH_2O	ad 20µl	1:100	0.5
		1:1000	0.05

5.1.4.4 4C-Seq PCR

4C samples were amplified using the Expand Long Template System PCR Kit (Roche) according to the specification of the manufacturer. Reactions inputs were 25-200ng of sample/reaction.

Cycler Program		
94°C	2min	
94°C	15s	
55°C	1min	30x
68°C	3min	
68°C	7min	
4°C		

5.1.5 Southern Blot

Southern Blot was used to detect the insertion of the Sleeping Beauty cassette (SB). Following an enzymatic digest and separation of the genomic DNA by gel electrophoresis, the DNA was blotted on a nylon membrane. A DIG-labeled probe specific for the SB cassette was hybridized to the membrane. Probe labeling was

performed using the PCR DIG Probe Synthesis Kit (Roche) according to the specification of the manufacturer.

Buffer	Ingredients
10xDIG1	in ddH_2O:1M Maleic acid,1.5M NaCl (pH7.5)
1M NaPi	in ddH_2O: 1MNa2HPO4 (pH7.2; autoclaved)
10xSSC	in ddH_2O: 50% 20xSSC
50mM	in ddH_2O:5% 1M NaPi
Anti-DIG solution	in DIG2: 1:20 000 Anti-Digoxygenin-AP (from sheep)
Church-Hybridization buffer	in ddH_2O:0.5M NaPi, 7% SDS, 4µM EDTA
Church-Washing buffer	in ddH_2O: 40mM NaPi 1%SDS
Denaturation solution	in ddH_2O: 0.5M NaOH, 1.5M NaCL
DIG1	in ddH_2O: 10% 10xDIG1
DIG2	in DIG1 10%BBR (autoclaved)
DIG3	in ddH_2O: 0.1M Tris (pH9.5), 0.1M NaCl
Neutralization solution	0.5M Tris-HCl (pH7.5) 1.4M NaCl

Day1 – Enzymatic Digest

The genomic DNA of the SB clones was digested at 37°C ON.

Day2 – Blotting

The digested DNA was separated on a 1% agarose gel for 6h at 80V. To improve binding of the DNA to the membrane, the gel was treated for 30min with denaturing buffer and neutralized for 2x 20min. Then, the negatively charged DNA was blotted on a positively charged nylon membrane by ion exchange, applying capillary force.

Day3 – DNA Fixation and Hybridization

To visualize the successful transfer of the genomic DNA to the nylon membrane, the membrane was exposed to UV light. The membrane was neutralized in 50mM NaPi. Subsequently, the DNA was fixed to the membrane by baking it for 2h at 80°C. Pre-hybridization in Church-Hybridization buffer was performed in a rotating glas vial for 1h at 65°C. Prior application to the membrane, the DIG-labeled probe was denatured in 25ml Church-Hybridization buffer at 100°C for 10min. Hybridization of the DIG-labeled probes was performed at 65°C ON.

Day4 – Developing

The membrane was washed 1x 10min in pre-warmed Church-Washing buffer at 65°C, rotating, and 1x 10min at room temperature (RT) in an incubation chamber, shaking. After 5min incubation in 1xDIG1/0.3% TWEEN, the membrane was heat-sealed in a plastic bag with blocking solution DIG2 and incubated at high speed for 30min. For antibody binding, DIG2 solution was replaced by Anti-DIG solution. The membrane was incubated at high speed for 30min. The membrane was washed 2x 20min with 1xDIG1/0,3%TWEEN and for 5min with DIG3 in an incubation chamber. For developing, the membrane was placed in an exposure cassette being covered with DIG3+CDPstar (1:100) for 5 minutes. The solution was removed prior detection.

Detection was performed applying X-ray films for 2-20 minutes depending on the efficiency of the probe and developed with Imaging system Curix 60.

5.1.6 Cloning

For cloning, linear plasmids were dephosyphorylated using alkaline phosphatase while the DNA insert was phosphorylated with T4-Phosphorkinase according to the specification of the manufacturer. Ligation and transformation was performed in *E.coli* TOP10 as described in Sammbrock et al. (2002).

5.1.6.1 CRISPR/Cas9 Cloning

CRISPR/Cas9 Design

The CRISPR /Cas9 technique was used to generate targeted deletions and duplications of regulatory elements at the *Ihh* locus. For this the corresponding position in the mouse genome was determined using the UCSC browser. The CRISPR Design Tool (http://crispr.mit.edu)[226] was used to design the guide RNA for the regions of interest. The settings sequence type 'other region' and the target genome 'mouse' were selected. The software was providing different guide sequences, which were ranked based on their predicted specificity (main score) and their potential off-targets (exonic and genomic off-target score). The off-target scores reflect the likelihood of unspecific binding, which is influenced by the number and position of guide mismatches. The higher the main score and the lower the off-target scores, the more specific and efficient is the guide sequence. All guides selected showed a main score of 80 or higher and the lowest number and value of off-target scores possible. Furthermore, the off-targets were evaluated using the genomic position that was given by the CRISPR Design Tool. I avoided choosing guides with off-targets that could potentially affect vital genes or genes that play a role in the patient syndrome.

Two oligonucleotides of the chosen guide sequence were ordered with overhangs to enable the insertion into pX459, for the sense sequence 'CACCG' (5' side) and for the antisense sequence 'AAAC' (5' side) and 'C' (3' side).

Cloning of CRISPR /Cas9 Guides

The oligonucleotides of each guide were diluted to 100pmol/µl as recommended by the company. To anneal the guides 10µl of each oligonucleotide were mixed with 10µl of 10xLigation Buffer and 70µl ddH_2O. The reaction was placed in a metal heating block that was preheated to 95°C and switched off after 5min. After 15 minutes this block was taken out of the heater and cooled down for another 45min at RT. The annealed guides were phosphorylated in T4-PNK buffer A as described above. The Cas9 expression vector pX459 was linearized with BbsI and dephosphorylated with alkaline phosphatase.

Ligation and transformation were performed as described by Sammbrock et al. (2002). Minis were inoculated with five clones, cleaned up and sequenced with the primers CR-Seq-f1 and CR-Seq-r2.

5.1.6.2 Cloning of LacZ Reporter Constructs

Enhancer-Trap/Sleeping Beauty Reporter

The vector backbone (pBSII-SB) including the SB reporter cassette (minimal ß-Galactosidase promoter, LacZ reporter gene, loxP site and neo resistance) was provided by the laboratory of Francois Spitz[71]. To introduce the cassette in the mouse genome, homologous arms flanking the target region were amplified using Phusion High Fidelity Polymersa kit (Thermofisher Scientific). The amplified PCR products were purified using NucleoSpin Gel and PCR Clean-up and cloned in the vector backbone. The restriction sites SalI/SacII and XhoI/NotI were used. Ligation and transformation were performed as written in Sammbrock et al. (2002). Screening of the clones was performed by genotyping PCR and Sanger Sequencing.

Enhancer-Reporter

The cloning of the LacZ reporter constructs to test the individual enhancers was performed as described previously[227].

5.2 Cell Culture

Media	Composition	Application
2xESC Freezing	in KO-DMEM: 20%FCS, 20%DMSO	Freezing ESCs
2xFeeder Freezing	in DMEM: 20%FCS, 20%DMSO	Freezing Feeder cells
2xPlate Freezing	in Bicarbonate-free DMEM: 20%FCS, 10%DMSO	Freezing ESC plates
20% FCS Feeder	in DMEM: 20%FCS	Freezing Feeder cells
20% FCS ESC	in KO-DMEM: 20%FCS	Freezing ESCs
ESC Media	in KO-DMEM: 15%FCS, 2mM Glutamine, 0.05U/ml Penicillin/Streptomycin, 1xNon-essential Aminoacids, 0.1mM Beta-Mercaptoethanol, 1x Nucleosides, 1000U/ml LIF	Cultivating ESC
Feeder Media	in DMEM: 10%FCS, 2mM Glutamine, 0.05U/ml Penicillin/Streptomycin	Cultivating Feeders
Gelatine	in water: 0.1%Gelatine	Coating dishes

5.2.1 Splitting Cells

Cells were washed twice in 1xPBS and trypsinized in the appropriate volume Trypsin-EDTA for 5-10min at 37°C in a cell incubator. Adding two volumes of media to the cells stopped the reaction. The Cells were suspended using a 1000µl pipette. The suspension

was centrifuged for 5min at 1000rpm. After discarding the supernatant the cell pellet was re-suspended in the appropriate volume of media and seeded on the required area.

5.2.2 Thawing Cells

Cell-vials wee thawed in 37°C water bath, then 1ml media was added to the cells drop wise and the cell suspension was transferred to a 15ml falcon including 5ml media. After centrifugation at 1000rpm for 5min the supernatant was discarded and the cell pellet was re-suspended in the appropriate volume of media and seeded on the required area.

5.2.3 Cryoconservation of Cells

For cryoconservation in vials the cells were trpysinized. After centrifugation the cell pellet was re-suspended in 1ml 20%FCSMedia per vial. To freeze one vial 1ml of this cell suspension was added to 1ml freezing media. The vials were frozen in a cryo-box with Isopropanol at −80°C and after 24h transferred to liquid Nitrogen for storage.

5.3 Embryonic Stem Cell Culture

5.3.1 Culturing Embryonic Stem Cells

Embryonic Stem Cells (ESC) were in general cultured in gelatinized dishes on a layer of mitosis inactivated murine embryonic fibroblasts (MEFs or MEF cells), also referred to as feeder cells. MEFs help to maintain the pluripotency of ESC by secreting Leukemia Inhibitory Factor (LIF). For this study feeder cells were derived from either wildtype CD1 mice or transgenic DR4-mice carrying resistance against Hygromyicin, Neomycin, Puromycin and 6-Thioguanin. If not described differently ESCs were cultured on CD1-feeder cells. A day before seeding ESCs the dishes were coated with 0,1%Gelatine and inactivated feeder cells were plated. The ESCs were plated after the feeder cells were attached to the bottom and reached 80% confluence (4-12h). ESCs grow rapidly and divide every 18-24h. For this reason the medium was changed every day.

If ESC colonies are too big or too close to each other, cells start to differentiate. This can be observed by irregular shape and structure of the colonies. To prevent differentiation of ESC colonies the cells were split every 48h. To reduce stress medium was changed 2h before trypsinization.

5.3.2 Preparation of Murine Embryonic Fibroblast

For preparation of MEFs E13.5-14.5 mouse embryos were dissected from the uterus and extra-embryonic membranes and placed into 1xPBS. The head and soft tissue were discarded and the remaining carcasses were washed in 1xPBS and minced into cubes

about 2-3mm with a pair of fine scissors. The cubes were transferred into a 50ml falcon with 5-20ml Trypsin-EDTA depending on the number of embryos. The cubes were further minced by pipetting and then incubated for 10min at 37°C. Pipetting further minced the tissue cubes. Adding two volumes of Feeder Media to the cells stopped the reaction. To achieve a single cell suspension, cells were mixed by pipetting. The cells were plated on a 15cm dish (2-3 embryos per dish) and incubated ON. The next day an aliquot of medium was taken for the Mycoplasma-test that was performed with the Myoalert Detection Kit according to the specification of the manufacturer. The cells were cryopreserved. When MEFs reach confluence, further proliferation was inhibited by mitosis inactivation with Mytomycin treatment. Mytomycin C inhibits DNA synthesis by CpG cross-linking and in this way prevents cell division without further effects on protein synthesis and secretion. After 3h Mytomycin was removed, cells were washed thrice with 1xPBS and cultivated for another day. The inactivated MEFs were cryopreserved.

5.3.3 Transfection of ESCs – CRISPR/Cas9 System

ESCs (wt-line G4) were thawed from vials and seeded on two 6cm dishes in two different concentrations: 70% and 30%. After two days the dish with 80% confluence was trypsinized. The cell number was determined with a hemocytometer and 400.000 cells were plated on one well of a 6-well plate. After one day of recovering the ESC-Media was replaced with 1.75ml ES Media without Penicillin and Streptomycin. All reagents were warmed at RT 30min before the transfection. For the transfection 125µl DNA-solution (Optimem containing 5–8µg of each ligation product) and 125µl FuGENE-solution (100µl Optimem, 25µl FuGene) were combined and incubated for 15min at RT to enable the formation of particles. The transfection mixture was added drop-wise to the cells. The cells were incubated for 12h in the cell incubator at 37°C. Replacing the media with normal ESC Media stopped the reaction. The cells were allowed to recover for 24h. After recovery, ESCs were split to four 6cm-dishes with DR4-feeder layer. The selection started directly by adding Puromycin (2µg/ml final concentration) to the ESC Media. The selection was stopped after 48h by switching to normal ESC Media and the cells were allowed to recover for 4–6 days, depending on the occurrence of colonies.

5.3.4 Electroporation of ESCs – Sleeping Beauty System

G4 cells were grown for 2 days in ES+LIF, trypsinized and counted prior electroporation. After washing in PBS, a pellet of $9x10^6$ cells was suspended in 0.8ml PBS. The cells were incubated with 30µg of the linearized construct for 5min at RT. After gentle mixing, the reaction was transferred to a Gene Pulser Cuvette (Bio-Rad #165-2088). Electroporation was performed with 240V and 500µF pulse. The cells were diluted in 12ml ES+LIF and seeded on 4x 6cm dishes containing Neo-MEF feeder cells. Selection started after 36h by applying Neomycin (day 1-2: 200µg/ml, after day2:

250µg/ml) to the ESC Media. During the selection period dead cells are removed by washing the plates with PBS prior feeding. Selection was stopped after 6-8 days and clones were picked in a 96-well plate.

5.3.5 Picking of ESC Clones

Each ESC colony represents one clone, which potentially includes the desired genetic modification. ESC colonies can be picked when they are visible with a stereo microscope. One day before picking ESC clones 96-well-F-bottom-plates were coated with 0.1%Gelatine and feeder cells were seeded. The ESC media of the growing ESCs was changed 2h before picking. Dishes with ESC colonies were washed twice with 1xPBS, then 3ml 1xPBS were added. Colonies were picked with the 10µl pipette under a microscope and single colonies were transferred to one well of a 96-well U-bottom-plate containing 30µl Trypsin-EDTA. The U-bottom-plates were incubates for 10min in the incubator after the colonies have been picked (but not more than 10 minutes after picking). Then 60µl ESC Media was added to each well to stop the trypsization. The colonies were suspended by pipetting and transferred to the 96-well-plates with the prepared feeder layer.

5.3.6 Split and Freeze in 96-Well-Plates

The clones that had been picked needed to be frozen and genotyped. For this purpose the 96-well-plates containing the clones were split in three parts after two or three days depending on their size. Two parts were frozen in two U-bottom-plates while one part was kept in ESC Media without LIF. The cells in this 'DNA-plate' were allowed to differentiate since they were used for DNA isolation and genotyping only. Before splitting two 96-well-U-bottom-plates with 50µl 2xPlate Freezing Media were prepared. The ESCs in 96-well-plates were trypsinized, but trypsinization was stopped by adding 100µl Bicarbonate free Media containing 20%FCS to each well. After trypsinization, 50µl of the cell suspension was added to each of the previously prepared 96-well-U-bottom-plates. These 'freezing-plates' were sealed with autoclave tape, packed in polystyrene boxes and kept at −80°C. After two days the freezing plates were stored at −80°C without the polystyrene box. The remaining 50µl in the initial clone plate were filled up with 200µl ESC Media without LIF ('DNA-plate'). The medium of these cells was exchanged every second day until they were harvested at 90% confluence for DNA extraction.

5.3.7 Thawing and Expansion of Cells from 96-Well-Plates

ESCs in 96-well-U-bottom-plates were thawed by adding 100µl pre-warmed ESC Media to single wells. The cell suspension was then transferred to a 15ml falcon containing 1ml ESC Media and centrifuged at 1000rpm for 5min. After discarding the supernatant the

cell pellet was re-suspended in 200µl ESC Media and seeded on a 96-well-F-bottom-plate containing CD1 feeder. After two days the cells were transferred to a 24-well-plate and after another two days to 6-well-plate. As the cells reached 80% confluence they were frozen in three vials. About 100µl of the cell suspension was used to culture ESCs feeder cell free for confirming the genotype.

5.4 Generation of Transgenic Mice using Diploid Aggregation

All mouse lines of this work were generated by the animal facility (Max Planck Institute for Molecular Genetics, Berlin) via diploid aggregation as described previously[228]. In this method zygotes are isolated from pseudo pregnant mice and cultured until they reach the morula stage. ESCs are injected into the morula and retransferred in a foster animal. The chimeric offspring is carrying wildtype and mutant cells.

5.5 Histological Methods

5.5.1 Micro-CT

For the visualization of skeletal features of adult mice Micro-CT was conducted at the Institute of Human Genetics at Charité (Campus Virchow) with the support of Hardy Chen. The samples were dissected and subsequently stored for at one day in 100% technical Ethanol. Due to the high electron density of the bones, no contrast agent was applied. Forelimbs and Skulls of control and mutant mice (P70) were scanned using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium) with a 0.5mm aluminium filter at 81kV and 124µA. 3D model reconstruction and length measurements were performed with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium). Cross-sections were performed at 10µm resolution. The relative length was determined relative to wildtype controls.

5.5.2 Skeletal Staining

Skeletal staining was used to reveal the structure of every bone in the complete skeleton and to distinguish bone from cartilage in order to determine the ossification status. For this purpose sacrificed mice from the embryonic stage E17.5 were treated with Alcian Blue (cartilage staining) and Alizarin Red (bone staining). All steps were performed at RT as not mentioned else.

First, the skin was disrupted by 1min incubation in a 65°C water bath. Then the outer layer of the skin was removed using forceps from hand and feet. At other parts of the body all skin layers were removed as well as all the inner organs apart from the brain. For adult stages also muscle tissue was partly removed to allow better accessibility. The carcass was fixed in 100% Ethanol ON. The skeleton was then stained with Alcian Blue

Solution for 12-24h, rinsed with and washed again in 100% Ethanol ON. The samples were treated with 1%KOH for up to 1h. Alizarin Red Solution was added for ON incubation to the skeleton (in 0.2% KOH). Further tissue was removed by KOH treatment (0.2% KOH). In general the KOH steps have to be adapted to the amount of tissue left. KOH digestion is stopped when the tissue is transparent. Then skeletons are dehydrated in 30%, 60% and 80% Glycerol for at least 24h per step. They are stored in 80% Glycerol at RT in the dark.

5.5.3 RNA *in situ* Hybridization

Whole Mount RNA *in situ* Hybridization is used to visualize the expression pattern of genes in the whole embryo. Wildtype controls were used in each experiment to reduce the variability of the staining intensity. If not mentioned else all reaction steps were performed in a 12-well-plate (3-4 embryos per well) on a shaker to allow proper homogenization. Embryos were dissected in 1xPBS-DEPC and fixed ON in 4%PFA-DEPC at 4°C for RNA protection. For E17.5 skulls, the skin was removed. Afterwards, embryos were washed twice in PBST and dehydrated at 4°C through a series of Methanol dilutions (in PBST): 25%, 50%, 75% and two times 100%Methanol. The embryos were stored in 100%Methanol at -20°C.

Buffer	Ingredients
6%Hydrogen Peroxide	dilute 30% Hyrdogen Peroxide in PBST
Blocking Solution	20µl Serum, 20µl BSA, add 1ml TBST
H1 Buffer	30ml L1 Buffer, 300µl tRNA, 15µl Heparin
Heparin	100mg/ml 4xSSX-DEPC
L1 Buffer	25ml de-ionized Formamide, 12.5ml 20xSSC pH 4.5, 2.5ml 20% SDS, 50µl Tween 20 add 50ml H_2O-DEPC
L2 Buffer	25ml de-ionized Formamide, 5ml SSC 20x pH 4.5, 50µl Tween 20, add 50ml H_2O-DEPC
L3 Buffer	5ml 20xSSC pH 4.5, 50µl Tween 20 add 50ml H2O-DEPC
NTMT	1ml NaCl 5M, 2.5ml Tris 2M pH 9.5, 500µl Tween 20, add 50ml ddH_2O
PBST	in DEPC-H_2O: 10% 10xPBS, 1% 10%-TWEEN
PBST/Glycine	2mg/ml Glycine in PBST
Proteinase K	20mg/ml stock
Staining Solution	0.9µl NBT, 3.5µl BCIP, add 1ml NTMT
TBS 10x	80g NaCl, 2g KCl, 30g Tris-Base, adjust pH 7.4 with HCl, add 1l ddH_2O
TBST	1ml TBS 10x, 100µl Tween 20, add 10ml ddH_2O
tRNA	10mg/ml RNA type III in DEPC-H_2O, Sigma R6750

Day 1 – Hybridization

Embryos were rehydrated on ice through a series of Methanol dilutions (in PBST): 75%, 50%, 25% and twice in PBST-DEPC. The embryos were bleached in 6%Hydrogen Peroxide (in PBST) for 1h on ice and protected from light. After three 5min steps of washing in PBST at 4°C the embryos were digested with Proteinase K (in Proteinase K Buffer) at RT. Embryos at stages E12.5 and E13.5 were digested with 20µg/ml

Proteinase K for 5 or 8min respectively. E17.5 skulls were digested with 20µg/ml Proteinase K for 10min. The samples were washed twice for 5min in PBST/Glycine, twice for 5min in PBST and fixed for 20min in 4%PFA-DEPC at 4°C, followed by five washing steps for 5min in PBST at 4°C. The embryos were incubated in PBST/L1 Buffer (1:1) for 10min. The pre-hybridization was performed in two steps. First, embryos were incubated in pre-heated L1 Buffer for 10min at 68°C. Then, embryos were incubated H1 Buffer for 90min at 68°C. For hybridization, the DIG-probe was diluted in H1 Buffer (1ml per well) and denatured at 80°C for 10min. The embryos were incubated in H1 Buffer including the DIG-probe at 68°C ON.

Day 2 – Removing of Unbound Probe

The Buffers L1, L2 and L3 were preheated to 68°C.

The embryos were transferred to a fresh 12-well-plate and washed at 68°C: three times for 30min with L1 Buffer, three times for 30min with L2 Buffer and once for 15min with L3 Buffer. After 5min at RT the plate was cooled down. To adjust the pH the embryos were washed three times for 5min at RT in TBST followed by 2h of pre-incubation in Blocking Solution at RT. The embryos were incubated with Anti-Dig (1/3000 Roche #11 093 274 910) in Blocking Solution ON at 4°C.

Day 3 – Removing of Unbound Antibodies

Unbound antibodies were removed by washing the embryos in TBST: five times for 5min at RT, five times for 90min at RT and ON at 4°C.

Day 4 – Staining

Prior staining, the embryos were washed three times for 15min in NTMT at RT. The embryos were transferred to a fresh 12-well-plate containing Staining Solution. The embryos were checked every 30min and the Staining Solution (yellow) was replaced as soon as it turned pink. The Staining Solution was replaced by NTMT to pause the reaction and to remove background. To stop the reaction, the embryos were washed twice for 15min with PBST and fixed for long-term storage in 4%PFA.

5.5.4 X-Galactosidase Assay

Buffer	Ingredients
4%PFA	As described in 4.5
X-Gal staining solution	46.75ml 1xPBS, 1.25ml X-Gal in DMSO (40mg/ml), 0.5ml 500mM $K_3Fe(CN)_6$, 0.5ml 500mM $K_4Fe(CN)_6$

The X-Gal assay was used to detect the regulatory potential of putative enhancer elements *in vivo*. Mouse lines carrying a SB cassette were assayed at the Max Planck Institute. Injections of reporter constructs were performed by Marco Osterwalder (Lawrence Berkeley Nactional Laboratory, California).

The embryos were harvested at E14.5 or E17.5 and dissected in cold 1xPBS. The neckfold of E14.5 embryos was opened and the skin of E17.5 embryos was removed to

facilitate the penetration of the fixation and staining solutions. X-Gal staining solution was applied and the embryos were incubated at 37°C in the dark upon color reaction. Then, the embryos were washed 2x in 1xPBS and stored in 4% PFA at 4°C upon documentation.

5.5.5 Lysotracker – Apoptosis Assay

To assay apoptotic cells during interdigital cell death limbs of E13.5 embryos were dissected in 1xPBS, washed 2x in 1xPBS for 5min and incubated in PBS containing 500mM LysoTracker w Red DND-99 for 1h at 37°C, shaking. The limbs were washed 2x in 1xPBS for 5min on ice and imaged immediately.

5.6 4C-Seq

Buffer	Ingredients
Lysis Buffer	5ml buffer/sample: 500mM Tris pH7.5, 150mM NaCl, 5nM EDTA, 0.5% Nonidet P-400, 1.15% Triton X-100, 200µl proeinase inhibitors/sample, add 5ml H_2O
37%PFA	0.555g PFA in 1050µl 10%FCS/PBS and 15µl 1N NaOH, vortex, dissolve at 99°C for 10min
Crosslinking Buffer	4%formaldehyde in 10%FCS/PBS, prepare fresh
methyl green pyronin	staining solution; Waldeck, Pappenheim, #2C-186

This protocol for 4C library preparation was adapted from [229]. E14.5 mouse limbs were collected and 4C libraries were processed by the established 4C-Seq protocol. According to the quality measurements of the 4C protocol by [230], only samples that reached more than 1 million mappable reads after sequencing and for which most reads mapping in cis (>60 %) relative to the viewpoint were used for analysis. The 4C-Seq profiles displayed in this study show normalized read counts, representing the interaction frequency of a DNA fragment with a chosen genomic region (viewpoint).

Crosslinking and Nuclei Extraction

Hands of E14.5 embryos were prepared in 1xPBS. To obtain a single cell suspension, the tissue was digested in 0.1%collagensae treatment for 10min at 37°C, while disrupting the tissue by pipetting every 2min. The digestion was stopped by adding 5x volume of 10%FCS/PBS. To obtain a single cell suspension, the solution was poured through 40µm cells strainer. Cells were centrifuged at 1100rpm for 5min at RT. The cell pellet was resuspended in 5ml 10%FCS/PBS and fixed with 5ml crosslinking buffer for 10min, rotating. The crosslinking reaction was stopped with 1ml 1.425M glycine on ice.

After centrifugation for 8min at 1500rpm at 4°C, the cell pellet was resuspended in 5ml freshly prepared, cold lysis buffer and incubation for at least 10min on ice. To confirm cell lysis, a 3µl aliquot was mixed on a microscope slide with 3µl of methyl green pyronin staining solution, which stains cytoplasm pink and nuclei blue with blue nuclei. Lysis was completed when nuclei were visible under the microscope. After counting the cells,

nuclei were pelleted by centrifugation at 2000rpm at 4°C, washed with 1xPBS and aliquoted in tubes with 2.5-5x10^6 nuclei. After centrifugation at 2600rpm for 2min, nuclei were snap-frozen in liquid nitrogen. Sample can be stored at -80°C for up to 6month.

Preparation of the 3C Library

The nuclei pellet was resuspended in 360µl of water and mixed with 60µl of 10x restriction buffer. The digestion strategies was designed using Csp6I as the 1st and BfaI as the 2nd restriction enzyme. Samples were mixed at 37°C at 900rpm. Next, 15µl of 10% SDS was added and incubated for 1h with occasional pipetting to dissolve the nuclei aggregates. The remaining SDS was separated from the solution by adding 150µl of 10% Triton X-100. After 1h incubation, 600µl of 1x restriction buffer and 400units of restriction enzyme were added. The samples were incubated at 37°C at 900rpm. Another 200units of restriction enzyme were added after 4h and again after ON incubation. Meanwhile, a digestion control was analyzed by agarose gel electrophoresis. For this a 5µl aliquot of the reaction was mixed with 90µl 10mM Tris (pH 7.5) and 2µl RNaseA (10mg/ml), which was incubated for 1h at 37°C. Chromatin was de-crosslinked by adding 5µl proteinase K (10mg/ml) and incubation at 65°C for 4h. For DNA extraction, 100µl phenol-chloroform were added and samples were mixed by inversion. Phase separation was achieved by centrifugation for 10min at 13200rpm and RT. The upper water phase was transferred into a fresh tube and analyzed on a 1% agarose gel. The restriction enzyme was heat inactivated according to manufacturer's instructions. The sample was transferred to 50ml falcon tubes and 700µl 10x ligation buffer were added. The volume was filled to 7ml with using ddH_2O and 50units of T4 DNA ligase (Thermo Fisher Scientific, #EL0013) were added. The ligation was incubated at 16°C ON. A 100µl aliquot of de-crosslinked DNA was analyzed on an agarose gel to monitor successful ligation. The ligation control was prepared as previously described (for the digest control).

The DNA of the final 3C library was subsequently de-crosslinked and precipitated. For this, 30µl of proteinase K (10mg/ml) were added to the ligation reaction and incubated at 65°C ON. Next, 30µl RNaseA (10mg/ml) were added and the sample was incubated for 45min at 37°C. For DNA precipitated, 7ml phenol-chloroform were added and samples were mixed by inversion. Phase separation was achieved by centrifugation for 15min at 3750rpm and RT. The following reagents were added to the water phase: 7ml ddH_2O, 1.5ml 2M NaAc (pH 5.6), 140µg glycogen, 35ml 100%ethanol. All reagents were mixed and placed at -80°C until the sample was completely frozen (24-48h). The sample was thawed and centrifuged for 20min at 8350xg and 4°C. The DNA pellet was washed with 30ml cold 70%ethanol and centrifuged for 15min at 3300xg and 4°C. The dried pellet was dissolved in 150µl 10mM Tris (pH 7.5). The 3C library was subsequently used for 4C library preparation.

Preparation of the 4C Library

The 3C library sample (150µl) was mixed with 295µl H_2O, 50µl of 10x restriction buffer and 60units of the 2nd restriction enzyme (BfaI). The restriction reaction was incubated at 37°C ON. Using a 5µl aliquot of the sample successful digestion was analyzed by gel electrophoresis (1% agarose gel). After digestion the enzyme was heat inactivated by

incubation following manufacturer's instructions. The sample was transferred to a 50ml Falcon tube. Subsequently, 12.1ml H_2O, 1.4ml 10x ligation buffer and 100units T4 DNA ligase (Thermo Fisher Scientific, #EL0013) were added. The ligation reaction was incubated at 16°C ON. The successful ligation was analyzed on a 1% agarose gel. After confirming DNA ligation the DNA of the final 4C library was precipitated and purified. For this 1.4ml 2M NaAC (pH 5.6), 35µg glycogen and 35ml 100% ethanol were added to the sample and placed at -80°C until The samples was frozen (24-48h). The DNA was pelleted by centrifugation for 45min at 8350xg and 4°C. The pellet was then washed with 30ml 70% ethanol and centrifuged for 15min at 3300xg and 4°C. The dry pellet was dissolved in 300µl 10mM Tris (pH 7.5). The final purification step was carried out using the QIAquick PCR purification kit (Roche, #28106) with one column per $1x10^6$ cell input. The concentration was measured on a Nanodrop and on column purification was repeated for samples with A260/A230 ratios higher than 3.0. The purified 4C libraries were then stored at -20°C.

Inverse PCR and Sequencing of 4C Samples

4C libraries were used as template for inverse PCR. The primers were designed according to the region of interest (viewpoint fragment), e.g. the transcription start site of a gene. The primer sequences for inverse PCR and corresponding digestion strategy used in this study are listed in Materials 4.8.9. Primers were designed as described previously[231]. The read primer 1, next to the 1st restriction site and primer 2, close to the 2nd restriction site contained Illumina TrueSeq adapter sequences at their 5' end. Each 4C template and primer pair was tested for amplification in the linear phase of PCR reaction. For this purpose, different concentrations of template DNA (final concentration: 0.5, 1, 2 and 4ng/µl) were tested in a 25µl PCR reaction and analyzed on a 1% agarose gel. The highest template concentration that resulted in a reproducible band pattern and still amplified in the linear phase of PCR, was selected for the final 4C PCR amplification. A total of 1.6µg of each 4C library was amplified in several 50µl PCR reactions, pooled and purified using the QIAquick PCR purification kit (Roche, #28106). One column was used for the purification of five PCR reactions. Correct PCR amplification was verified on a 1% agarose gel. Samples were multiplexed and sequenced with Illumina Hi-Seq technology according to standard protocols. Sequencing was done by the sequencing core facility of the Max Planck Institute for Molecular Genetics (Bernd Timmermann) or at the Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin (Ulrike Krüger and Jochen Hecht). Up to 35 4C libraries were multiplexed, pooled and sequenced on one lane of a Hi-Seq flow cell using 100bp paired-end or 100bp single-end sequencing kits. On average ten million reads per 4C library were obtained.

PCR conditions for inverse PCR

The amplification from 4C libraries the Expand Long Template PCR System (Roche, #11759060001) was used. The conditions are described in Methods 5.1.4.4.

4C-Seq Data Analysis

The pipeline for 4C data analysis was established in cooperation with Verena Heinrich (Department Computational Molecular Biology, Max Planck Institute for Molecular

Genetics, Berlin). Data processing was performed by Verena Heinrich. The primer sequences were clipped from short sequencing reads and quality assurance was done as previously suggested [230] with a customized Java program and the remaining reads were mapped to the reference sequences NCBI37/mm9 using BWA[232]. Each 4C library fulfilled the following quality criteria: more than one million mapped reads and *cis*/overall ratio of mapped reads ≥ 60%. BedGraph tracks for a normalized number of reads (reads per million) that overlapped with fragments of first restriction enzyme were created in a specified genomic range and smoothed over a specified number of fragments using customized Java programs. The viewpoint and adjacent fragments 1.5 kb up- and downstream were removed and a window of ten fragments was chosen to normalize the data per million mapped reads. BedGraph files of single 4C experiments were visualized as tracks in the UCSC browser. 4C tracks used in figures of this study visualize the following genomic coordinates (chr1:74,000,000-78,000,000 and 74,900,000-75,200,000). For comparing interaction profiles of different samples, the $\log_2$ fold change (ratio) was obtained for each window of normalized reads. Scaling differences during the normalization induced by the extra copy in mutants were corrected by excluding the duplicated part from the mutant as well as from the wt allele. To test the reproducibility of 4C-Seq experiments, biological and technical replicates were compared for the *Ihh* promoter viewpoint. Biological replicates represent the independent preparation of tissues and 4C-libraries whereas technical replicates used the same 4C library but independent 4C inverse PCRs. The correlation (Pearson) indicated high reproducibility between biological replicates, ranging from 0.79 to 0.9 (data not shown). Comparison of technical replicates showed an even higher reproducibility (from 0.89 to 0.94) indicating that tissue preparation and fixation contribute slightly to variation in independent experiments. In all figures a representative result is shown.

6 Results

6.1 The Regulatory Landscape of *Ihh*

6.1.1 4C-Seq Identifies *Ihh* Exclusive Chromatin Interactions

To study higher order chromatin interactions at the *Ihh* locus, publicly available Hi-C data of mouse ESC (40kb resolution) was applied[55]. Based on the directionality index that was developed by Dixon *et al* two TADs, including *Ihh* or *Epha4*, were called. Unlike *Epha4*, *Ihh* is located in a gene dense TAD together with more than 40 genes. The high density of CTCF BSs at this TAD indicates the presence of distinct interaction compartments (sub-TADs) that might separate individual genes and their regulatory landscapes from another to facilitate gene specific interactions[55]. In addition, the Pol II Chip-Seq track reflects the high gene density as well as the activity of genes in the TAD that includes *Ihh* [105].

To investigate the regulatory landscape of *Ihh* in mice, 4C-seq was performed in E14.5 wt hands. 4C-Seq libraries of two biological replicates were generated, using the *Ihh* promoter as viewpoint. The *Ihh* promoter contacts a 250kb window with a moderate basal interaction level and individual strong and dispersed peaks, forming an isolated sub-TAD. This region will be referred to as *Ihh*-TAD form now on (Figure 14). The strongest interaction frequencies were mapped upstream of *Ihh*, mainly in the third intron of its neighboring gene *Nhej1* that will be referred to as *Nhej1*-intron (Figure 15A). Four duplications mapped at the *Ihh* locus have been reported to be associated with human disease[233,234]. Interestingly, these duplications align with the main interaction frequencies of the *Ihh* promoter (*Nhej1*-intron), suggesting the presence of potential *Ihh* specific enhancers at this region.

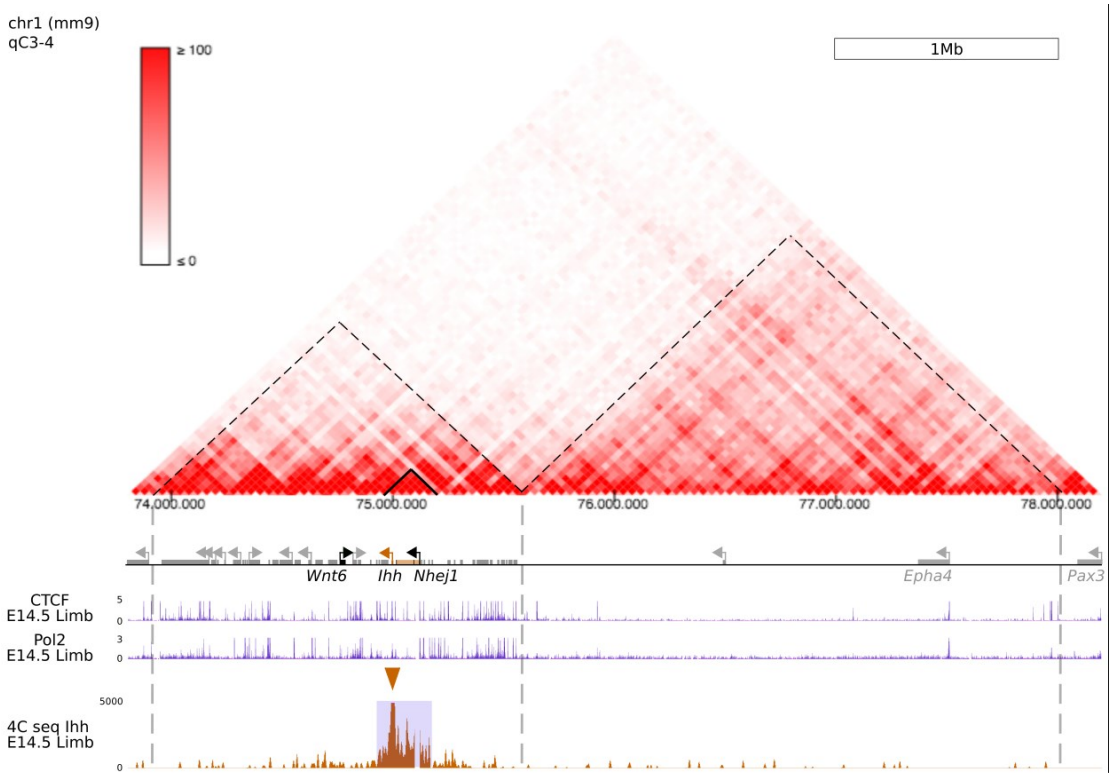

Figure 14. Higher order chromatin interactions at the *Ihh* locus

Hi-C of mouse ESC: strong (red) or weak (white) interaction frequencies. Triangles indicate TADs (dashed line) and the *Ihh*-TAD (solid line). CTCF and Pol II Chip-Seq profiles of E14.5 limbs align with the Hi-C TAD structure. 4C-Seq of E14.5 limbs revealed interaction frequencies of the *Ihh* promoter (arrow) within a 250kb window (*Ihh*-TAD, purple).

6.1.2 SB-Reporter Assay Reveals Presence of Potential *Ihh* Exclusive Enhancers

In order to analyze the regulatory potential of the *Ihh*-TAD *in vivo*, an SB reporter cassette[71], containing a minimal promoter and a *LacZ* reporter gene, was inserted at the center of the *Nhej1*-intron (Figure 15A). The insertion strategy of the SB reporter and the detection via Southern Blot are shown in Supplementary Figure 30. Heterozygous clones were chosen for blastocyst aggregation and mice were crossed to homozygosity.

Homozygous embryos were assayed at two developmental time points (E14.5 and E17.5). The reporter assay revealed a LacZ expression pattern that is consistent with previously published *Ihh* expression, i.e. in the condensations of the long bones, including the ribs, stylopods, zeugopods and autopods at E14.5, as well as in the growth plates and the skull sutures at E17.5. Based on the LacZ expression pattern and the CNV-induced human phenotypes (craniosynostosis, syndactyly and polydactyly), four regions of interest were defined for further enhancer analysis: fingertips, digit condensations, growth plates and skull (Figure 15B). No additional expression pattern that could be linked to interactions of this region with nearby genes was detected

(Supplementary Figure 31). This suggests that the *Ihh*-TAD includes potential enhancers that regulate the expression of *Ihh* exclusively.

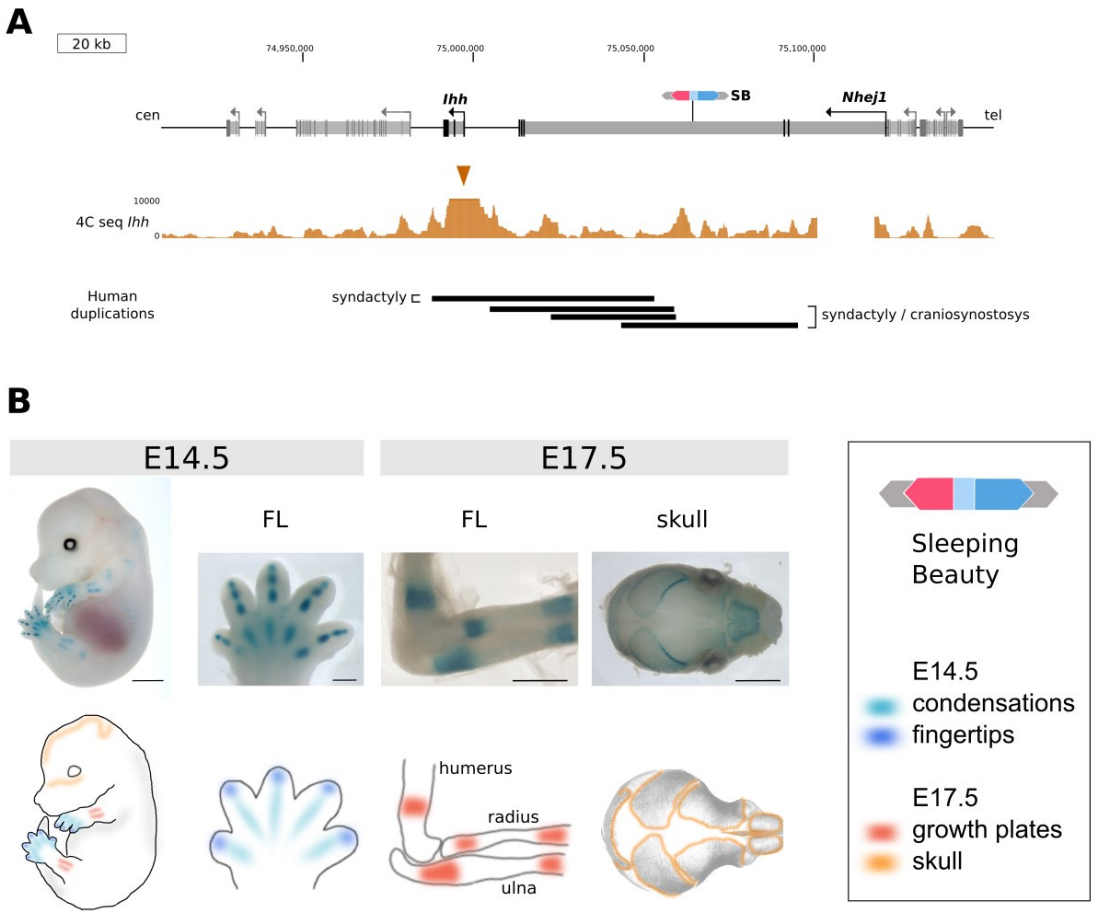

Figure 15. SB reporter assay reveals *Ihh* exclusive enhancers

(A) Schematic overview of the *Ihh*-TAD, introns and exons are indicated in grey and black, respectively (cen, centromeric; tel, telomeric). Magnification of 4C-Seq shows highest interaction frequencies of the *Ihh* promoter (arrowhead) inside the *Nhej1*-intron that co-localize with disease associated human duplications (black bars). (B) Inserted SB reporter (red/blue) reflects regulatory potential of *Ihh*-TAD components with expression domains in: digits (turquoise) and fingertips (blue) at E14.5 and in growth plates (red) and skull (orange) at E17.5. Bars: Embryo 2000µm, FL E14.5 400µm, FL E17.5 1000µm, head E17.5 2000µm.

6.1.3 The *Ihh*-TAD is Populated by Nine Potential *Ihh*-Specific Enhancers

To identify potential *Ihh*-specific enhancers, 4C-Seq interaction profiles as well as publicly available enhancer-associated histone marks (HeK4me1, HeK27Ac), Runx2 (*Ihh*-specific TF) ChIP-Seq and mammalian/amniote sequence conservation profiles were applied[235–237]. Regions with at least 50% sequence conservation that show enhancer-associated histone marks and/or Runx2 binding affinity were chosen to be individually tested in an enhancer-reporter assay[68,74,75]. In combination of these sources nine elements

(i1-9) were selected (Figure 16A). In Table 2 the elements are listed with the according coordinates, sizes and the number of transgenic embryos assayed. The tissue specificity of the elements was evaluated based on the previously defined regions of *Ihh*-specific expression (fingertips, condensations, growth plates and skull). Elements that induced a reproducible reporter signal in at least two embryos were considered as true positive.

Table 2. List of potential regulatory elements tested with LacZ enhancer-reporter assay.

Name	Coordinates [mm9]	Size [kb]	Transgenic embryos [positive|total]	
			E14.5	E17.5
i1	chr1:75,008,008-75,012,847	4.8	5|9	Skull sutures: 3 | 8 Growth plate: 7 | 8
i2	chr1:75,023,290-75,026,536	3.2	0|9	Skull sutures: 1 |10 Growth plate: 3 |10
i3	chr1:75,051,762-75,053,663	1.9	1|7	Skull sutures: 2 | 5 Growth plate: 4 | 5
i4	chr1:75,046,263-75,049,025	2.7	0|6	Skull sutures: 2 |11 Growth plate: 10 |11
i5	chr1:75,053,879-75,060,174	6.1	3|5	Skull sutures: 3 |7 Growth plate: 5 |7
i6	chr1:75,059,085-75,064,020	4.9	10|12	Skull sutures: 1 | 7 Growth plate: 7 | 7
i7	chr1:75,068,299-75,072,430	4.1	4|7	Skull sutures: 1 | 6 Growth plate: 6 | 6
i8	chr1:75,075,786-75,080,268	4.5	4|7	Skull sutures: 5 |10 Growth plate: 10|10
i9	chr1:75,085,302-75,089,234	3.9	0|5	Skull sutures: 3 | 6 Growth plate: 5 | 6

Five of the nine elements (i1, i5, i6, i7 and i8) showed reproducible reporter signals at the two tested developmental stages E14.5 and E17.5 (Figure 16B). The strongest and most prominent reporter signals were induced by the elements located in the telomeric region of the *Nhej1*-intron (i5-i8). The remaining elements (i2, i3, i4 and i9) induced weak but reproducible LacZ signals at E17.5 only (Supplementary Figure 32) and are distributed evenly throughout the *Nhej1*-intron. While most elements showed reporter activity in the growth plates and the skull, LacZ expression was observed in digits for four (i5-i8) and in fingertips for only two elements (i5 and i7). These observations suggest that the individual elements of the enhancer cluster are not fully redundant and that they might act in a modular fashion to meet the needs of different tissues during embryonic development.

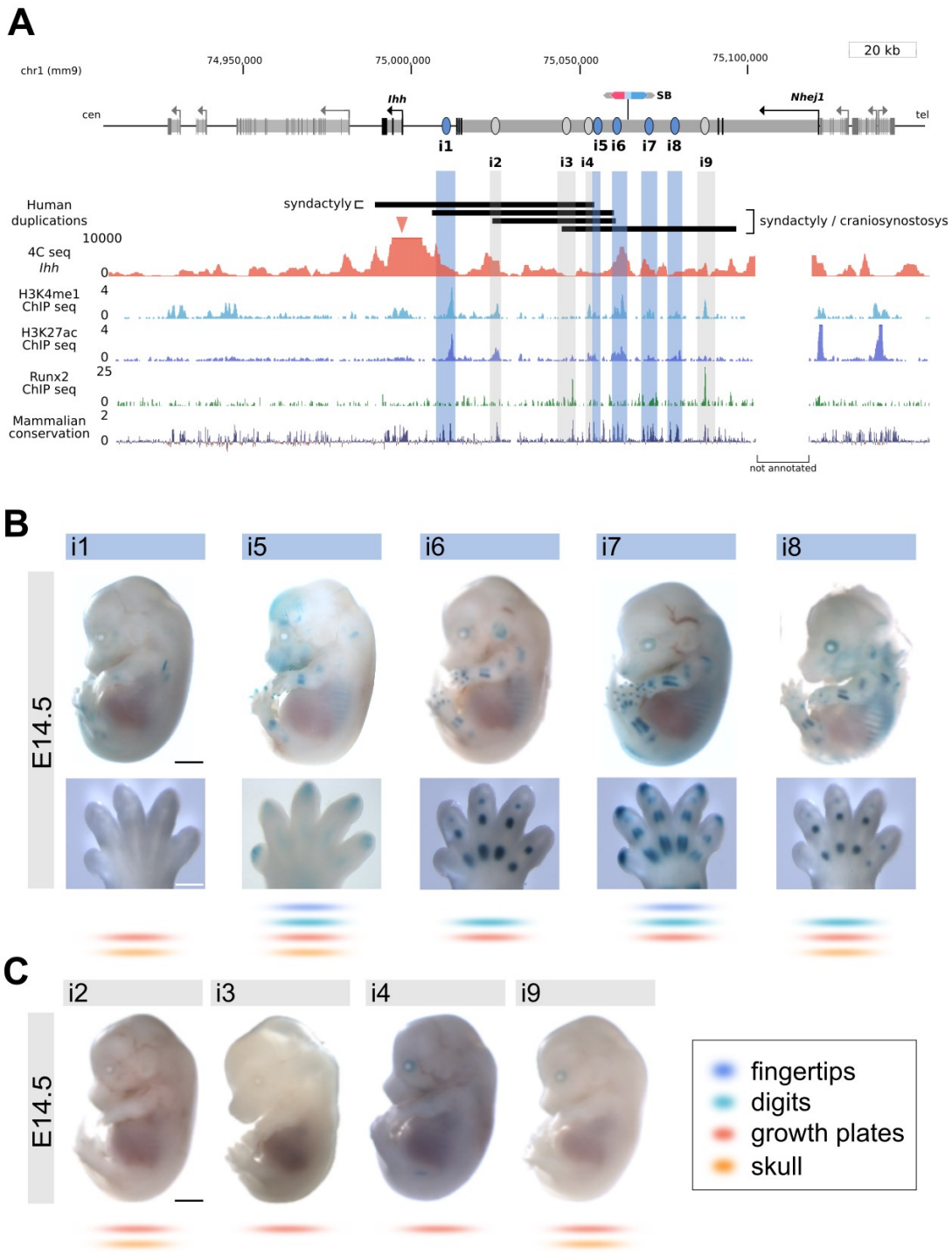

Figure 16. Identification of potential *Ihh*–specific enhancers

(A) Potential enhancer regions were defined using 4C-Seq, enhancer marks (H3K27Ac, H3K4me1) of E14.5 limbs, Runx2 Chip-Seq of osteoblastic cells and mammalian/amniote conservation (human, mouse, rat, orangutan, dog, horse, opossum, chicken). Alignment of disease causing human duplications (black bars). Regions of interest are highlighted (blue: E14.5 positive, grey: E14.5 negative). (B) Enhancer-reporter assay of E14.5 positive enhancers (blue): i1, i5, i6, i7 and i8. Positive expression domains are scored based on E14.5 and E17.5 reporter signals (Supplementary Figure 32) and indicated by color-coding: digits (turquoise) and fingertips (blue) at E14.5, growth plates (red) and skull (orange) at E17.5. (C) X-Gal reporter assay of negative/weak enhancers (grey): i2, i3, i4 and i9. Bars: 2000µm embryo and 400µm hand.

6.2 Generation of CNVs at the *Ihh*-TAD Using CRISVar

CNVs (deletions and duplications) of the non-coding genome have been reported to result in phenotypic effects[238]. To determine the effects of CNV-induced changes in enhancer dosage on skeletal development, CRISVar[102] was applied to modify the enhancer dosage at the *Ihh*-TAD. Deletions and in tandem duplications of various sizes (13-75kb) were generated using two sgRNAs simultaneously to induce double strand breaks at the target region (Figure 17A). Five deletion constructs were generated based on the results of the enhancer-reporter assay, each including a defined set of enhancers. The duplication constructs were engineered to target the *Nhej1*-intron [Dup(int)] and to reconstruct two disease-associated human duplication [Dup(csp) and Dup(syn)]. All constructs are listed in Table 3, with the genomic position, size, included elements and efficiency of genomic rearrangement.

Table 3. Deletion and duplication constructs generated using CRISVar.

Name	Coordinates [mm9]	Size [kb]	Components	Efficiency [%]	
				Del	Dup
Del(i2-3)	Chr1:75,015,710-75,050,992	34	2 enhancers: i2, i3	40.6	0.0
Del(i4-6)	Chr1:75,050,992-75,063,567	13	3 enhancers: i4, i5, i6	44.1	0.0
Del(i7-9)	Chr1:75,063,567-75,091,187	28	3 enhancers: i7, i8, i9	27.1	0.7
Del(i4-9)	Chr1:75,050,992-75,091,187	41	6 enhancers: i4-i9	29.5	0.4
Del(i2-9) / Dup(int)	Chr1:75,015,710-75,091,187	75	8 enhancers: i2-i9	8.3	3.5
Dup(csp)	Chr1:75,005,921-75,060,430	54	5 enhancers: i1-i5	n.a	1.0
Dup(syn)	Chr1:74,989,792-75,054,809	65	1 gene: *Ihh*, 5 enhancers: i1-i5	n.a	2.1

PCR-genotyping revealed high efficiencies of genomic rearrangements for all target regions (Table 3), indicating that this locus is well amenable for DNA modifying enzymes such as Cas9. This further shows that NHEJ repair was not affected by deleting the *Nhej1*-intron or by duplicating the gene body partially, as a functional copy of the gene was preserved at all times. Interestingly, deletion and duplication efficiencies correlated with the size of the target region. While the deletion efficiency was observed to be highest for smaller rearrangements (up to 44%), the duplication efficiency increased with the size of the target region (up to 3.5%). In 50% of all cases, duplications were observed in accordance with deletions, indicating that the gained genomic information originated from the deleted homolog.

To determine the exact position of the rearrangement, breakpoint analysis was performed. Sanger sequencing of the breakpoints revealed minor sequence variations, i.e. single base pair insertions or deletions. The copy number of the genomic rearrangements (wt=2) was determined by qPCR analysis to determine the genotype of the clones and to ensure the completeness of the rearrangement. For clones with a heterozygous deletion [Del(i2-3) and Del(i4-6)], the copy number was reduced by 50%, whereas for clones with a homozygous deletion [Del(i4-9), Del(i2-9)] no signal was detected. Clones with a duplication contained one allele with the deleted and one with the duplicated region and thus, showed no changes in the copy number [Dup(int)/Del(i2-9), Dup(syn) and Dup(csp)] (Figure 17C). Both alleles were separated by mating the founder animals (F0) with wt mice.

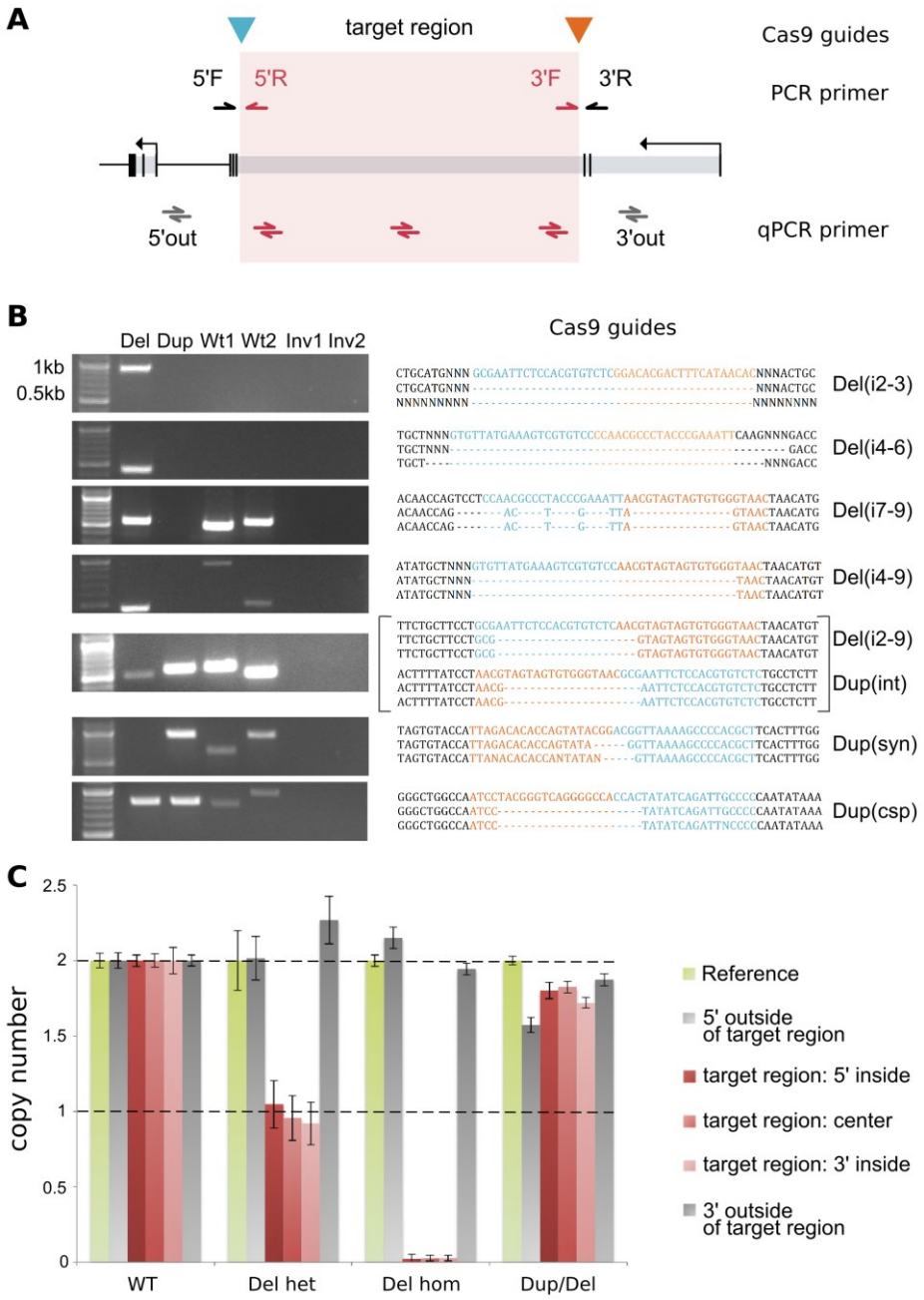

Figure 17. Detection of CNVs generated using CRISVar

(A) Schematic of CRISVar targeting and genotyping strategy. Binding position of Cas9 guides (blue and orange arrowhead). PCR primers and qPCR primer pairs bind inside (red arrow) or outside (black arrow) of the target region (light red). (B) The left panel shows the genotyping PCRs applied to determine the genomic rearrangement of the clones: Del: deletion (5'F+3'R), Dup: duplication (5'R+3'F), Wt1/2: wildtype (5'F+5'R/3'F+3'R), Inv1/2: inversion (5'F+3'F/5'R+3'R). The right panel shows the sequences of the breakpoints at the rearrangement. 5' and 3' guide sequence are indicated in blue and orange, respectively. (C) Copy numbers analysis of the aggregated clones using qPCR. Wt carry two allelic copies, heterozygous deletions one copy and homozygous deletions no copy of the deleted fragment. Clones with one deletion allele and one duplication allele show two copies. The color-coding is indicated in the legend.

6.3 Functional Characterization of the *Ihh* Enhancer Cluster

6.3.1 *Ihh* Expression is Primarily Controlled by an Intronic Enhancer Cluster

To prove the regulatory potential of the enhancer cluster all enhancers located inside the *Nhej1*-intron were deleted [Del(i2-9)]. In order to avoid an impact on *Nhej1* expression, i1 was not included in the functionality study. The skeletal phenotype of homozygous embryos was assayed at E17.5 in parallel with wt and *Ihh* k.o. controls. The phenotype of Del(i2-9) embryos recapitulated and was almost as severe as the previously reported *Ihh* k.o. phenotype[79]. Homozygous pups were not viable and died after birth. They showed strong reduction in bone length, reduced ossification of the skull, absence of cortical bone and fused joints of the autopods (Figure 18C). Whole mount in situ hybridization (WISH) and qPCR analysis of E13.5 embryos revealed, that these abnormalities originated from a drastic reduction of *Ihh* expression by 98-99% (Figure 18C and Figure 21A). These observations confirmed the *in vivo* functionality of the enhancer cluster and further suggest that i1 may as well contribute to *Ihh* expression to a limited extent.

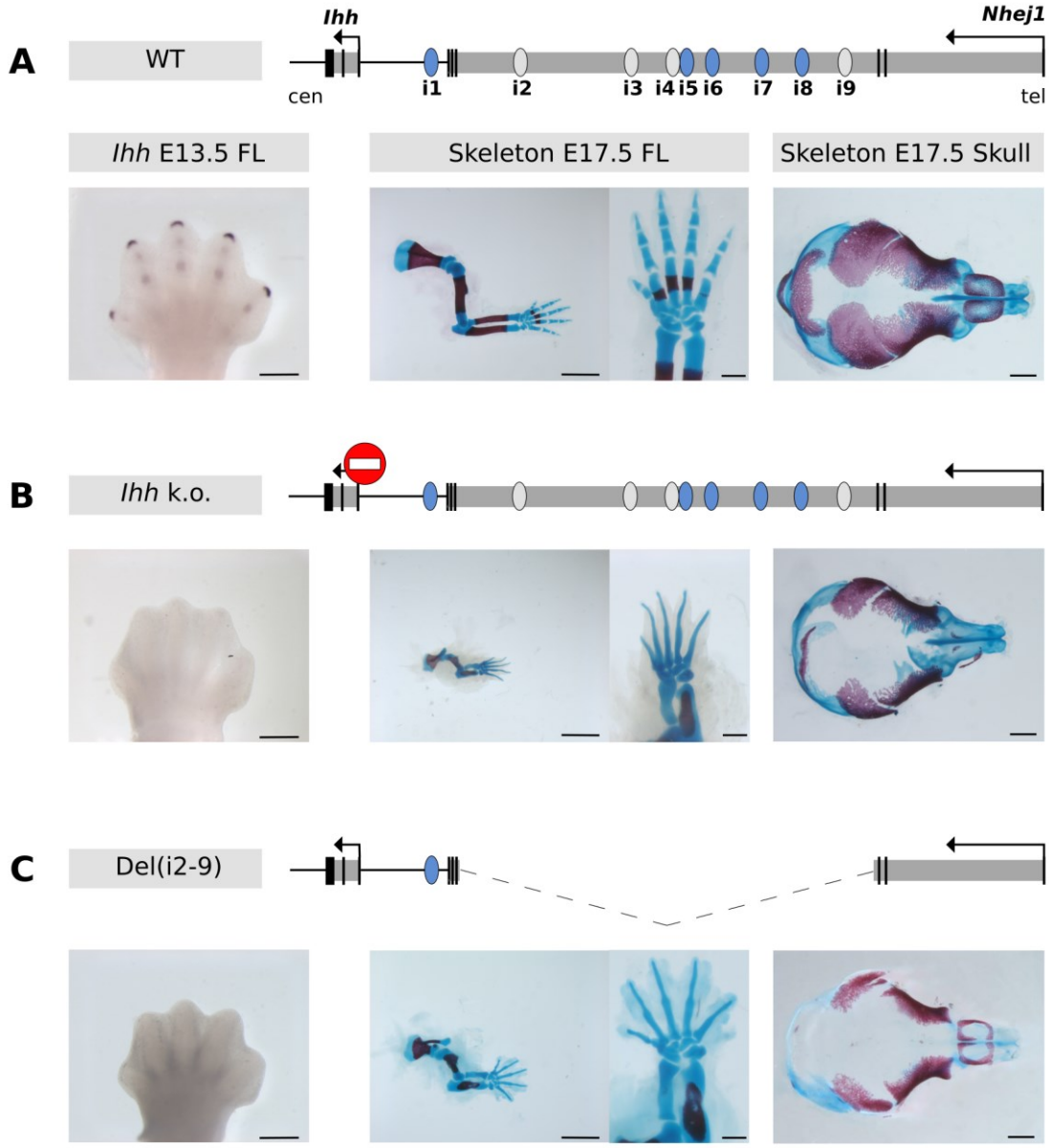

Figure 18. Functional characterization of the intronic enhancer cluster

The schematic of the genotype is shown on top. WISH of E13.5 forelimbs on the left and the skeletal phenotype of E17.5 embryos on the right. (A) wt, (B) *Ihh* k.o. and (C) deletion of the *Nhej1*-intron. cen: centromeric, tel: telomeric. Bars: 400µm (WISH) and 2000µm forelimbs, 500µm autopods, 1000µm skull (skeletal staining).

6.3.2 Consecutive Deletions Reveal Additive Functionality of the Enhancer Cluster

To address the functional redundancy of this cluster, three domains (i2-3, i4-6 and i7-9) were assigned to the *Nhej1*-intron and deleted individually or in combination. The enhancer-reporter assays suggested that individual elements might activate tissue-specific *Ihh* expression at certain developmental time points. The enhancers i2, i3, i4 and i9 showed no LacZ reporter signal at E14.5 and only weak signals at E17.5 in two (i3 and

i4) or one (i2 and i9) of the four assigned tissues. The homozygous deletion of i2 and i3 [Del(i2-3)] resulted in no detectable changes of the *Ihh* expression pattern in E13.5 limbs. The skeletal features of these embryos developed normally (Figure 19A).

In Del(i4-9) embryos six elements of the *Nhej1*-intron were deleted, but i2 and i3 remained intact. In E13.5 forelimbs, *Ihh* expression was reduced by 90%, being comparable with the deletion of the complete *Nhej1*-intron [Del(i2-9)] (Figure 19B and Figure 21A). However, the skeletal features of Del(i4-9) embryos were slightly more developed as compared to Del(i2-9), i.e. showing the formation of joints and increased ossification of the radius and the skull (Figure 19B). These observations verify, that i2 and i3 are dispensable for skeletal development but may contribute to chondrocyte maturation at later developmental stages, as suggested by the LacZ reporter staining. Furthermore, this shows that the most relevant *Ihh* enhancers (i4-9) are localized at the telomeric region of the *Nhej1*-intron.

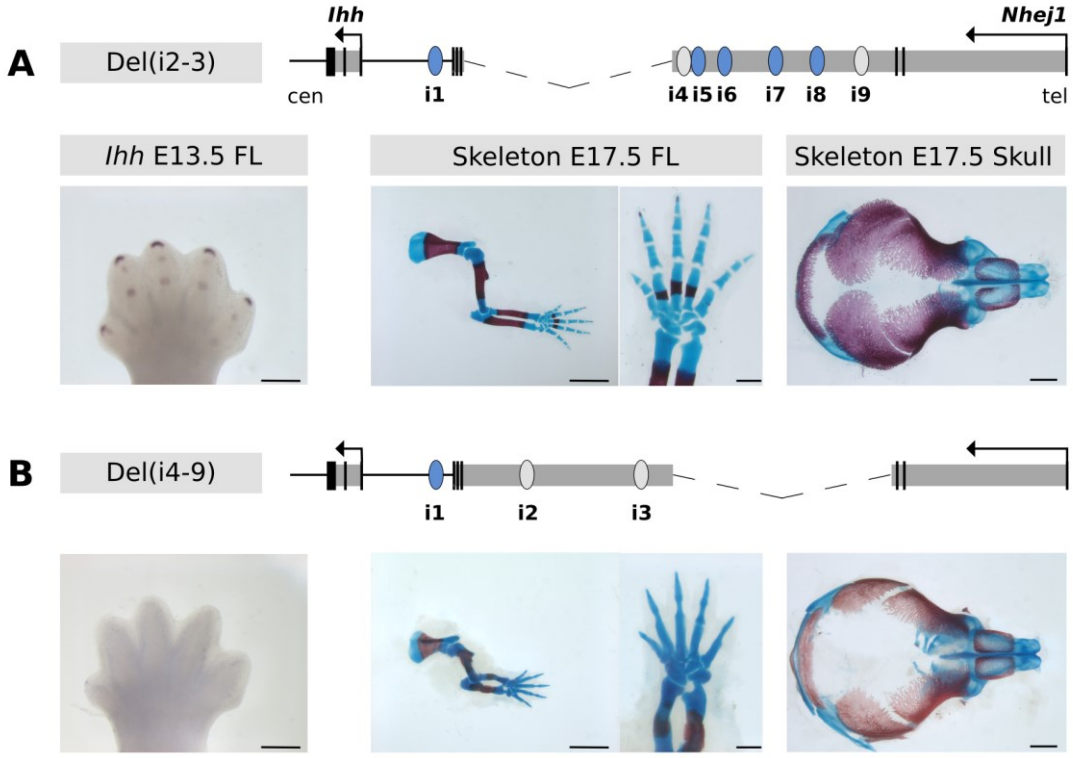

Figure 19. Functional characterization of redundant enhancer domains

The schematic of the genotype is shown on top. WISH of E13.5 forelimbs on the left and the skeletal phenotype of E17.5 embryos on the right. (A) Del(i2-3), (B) Del(i4-9). cen: centromeric, tel: telomeric. Bars: 400µm (WISH) and 2000µm forelimbs, 500µm autopods, 1000µm skull (skeletal staining).

In Del(i4-6) , the central part of the *Nhej1*-intron was deleted. In homozygous embryos, minor skeletal effects were detected at E17.5, mainly affecting the ossification of the metacarpals, the size of the phalanges and the ossification of the skull (Figure 20A). In Del(i7-9) embryos skeletal development of the limbs was not affected. However, ossification of the skull was slightly delayed. To quantify tissue-specific effects on skeletal development, Micro-CT analysis of Del(i4-6) and Del(i7-9) adult mice was performed. Only Del(i4-6) mice showed a significant reduction in length of the ulna (10%) and of the nasal suture (15%) compared to wt. The length of other cranial sutures remained nearly unaffected (Supplementary Figure 33). Strong differences in the volume of the distal phalange (third phalange of digit 3) were determined for both deletion lines. In Del(i7-9) mice, the phalange volume was decreased by 20% and in Del(i4-6) mice by 35% (Figure 20C).

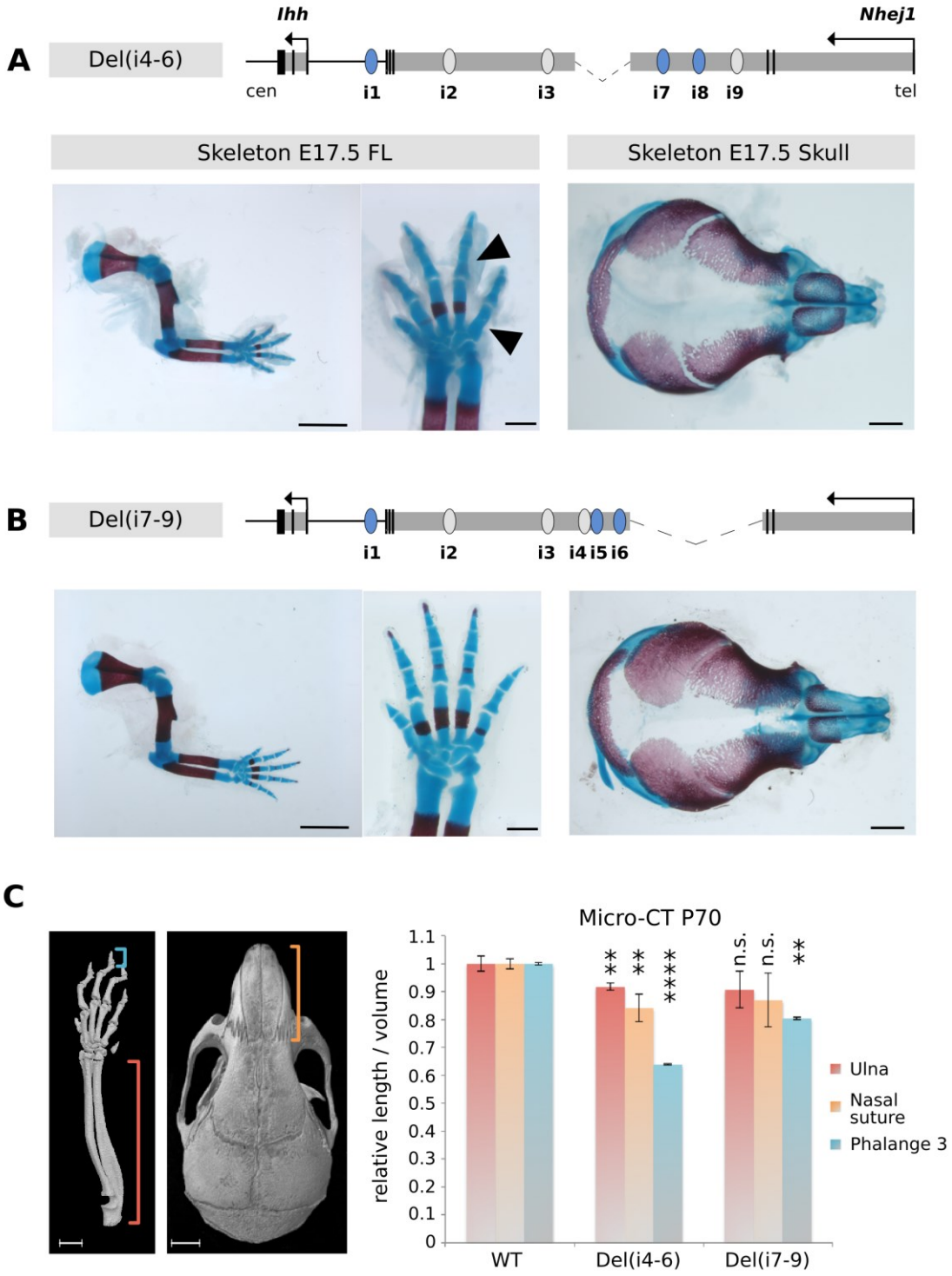

Figure 20. Phenotype analysis of the central and telomeric enhancer domains

The schematic of the genotype is shown on top and the skeletal phenotype of E17.5 embryos at the bottom. (A) In Del(i4-6) mutants, ossification of the phalanges and metacarpals is delayed (arrowheads). (B) Del(i7-9) embryos show normal ossification at E17.5. cen: centromeric, tel: telomeric. Bars: 2000µm FL, 500µm autopods, 1000µm skull. (C) Micro-CT analysis and length/volume measurements, normalized to wt. The measured region is indicated in the schematic on the left. *P* values were calculated with two-sided Student's t-test (n≥3).

Quantitative expression analysis of Del(i4-6) and Del(i7-9) embryos was performed in hands (E13.5), growth plates and skulls (E17.5). All embryos were assayed in parallel with Del(i2-9), Del(i4-9) and wt controls. The deletion of all six elements [Del(i4-9)] showed a strong reduction of *Ihh* expression in hands (90%) and was almost absent in growth plates and skulls (98%), being comparable with Del(i2-9) mutants in which *Ihh* expression was reduced by 98-99% in all tissues. *Ihh* expression was significantly reduced in the hands, the growth plates and the skull of Del(i4-6) and Del(i7-9) embryos by 70% and 60-65%, respectively (Figure 21A). WISH of E17.5 skulls and E13.5 forelimbs confirmed the reduction of *Ihh* expression in both transgenic lines (Figure 21C and D). Interestingly, Del(i4-6) embryos showed a strong reduction of *Ihh* expression throughout the limb. In Del(i7-9) limbs, however, *Ihh* expression was reduced in the proximal condensations but was not affected at the fingertips. This observation is consistent with the phenotype of Del(i4-6) mice that showed a strong developmental impairment of the distal autopods as shown by Micro-CT analysis of phalange 3. As i5 is the only element included in Del(i4-6) that showed LacZ reporter expression in the fingertips, this element might be a major regulator of *Ihh* expression in that tissue.

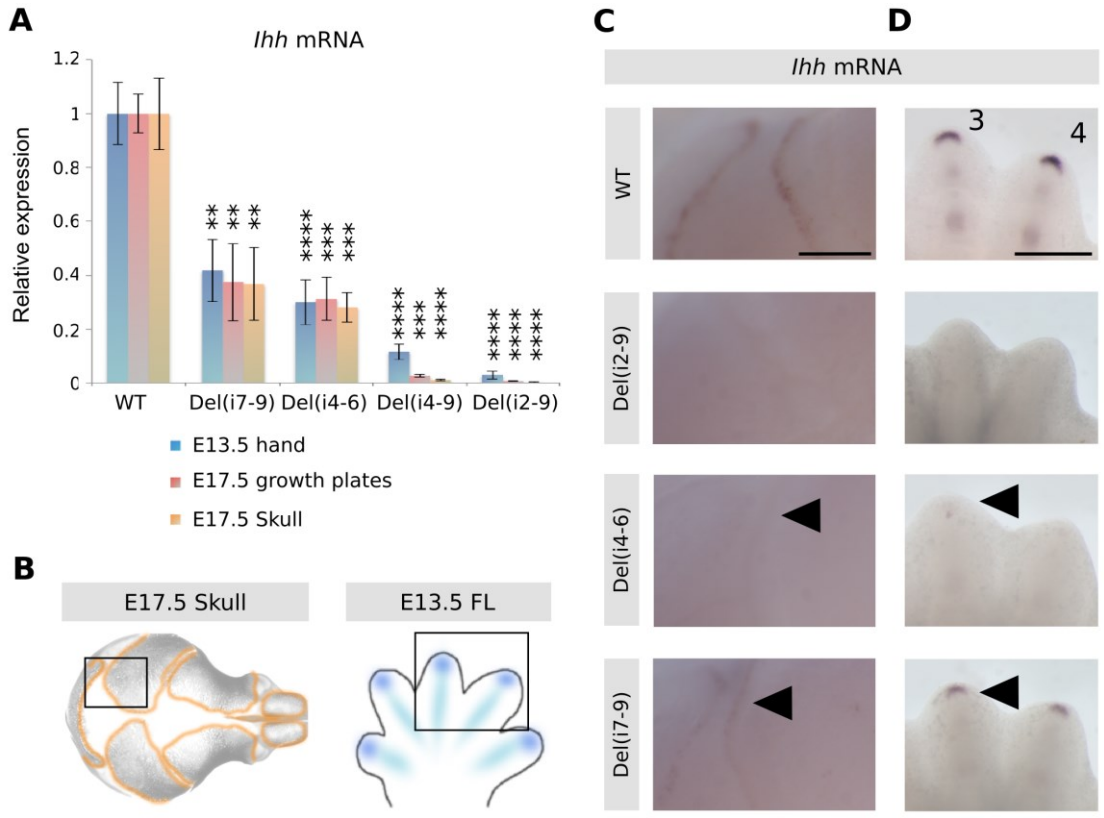

Figure 21. Expression analysis of the central and telomeric enhancer domains

(A) qPCR analysis of *Ihh* expression. Analyzed tissues are indicated. *P* values were calculated with two-sided Student's t-test (n≥3). (B-D) WISH of E17.5 skulls and E13.5 forelimbs (FL). Schematic indicates magnified region (B and C). In Del(i4-6) embryos, *Ihh* expression is strongly reduced along the skull sutures and in the complete forelimb. In Del(i7-9), *Ihh* expression remains present at the fingertips (arrowheads). Bars: 500µm skull, 400µm FL.

Although, the deletion of one of the two domains (i4-6 or i7-9) induced a reduction of *Ihh* expression that resulted in a minor skeletal phenotype, a severe phenotype was induced only if both domains were deleted (i4-9). This demonstrates that a certain threshold of *Ihh* expression is necessary to facilitate proper skeletal development.

In summary, these results highlight the spatio-temporal specificity of individual *Ihh* enhancers and further demonstrate that *Ihh* expression is controlled by a cluster of enhancers that act in an additive fashion.

6.4 Duplications of *Ihh* Enhancers Reveal Tissue-Specific Pathomechanisms

6.4.1 Functional Characterization of Increased Enhancer Dosage

To investigate the effects of increasing enhancer dosage on *Ihh* expression three duplications with a defined number of enhancers were generated using CRISVar. Dup(int) comprises the intronic enhancers (i2-9), which were identified to be the main *Ihh* regulators. Dup(csp) and Dup(syn) were generated to reconstruct the previously identified disease-associated human duplications craniosynostosis Philadelphia type and syndactyly type 1[196,197], comprising the elements i1-5 or *Ihh* and i1-5, respectively. To quantify the effects of the duplications on *Ihh* expression, qPCR analysis was performed in hands at E13.5 and in growth plates and skulls at E17.5 (Figure 22). Dup(csp) embryos showed a modest upregulation in the growth plates and skull of approximately 1.5-/1.8-fold, but not in hands. In Dup(int), qPCR analysis revealed a strong upregulation of *Ihh* expressions in all tissues, i.e. 2-fold in growth plates and up to 5-fold in hands and skulls. To investigate the effect of the additional copy of the *Ihh* gene, heterozygous [Dup(syn)/+] and homozygous [Dup(syn)] embryos were analyzed for this line. In Dup(syn)/+ mice, *Ihh* expression was moderately upregulated in hands and skulls (1.7-/1.5-fold), but not in the growth plates. In Dup(syn) mutants an increase of *Ihh* expression was observed in all tissues, being double as high as compared to Dup(syn)/+ mutants but not as high as compared to Dup(int) (Figure 22). These results indicate that increased gene dosage has a quantitative effect only, while increased enhancer dosage has quantitative as well as site-specific effects on gene expression.

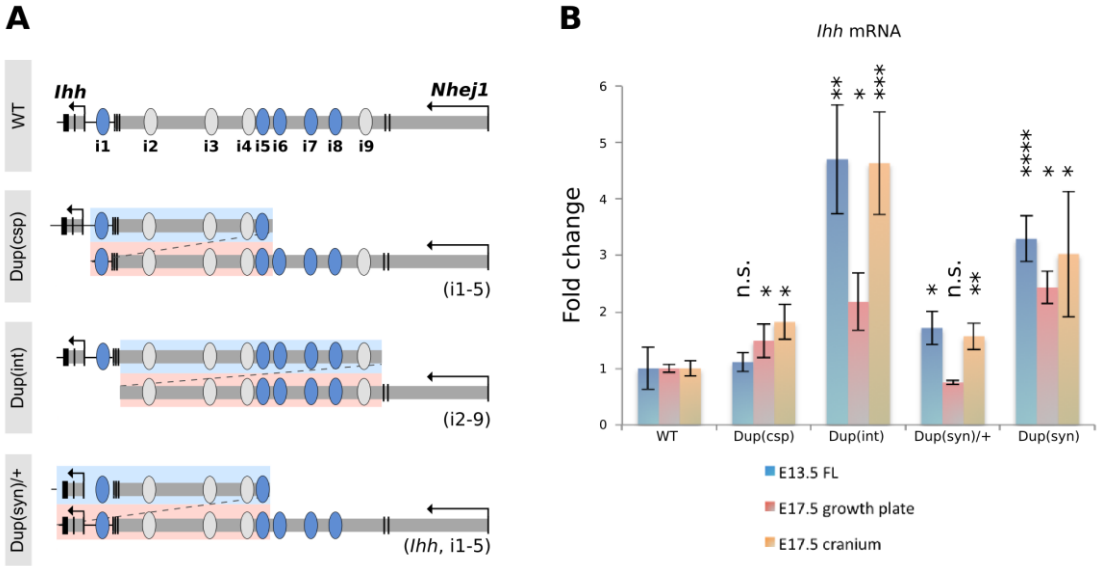

Figure 22. Expression analysis of increased enhancer dosage

(A) Schematic of duplications generated using CRISVar. Dup(csp) and Dup(syn) reconstructed disease-associated human duplications of craniosynostosis and syndactyly patients, respectively. (B) qPCR analysis of *Ihh* expression. Analyzed tissues are indicated by color coding. *P* values were calculated with two-sided Student's t-test (n≥3).

6.4.2 Enhancer Duplications Induce Diverse Phenotypic Effects

6.4.2.1 Bone Length and Skull Phenotypes

To investigate the effect of increased *Ihh* expression on bone length, i.e. ulna and skull, Micro-CT analysis was performed at P70 (Figure 23). In contrast to the deletion constructs, no significant changes were observed by length measurements of ulna, skull and nasal suture in any transgenic line (Supplementary Figure 34). Cross-sections of the metopic suture revealed craniosynostosis in Dup(csp), in consistence with the human phenotype[196,197]. Interestingly, in Dup(syn)/+ skulls that showed slightly lower *Ihh* expression levels as compared to Dup(csp) mutants, no craniosynostosis was detected. However, homozygous Dup(syn) mice showed fused sutures. Suture fusion was most pronounced in Dup(int) mice, suggesting a dosage effect as Dup(int) mice showed highest expression levels of *Ihh* in the skull (Figure 22).

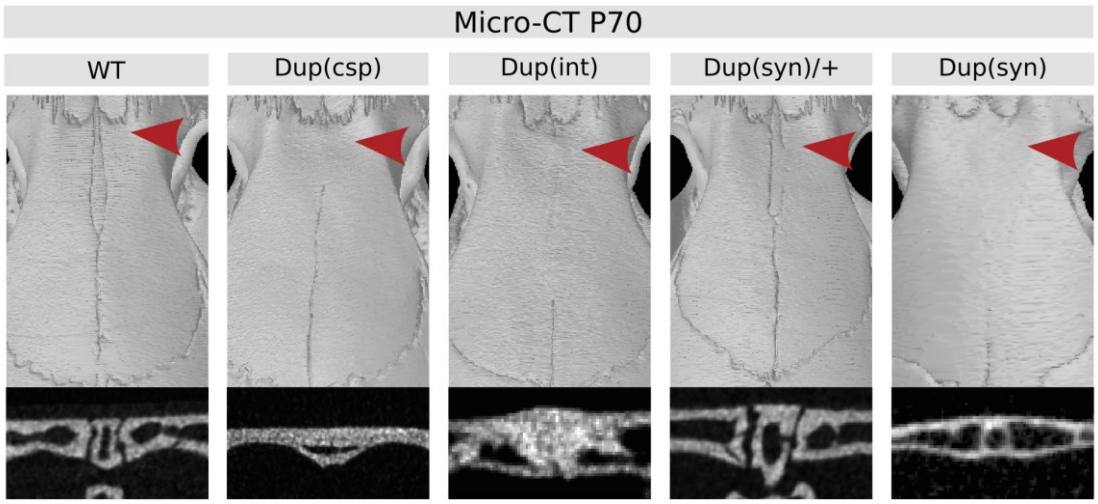

Figure 23. Micro-CT analysis of skulls in duplication lines

Micro-CT analysis of P70 mutants. Magnification of the metopic suture (red box) and cross-sections show suture fusion in homozygous mutants, Dup(csp), Dup(int) and Dup(syn), but not no fusion in wt mice and heterozygous Dup(syn)/+ mutants. Red arrowheads indicate the region of the cross-section. Bar: 2cm.

6.4.2.2 Limb Phenotypes

To investigate skeletal malformations of the limbs, mice were examined at P7 and E17.5. Dup(csp) and Dup(int) showed no skeletal abnormalities in fore- or hindlimbs (Supplementary Figure 35). However, Dup(syn)/+ and Dup(syn) mice showed complete penetrance of cutaneous syndactyly of the hands (digits 2-5) and feet (variable) at P7 (Figure 24A). Skeletal stainings revealed normal development of digits and joints, but broadening of the terminal phalanges, being consistent with the human phenotype [233,234]. In addition, preaxial polydactyly was observed in Dup(syn)/+and Dup(syn) mice with 50% penetrance (Figure 24B). Since heterozygous and homozygous mice of this line showed the same penetrance and severity in the limb phenotypes, all following

experiments were performed with heterozygous mutants [Dup(syn)/+] following the instructions of the german authorities for laboratory animal science (Lageso).

WISH of Dup(syn)/+ hindlimbs showed an increase and broadening of the *Ihh* expression pattern in the distal digit condensation as well as the formation of an additional digit at the anterior side of the limb. In E12.5 hindlimbs, a prominent increase of *Ihh* expression was observed in the distal zeugopod, which might interfere with the hedgehog gradient and thus could induce the formation of an additional digit. These observations show that enhancer duplications can result in phenotypes that do not strictly correlate with the number of duplicated elements. The diversity of phenotypes might rather be explained by changes in the spatio-temporal expression of *Ihh*, induced by the duplication of tissue specific enhancers.

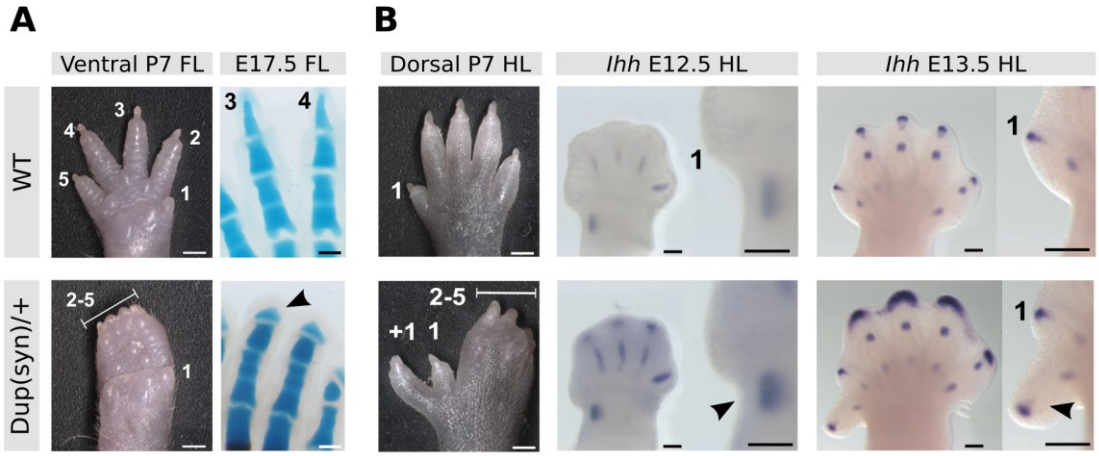

Figure 24. Dup(syn)/+ mutants phenocopy syndactyly and polydactyly

(A) Phenotype analysis of P7 mutants shows cutaneous syndactyly of the digits 2-5 in forelimbs. Skeletal staining reveals broadening of the terminal phalange at E17.5 (arrowhead). (B) P7 hindlimbs show cutaneous syndactyly as well as preaxial polydactyly. WISH of E12.5 and E13.5 hindlimbs. *Ihh* overexpression at the anterior side of the limb (arrowhead) might induce the formation of an additional digit. Bars: 1000µm P7, 200µm E17.5 skeletal staining; E12.5 and E13.5 WISH.

6.4.3 *Ihh* Misexpression in the Distal Limb Results in Apoptosis Suppression and Syndactyly

To investigate the pathomechanism resulting in syndactyly, E13.5 forelimbs were assayed using WISH (Figure 25A). In Dup(csp) limbs, a mild increase and broadening in expression of *Ihh* and its downstream targets *Ptc1* and *Gli1* was observed, an effect that was more pronounced in Dup(int) embryos. In Dup(syn)/+ mutants, misexpression of *Ihh*, *Ptc1* and *Gli1* was most prominent in the distal limb condensations and extended into the interdigital mesenchyme, showing a clear misexpression of the IHH pathway.

Apoptosis of the interdigital mesenchyme has been reported to induce digit separation[177]. To determine the effect of the duplications on interdigital apoptosis Lysotracker

apoptosis assay was applied (Figure 25B). With Lysotracker, apoptotic cells are labeled by fluorescent probes that are highly selective for acidic organelles. Fluorescence of Dup(csp) and Dup(int) mice showed strong fluorescent signals in the interdigital mesenchyme, comparable with wt limbs. In Dup(syn)/+ limbs, apoptosis was absent in the distal interdigital mesenchyme, showing an overlap with the *Ihh* pathway misexpression. This indicates that the duplication-associated syndactyly might be induced by absent apoptosis in the distal interdigital mesenchyme, which might be induced by the over- and misexpression of *Ihh* in this region.

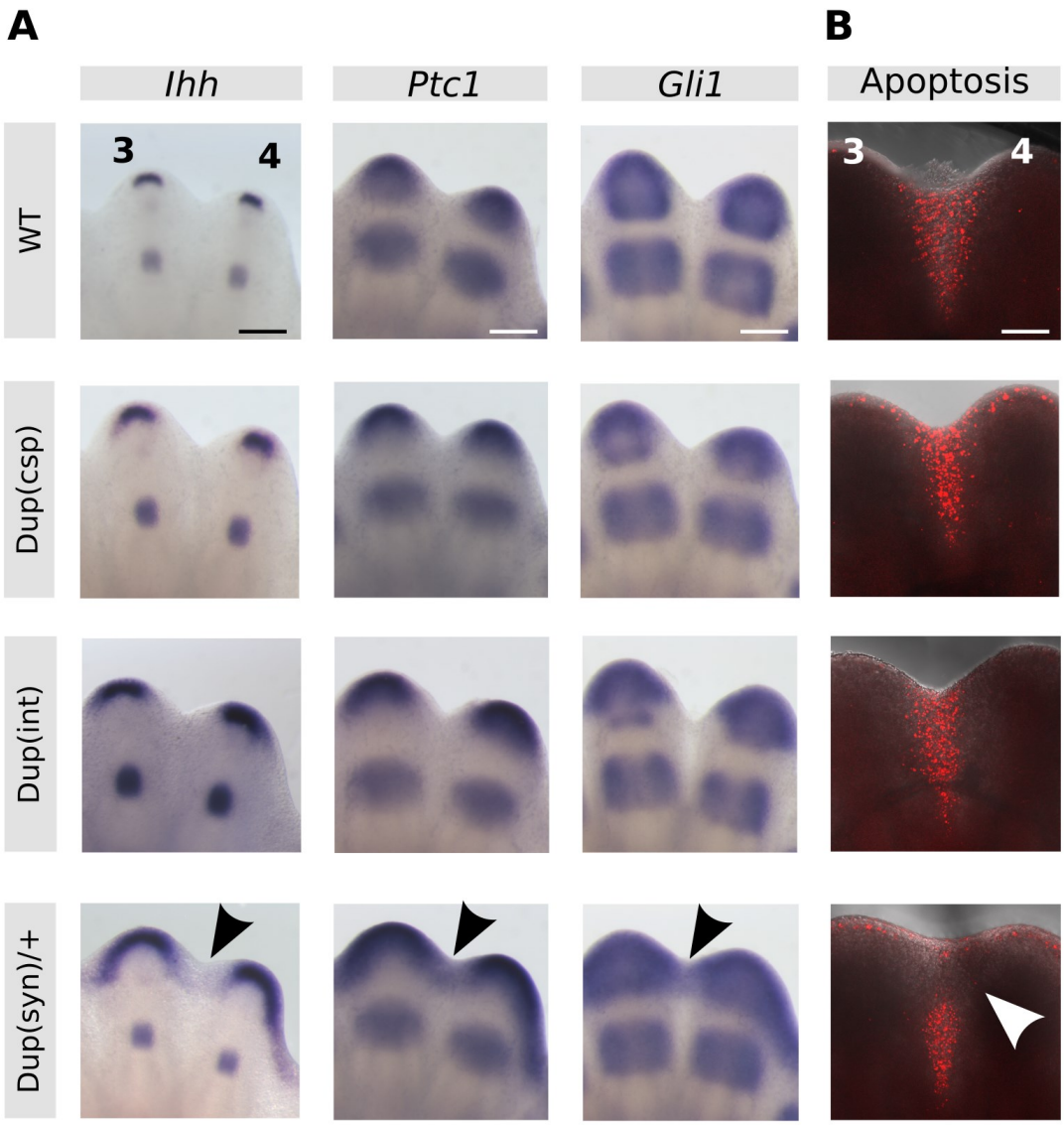

Figure 25. Expression analysis and apoptosis assay of duplication mutants

(A) WISH of *Ihh* and the downstream targets *Ptc1* and *Gli1* in E13.5 forelimbs shown as a magnification of the expression pattern (digits 3 and 4). Dup(csp) and Dup(int) mutants show overexpression of *Ihh*, *Ptc1* and *Gli1*. Dup(syn)/+ mutants reveal overexpression and misexpression of all genes, protruding into the interdigital mesenchyme (black arrowhead). (B) Lysotracker apoptosis assays in E13.5 forelimbs. In Dup(syn)/+ forelimbs, apoptosis is absent in the distal interdigital mesenchyme (white arrowhead). Bars: 200µm.

To test the effect of *Ihh* misexpression on apoptosis induction, WISH of the main regulators of interdigital apoptosis was performed in Dup(syn)/+ fore limbs and wt controls (Figure 26). Interestingly, no increased or prolonged expression of the apoptosis-repressing factor *Fgf8* as well as no reduction of the apoptosis-inducing factors *Aldh2*, *Bmp2, 4* and *Bmp7* was observed that could be linked with the syndactyly phenotype[189,192,193]. In consistency with the increased proliferation of the terminal phalanges, a broadening of the expression patterns of *Bmp4* and *Nog* was observed in the distal limb[239]. These observations suggest that upregulation and misexpression of *Ihh* and its downstream targets *Ptc1* and *Gli1*, has the potential to inhibit apoptosis in the distal interdigital mesenchyme and that IHH signaling can act independently of the major apoptosis regulators, resulting in syndactyly and broadening of the terminal phalanges.

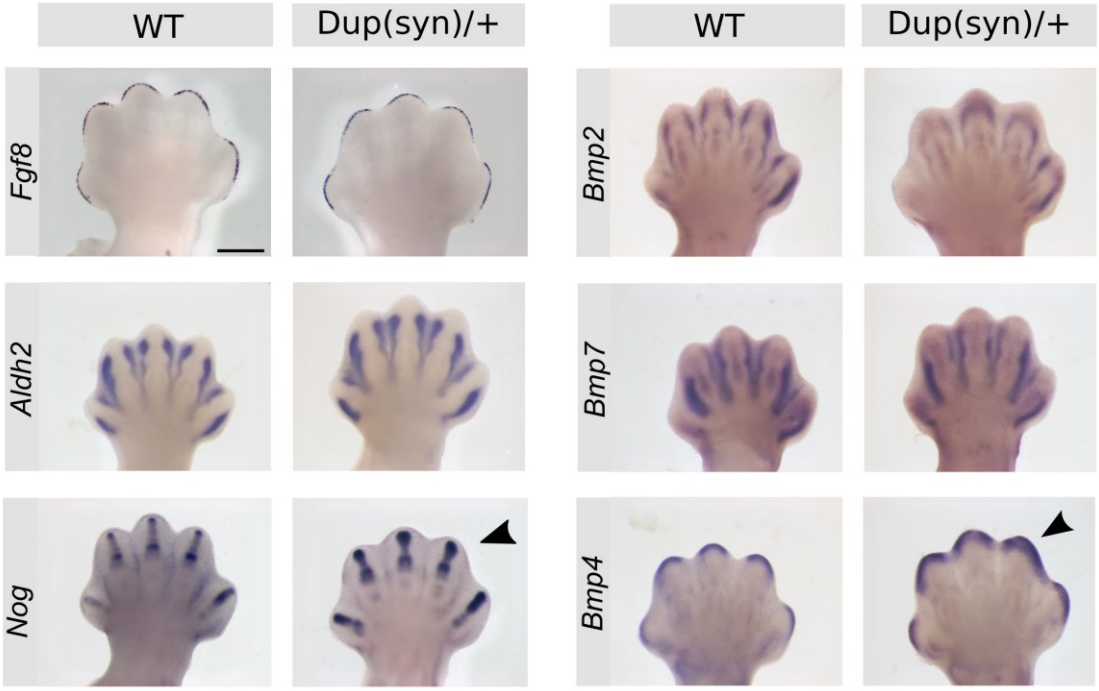

Figure 26. Expression analysis of apoptosis repressing and inducing factors

WISH of apoptosis repressing (*Fgf8*) and inducing factors (*Aldh2, Bmp2, 4* and *Bmp7*) showed no alterations in the expression pattern at E13.5 that could be linked with the syndactyly phenotype. *Nog* and *Bmp4* showed an overexpression in the distal fore limb (arrowhead). Bars: 200µm.

6.4.4 Duplication-Induced Alterations in Chromatin Configuration Define the Pathogenicity of Syndactyly

To investigate the effect of increased gene dosage in mice with a duplicated copy of *Ihh* on the limb phenotypes, Dup(syn)/+ mice were mated with mice heterozygous for the *Ihh* k.o. or the enhancer deletion Del(i2-9). Complete penetrance of cutaneous syndactyly and broadening of the terminal phalanges was observed in double heterozygous mice (Figure 27). Penetrance of polydactyly was incomplete (approximately 50%), as observed before in Dup(syn)/+ mice (not shown). This indicates, that the increased gene dosage was not sufficient to drive the syndactyly (and polydactyly) phenotype. Instead, the duplication-induced modification of the enhancer composition might drive the pathogenicity of this phenotype.

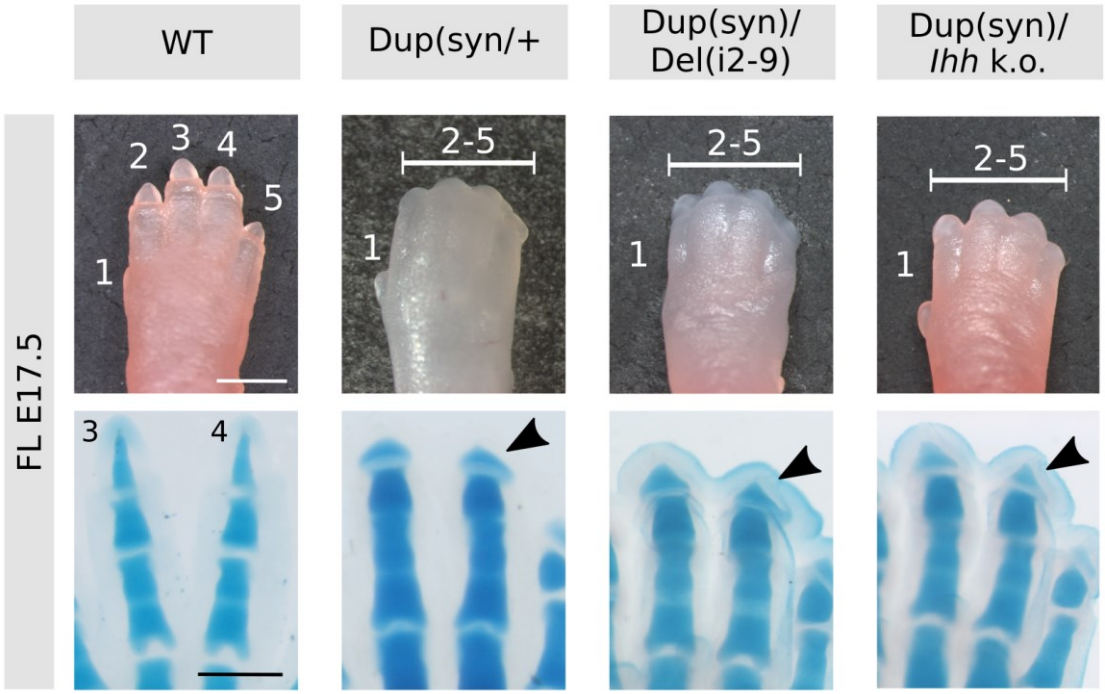

Figure 27. Phenotype analysis of double-heterozygous mutants reveals gene dosage independent syndactyly

All mutants show cutaneous syndactyly (digits 2-5) with full penetrance independent on *Ihh* copy number. Skeletal preparations of E17.5 forelimbs with the magnifications of the digits 3 and 4 show broadening of the terminal phalange (arrowheads) in all mutants, being consistent with the clinical phenotype. Bars: 1000µm limbs, 500µm skeletal staining.

To investigate the effects of the duplications on enhancer- promoter interactions, 4C-Seq was performed in E14.5 hands (digits 2-5). The *Ihh* promoter was used as viewpoint (Figure 28A). In Dup(int)/+ mice, subtraction analysis revealed an increase of enhancer-promoter contacts across the entire *Nhej1*-intron, suggesting that the *Ihh* promoter is able to contact all enhancers of the centromeric and the telomeric copy (i2-9' and i2-9). However, in Dup(syn)/+ embryos, increased interaction frequencies were only observed at the duplicated region (enhancers i1-i5) but not at the unaffected part of the *Nhej1*-intron (enhancers i6-9), (Figure 28B). Interestingly, this duplication generated a cluster of divergent CTCF binding sites at the boarder of the centromeric and the telomeric copy that might insulate chromatin interactions between the two copies. These observations suggest that the centromeric gene copy (*Ihh*') is not able to contact the enhancers of the telomeric cluster (i1-9) and builds an isolated interaction domain (*Ihh*' with i1-5'). However, the duplicated enhancer i5' might indeed possess the ability to interact with the telomeric gene copy (*Ihh*) due to its close proximity to the gene (2.5kb) and its unique position at the center of the divergently oriented CTCF pair (Figure 28C). This observation indicates that the observed misexpression of *Ihh* might be induced by specific enhancer-promoter interactions, rather than by increased gene or enhancer dosage.

In summary, these results highlight the divergence of the underlying pathomechanisms of phenotypes that are associated with CNVs of the non-coding genome at the *Ihh* locus. While some are purely dose dependent (craniosynostosis), others are induced by alterations of regulatory configurations that may result in tissue specific misexpression (syndactyly).

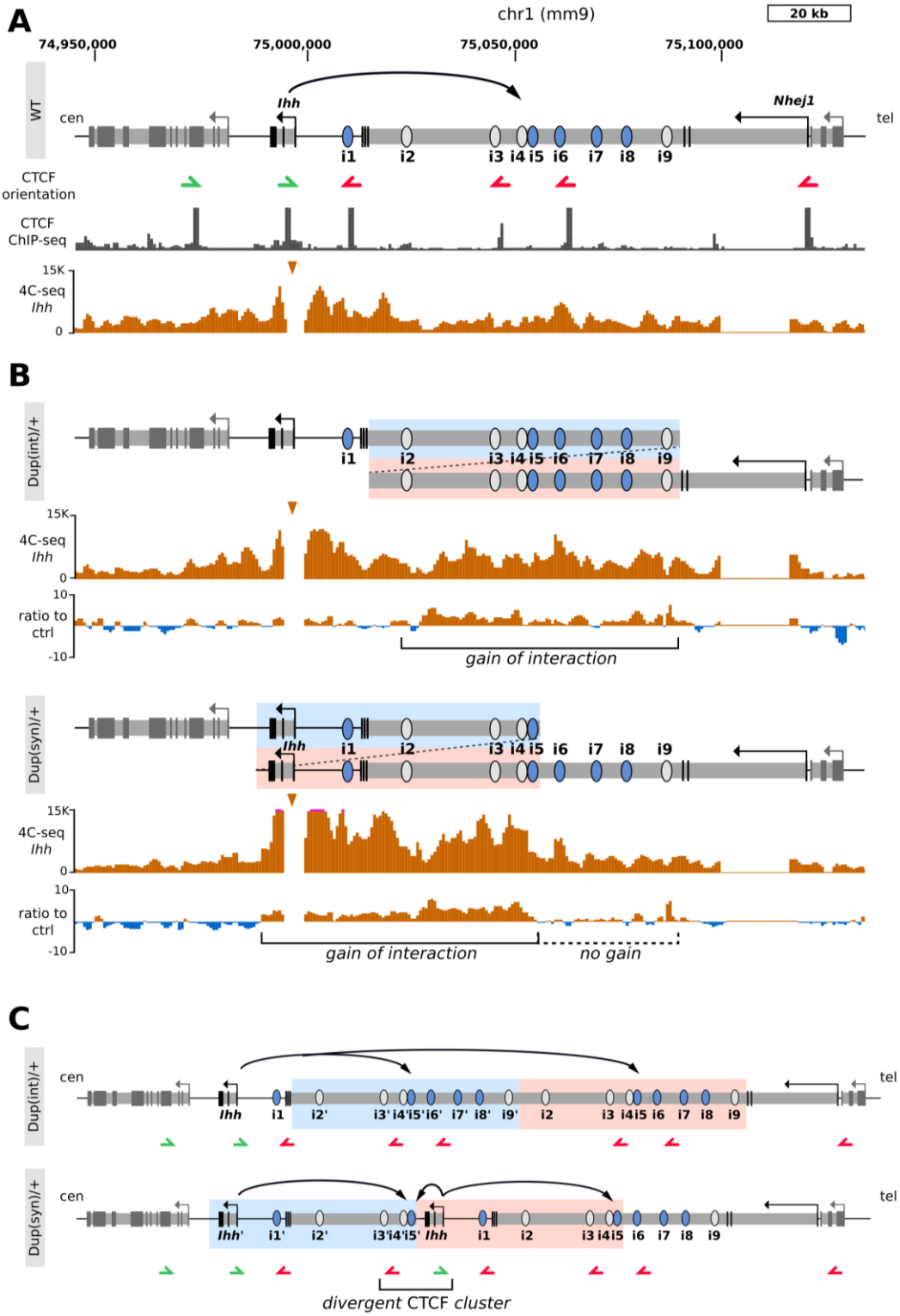

Figure 28. Reorganization of the *Ihh* enhancer cluster induces tissue-specific enhancer-promoter contacts

(A) Schematic overview of the *Ihh* locus. Introns and exons are indicated in grey and black, respectively (cen, centromeric; tel, telomeric). Black arrow indicates the interaction of the *Ihh* promoter with the main fingertip enhancer (i5) in wt configuration. CTCF Chip-Seq of E14.5 limbs shows the distribution of CTCF binding sites at the *Ihh*-TAD[104], arrow indicate CTCF motif orientation. 4C-Seq profile shows interaction frequencies of the *Ihh* promoter (arrowhead) in E14.5 wt hands. (B) Schematic of Dup(int) and Dup(syn) on top, duplicated regions are

highlighted (blue and red boxes). Below, 4C-Seq of E14.5 hands and ratio to control (wt) show that the gain of interaction (brackets) is limited to the duplicated regions, respectively. In Dup(syn)/+ embryos, no gain of interaction was detected for the enhancers i6-9. (C) Linear model of genomic interactions, duplicated regions are highlighted (blue and red boxes). In Dup(int), the *Ihh* promoter interacts with both copies of the main fingertip enhancer (i5' and i5). In Dup(syn), *Ihh'* is restricted to interact only with the centromeric i5' due to the newly formed divergent CTCF cluster. The telomeric copy of *Ihh*, however, is able to interact with the main fingertip enhancer i5 (long arrow) and with its centromeric copy i5' due to the genomic proximity of *Ihh* and i5' (short arrow).

7 Discussion

Genome-wide studies conducted by the ENCODE consortium and others revealed that a large fraction of the genome comprises non-coding elements, referred to as enhancers, that induce gene expression in many species[104]. Developmental genes are often regulated by multiple enhancers that arrange in clusters in the genomic vicinity of their target genes and share redundant characteristics in TF binding affinities and reporter assays[34,35]. However, the extent of redundancy among individual enhancers and how the activity of these elements is orchestrated to ensure complexity and precision of gene expression during developmental processes is highly discussed. Recent studies conducted in *Drosophila* and mammals suggest that redundancy among enhancers might be crucial to ensure spatio-temporal precision and robustness to gene expression[36,39,41,240].

Our current knowledge of how enhancers function during developmental processes is mainly based on targeted deletion studies. Deletion studies are a common tool to assess the *in vivo* functionality of individual enhancers and enhancer clusters, but fail to explain the underlying pathomechanisms of naturally occurring CNVs. In human, CNVs (deletions, duplications and insertions) of the non-coding genome have been frequently linked to disease. CNVs might induce alterations in enhancer dosage in either direction, by decreasing or increasing enhancer number, as well as alter the composition of an enhancer cluster through systemic rearrangements. In an attempt to answer the question of how redundancy and dosage respond to systematic variations in enhancer number and cluster composition, the regulation of *Ihh* during mouse skeletal development was investigated in an exemplary manner.

7.1 Identification of *Ihh* Exclusive Enhancers

In skeletal development, IHH plays a central role in regulating the process of chondrocyte differentiation[241]. In human, duplications upstream of *IHH* have been associated with highly localized phenotypes, i.e. craniosynostosis, syndactyly and polydactyly[196,197]. These phenotypes could not be linked to previously reported functions of IHH, derived from annotated human mutations and k.o. studies in mice that affect endochondral ossification, ranging from mild - in which only a single phalange is affected (Brachydactyly)[120] - to strong deficiencies of the skeletal structures with postnatal lethality[79]. Interestingly, only one of the four disease-associated duplications contains the *IHH* gene itself. However, all four duplications comprise more than 50kb of non-coding sequence upstream of *IHH*, covering the third intron of the neighboring *NHEJ1* gene. Klopocki *et al* identified a common overlapping region of 6.2kb that showed reporter activity resembling *Ihh*-specific expression patterns in the condensations of long bones, the fingertips and the growth plates at E13.5, E15.5 and E17.5[196]. However, the variability of the four duplications suggests two scenarios. First, the duplications might expand the regulatory landscape of *IHH* and could affect the regulation of neighboring

genes. Second, several more enhancers might be located upstream and downstream of the common overlapping region and might contribute to *IHH* expression as well.

In consideration that the organization of the genome in regulatory domains (TADs) is highly conserved between human and mice[55], the genomic position and context of *Ihh* was analyzed using 3C-based methods to investigate if the duplications affect the regulation of *Ihh* neighboring genes. The application of publicly available Hi-C data of mouse ESC[55] revealed that *Ihh* is localized in a TAD that comprises more than 40 genes. Recent studies have shown that structural rearrangements (intra- and inter-TAD) can affect TAD-organization and as a consequence disrupt well-established or create novel enhancer-promoter interactions[60,118]. Thus, it might be possible that these duplications, although they do not span TAD boundaries, interfere with the regulation of *Ihh* neighboring genes. To assess if the four reported duplications induce structural rearrangements that exceed the regulatory landscape of *Ihh*, 4C-Seq was performed in E14.5 mouse developing forelimbs using the *Ihh* promoter as the viewpoint. 4C-Seq revealed high contact frequencies of the *Ihh* promoter with the upstream *Nhej1*-intron, the genomic region to which the reported duplications have been mapped (Figure 29A). In order to capture the full regulatory potential of this region in its natural genomic context, an SB cassette containing a *LacZ* reporter gene[71] was introduced at the center of the *Nhej1*-intron. The LacZ assay revealed reporter activity in the digit condensations and fingertips at E14.5 as well as in the growth plates and skull sutures at E17.5 (Figure 15), which is consistent with the known expression pattern of *Ihh*[241]. Furthermore, no additional expression domains that could be linked to the interactions of this region with nearby genes were detected. In agreement with the domain-limited spread of *Ihh* promoter contact frequencies, this suggests that *Ihh* forms a regulatory sub-domain (*Ihh*-TAD) that comprises enhancers, which regulate the expression of *Ihh* exclusively. Additionally, the *Ihh*-TAD appears to be an isolated regulatory unit that functions independently of its genomic context. Since the orthologous duplications do not extend the *Ihh*-TAD, intra-TAD rearrangements and alterations in enhancer dosage might be disease causing.

The position of putative regulatory elements within the *Ihh*-TAD was determined by analyzing common enhancer features (enhancer-associated histone marks, similarities in Runx2 TF binding sites and sequence conservation) in combination with the generated 4C-Seq data. Thereby, nine regions (i1-9) with enhancer potential were identified. The *in vivo* activity of these elements was confirmed in enhancer-reporter assays[242] (Figure 16 and Supplementary Figure 32). Developmental genes are often regulated by a cluster of multiple enhancers that display overlapping expression patterns in enhancer-reporter assays, suggesting functional redundancy[34–36]. To assess the putative functional behavior of the enhancer cluster, the elements were scored for enhancer-activity in the previously identified tissues (digit condensations, fingertips, growth plates and skull sutures) at two developmental time points (E14.5 and E17.5). Based on this scoring system, three functional characteristics were observed for this enhancer cluster.

First, all of the tested regions showed reporter activity in at least one (and up to four) of the evaluated tissues, indicating that all elements possess *Ihh*-specific enhancer potential.

In consistency with the known regulatory function of IHH in chondrocyte differentiation[241], most elements showed reporter activity in the growth plates at E17.5. This suggestes that *Ihh* expression might be regulated by a cluster of at least nine enhancers.

Second, reporter activity was highly divergent among the individual enhancers in respect of the evaluated tissues and developmental time points and did not directly correlate with Runx2 binding and enhancer-associated histone marks. This was shown, as only five of nine elements (i1 and i5-8) activated reporter expression at E14.5 and E17.5, of which no more than two enhancers showed the same combination of activity. Another example of this discrepancy is the enhancer i2, which showed strong Chip-Seq signals for enhancer-associated histone marks, but proved to be a rather dispensable enhancer as it was active at E17.5 only. Thus, enhancer identification shall be conducted using a combination of enhancer-specific characteristics and need to be confirmed in enhancer-reporter assays that, in an optimal situation, are conducted at various time points as individual enhancers can get activated differently in different tissues and developmental time points.

Third, in each tissue reporter activity was driven by at least two enhancers, indicating for partial functional redundancy within the cluster, which might be crucial to ensure developmental robustness. Interestingly, in some tissues reporter expression was activated only by a subset of enhancers. Tissue-specificity was observed in fingertips in which only two enhancers (i5 and 7) induced reporter activity. These findings suggest that the enhancers in this cluster are not completely redundant and possess individual importance instead. This cluster appeared to act cooperatively and in a modular fashion by showing a variable degree of overlapping activity between tissues and developmental time points, which provides gene expression with developmental robustness and precision. This effect has previously been observed in patterning studies conducted in *Drosophila*, where enhancer-reporter assays revealed that different enhancers of a cluster contribute to different expression patterns in multiple gene loci, including snail, hunchback and knirps[41,243].

7.2 Generation of CNVs to Induce Alterations in Enhancer Dosage and Cluster Composition Using CRISVar

Variations of the human genome are tolerated in the normal population to a certain extent. Depending on the affected region and genomic context, copy number and structural variations can result in pathogenic alterations of gene expression, induced by deletions or duplications of *cis*-regulatory elements[111,112,114] or by reconstructing the regulatory organization of the genome[60,118]. Alterations that span neighboring regulatory units (TADs or sub-TADs) can disrupt well-established promoter-enhancer interactions or build novel regulatory contacts[60,118]. The four disease-associated duplications at the *Ihh* locus, however, do not extend the regulatory domain of *Ihh*, suggesting that variations in copy number and composition of the *Ihh* enhancer cluster might interfere with the natural dosage and expression domains of *Ihh* and as a consequence might induce pathogenic effects. To address the functionality of the *Ihh* enhancer cluster and the

effects of cluster intern CNVs, CRISVar[102] was applied to induce consecutive deletions and duplications in mouse ESCs. CRISVar is a derivative of the recently developed CRISPR/Cas9 technique and is frequently used to induce complex variations (deletions, duplications, inversion) of the genome by applying two sgRNAs simultaneously[102].

At the *Ihh* locus, genomic rearrangements were generated with high efficiency of deletions (8.3-44.1%) and tandem duplications (0.4-3.5%), being much higher than those obtained by recombinase- and nuclease-mediated targeting[92,244–246]. In agreement with previously published data[247,248], the efficiencies of deletions negatively correlated with the size of the target regions, being highest for the 13kb deletion Del(i4-6) and smallest for the 75kb deletion Del(i2-9), (Table 3). The duplication efficiency showed an opposing trend, reaching a maximum of 3.5% for the 75kb duplication Dup(int), an observation that has not been reported previously[102,248] and which might be simply dependent on the efficiency of the sgRNA. However, some false-negative clones may have been missed as NHEJ1-mediated repair can result in extensive rearrangements at the breakpoints and thus might interfere with the binding of genotyping primers. In general, tandem duplications were detected with a much lower frequency in comparison to deletions. Tandem duplications are induced by trans-allelic recombination between two homologous chromosomes, resulting in the loss of the target segment on one chromosome while the other one gains it. In contrast to *cis*-allelic recombination, this process requires the generation of four double strand breaks and further involves the stabilization and ligation of the correctly oriented target region. Therefore, tandem duplications appear less frequently [244,245].

Sanger sequencing revealed small sequence variations (insertions/deletions) of up to 100bp at the junctions of the deleted and duplicated alleles. This effect is referred to as 'scaring', which is typically induced by NHEJ-mediated repair[89] and might be problematic as it can induce frameshift mutations of coding regions. In this study CRISVar was exclusively applied to induce deletions and duplication of non-coding regulatory regions on which frameshift mutations show little functional effects[249]. To confirm the precision of the genomic modifications, the copy numbers of the target region and their 5'/3' flanking regions were determined. ESC clones with a clean genotype were selected for diploid aggregation to generate transgenic mouse lines for the assessment of functional effects of the generated modifications at multiple developmental time points. The generation of these CNVs substantially contributed to the development of the CRISVar protocol[102]. With the CRISVar protocol, transgenic mouse lines with structural and copy number variations are generated within ten weeks, a process that takes more than one year with conventional systems like TAMERE[71,95] which requires the engineering of three transgenic alleles. Thus, CRISVar represents a strong improvement in studying disease-associated structural variations and contributes substantially to gain a better understanding of human diseases. In this study, CRISVar was applied to induce consecutive CNVs (deletions and duplications) at the *Ihh* enhancer cluster in mouse ESCs.

7.3 Functional Characterization of the *Ihh* Enhancer Cluster

The previously performed enhancer-reporter assay revealed that the regulatory landscape of *Ihh* contains at least nine enhancers with regulatory potential, of which only one is located in close proximity of *Ihh* (i1) while most of them are located more distally, residing in the third intron of its neighboring gene *Nhej1*. Thus, the *in vivo* functionality of the enhancer cluster was determined by deleting the complete *Nhej1*-intron that contains eight of the nine identified enhancers, (i2-9). Embryos that were homozygous for this deletion [Del(i2-9)] were not viable and showed a severe skeletal phenotype at E17.5 being almost as severe as the previously reported *Ihh* k.o. phenotype[79] (Figure 18). In particular, Del(i2-9) embryos showed a strong shortening of the limbs, absence of cortical bone, fused joints of the autopods and reduced ossification of the skull. These abnormalities were the consequence of a drastic reduction of *Ihh* expression in E13.5 hands as well as in E17.5 skulls and growth plates by 98% to 99% (Figure 21). Therefore, the enhancer i1, which is located downstream of the *Nhej1*-intron, as well as other yet unidentified enhancers may as well contribute to *Ihh* expression, but to a rather negligible extent. This confirms that the *Nhej1*-intron contains most of the regulatory elements that are required for *Ihh* skeletal expression.

The tissue-based scoring that was applied in the previously performed enhancer-reporter assays revealed that *Ihh* enhancers share similar domains of activity, suggesting a partial functional redundancy. In addition, individual *Ihh* enhancers showed differential tissue-specific activities, implying a modular functionality of the enhancers within this cluster. Recent studies, mainly conducted in *Drosophila* have shown that partial redundancy among enhancers of developmental genes may be functionally relevant to provide gene expression with robustness and spatio-temporal precision[39,41,42]. This was demonstrated for several gene loci through enhancer deletion studies, which resulted in no observable or rather minor phenotypic effects. Thus, it was suggested that the loss of enhancer activity was balanced by the remaining redundant enhancers of a cluster[35,40,251]. Enhancer redundancy is also a characteristic feature of mammalian genomes. This was recently studied at the α-globin locus where individual enhancers of a cluster act independently and in an additive manner[240]. However, how complex patterns and precision of gene expression during development are achieved and why this involves elements with apparently redundant function remains elusive.

To assess the functional behavior of the *Ihh* enhancer cluster, consecutive deletions of three assigned cluster domains were generated: centromeric (i2-3), central (i4-6) and telomeric domain (i7-9). The centromeric domain contains the enhancer i2 and i3, which activated *LacZ* expression only slightly in one or two of the four evaluated tissues, suggesting that these enhancers might be purely redundant. The deletion of this domain [Del(i2-3)] showed no skeletal phenotype in E17.5 embryos nor changes of the *Ihh* expression pattern in E13.5 forelimbs. To assess, if these elements are truly dispensable for *Ihh* expression, the central and the telomeric domain of the enhancer cluster were deleted simultaneously [Del(i4-9)]. The deletion of i4-9 resulted in a drastic reduction of *Ihh* expression in E13.5 forelimbs and a strong skeletal phenotype with postnatal lethality (Figure 19). This confirmed that i1-3 are not able to induce a sufficient level of *Ihh*

expression during embryonic development, and that the most relevant *Ihh* enhancers are located in the central and telomeric domain of this cluster. Nevertheless, slight differences in *Ihh* expression levels as well as in the resulting phenotype between the two deletions Del(i2-9) and Del(i4-9) were observed (Figure 21). This suggests that i2 and i3 might contribute to chondrocyte maturation by fine-tuning/boosting *Ihh* expression at later developmental stages.

The central (i4-6) and the telomeric domain (i7-9) contain three enhancers, which collectively activate reporter expression in all evaluated tissue, respectively. Both, Del(i4-6) and Del(i7-9) mutants were viable and showed no obvious phenotypic effects, suggesting functional redundancy among these two domains. However, skeletal preparations and Micro-CT analysis revealed divergent phenotypic effects between homozygous mutants (Figure 20). In particular, Del(i4-6) embryos showed reduced ossification of the metacarpals and the skull as well as reduced size of the distal phalanges at E17.5. Furthermore, a significant reduction in length of the ulna (10%) and the nasal suture (15%) was detected in Del(i4-6) adult mice, but not in Del(i7-9). This phenotypic divergence can be explained by the decreased dosage of tissue-specific enhancers that resulted in reduced *Ihh* expression in all tested tissues, i.e. by 70% in Del(i4-6) and by 60-65% in Del(i7-9).

In both mutants, another phenotype was detected showing a reduced volume of the distal phalanges to a different extent, i.e. by 35% in Del(i4-6) and 20% in Del(i7-9) mice. This, however, does not correlate well with the similar expression levels that were observed among all three tissues along with the absence of other phenotypes in Del(i7-9) mice. As this phenotype is highly localized, it might be induced by the deletion of one of the two putative fingertip enhancers, that is i5 in Del(i4-6) or i7 in Del(i7-9). Noteworthy, quantifications of *Ihh* expression have been performed on tissues that contain different cell populations, thereby changes in expression that are induced in only a subset of cells might not have been detected, such as in this case for fingertip vs. handplate. Therefore, the expression pattern of *Ihh* was analyzed in E13.5 developing forelimbs by WISH. In Del(i4-6), the expression pattern of *Ihh* was equally reduced throughout the limb. Interestingly, in Del(i7-9) *Ihh* expression was only reduced in the proximal condensations, but remained nearly unaffected in the distal fingertips, an effect that has not been observed in any other deletion line (Figure 21). This suggests that i5, which remains active in Del(i7-9), may act as a major regulator of *Ihh* expression at the fingertips.

In summary, the deletion studies confirm the biological relevance of the *Ihh* enhancer cluster and show that this cluster is composed of enhancers that execute partially redundant functions. In consistency with the tissue-based scoring of the enhance-reporter assays, some enhancers have been proven to be less relevant (i1-3) to induce *Ihh* expression than others (i4-9). It appears that the most relevant *Ihh* enhancers are located in the central and telomeric domain of this cluster. Consecutive deletions of these two domains revealed that *Ihh* enhancers function in an additive fashion to maintain *Ihh* expression levels above a biological significant threshold. These levels appear to be less constrained in tissues in which *Ihh* expression is controlled by a small amount of

enhancers with tissue-specific activity. This was demonstrated by the deletion of each one of the two fingertip enhancers (i5 and i7), highlighting the spatio-temporal specificity of individual *Ihh* enhancers.

7.4 Effects of Increased Enhancer Dosage and Alterations in Cluster Composition

In human, duplications at the *IHH* locus have been shown to be associated with highly localized phenotypes, i.e. craniosynostosis, syndactyly and polydactyly[196,197]. The analysis of the *Ihh* regulatory landscape, using reporter assays and consecutive deletions, revealed that *Ihh* is regulated by a cluster of at least nine enhancers, whose copy number and composition is modified by these duplications. Interestingly, the four human duplications showed a small common overlap of 6.2kb at the center of the *Nhej1*-intron. This region contains two *Ihh* specific enhancers (i5 and i6) that have been identified in this study and proven to possess regulatory function. Since the duplications neither completely aligned elsewhere nor exceeded the regulatory landscape of *Ihh*, two scenarios shall be addressed to elucidate the underlying pathomechanisms of these human diseases. First, the duplication of the overlap region or a certain number of *Ihh* enhancers might be disease causing by simply increasing enhancer dosage, which results in *Ihh* overexpression. Second, it was shown that the duplications contain enhancers that are located up- and downstream of the common overlap region. This suggests that duplications of parts of this cluster might alter enhancer dosage as well as cluster composition, which in turn might affect the spatio-temporal precision and the level of *Ihh* expression, resulting in disease phenotypes.

To address these possibilities, three duplication mouse models were generated using CRISVar: Dup(int), Dup(csp) and Dup(syn). In Dup(int), the entire *Nhej1*-intron and with this the most relevant *Ihh* enhancers were duplicated to maximize the enhancer dosage. With Dup(csp) and Dup(syn) two of the human duplications were re-engineered to induce alterations in cluster composition. Dup(csp) comprised the enhancers i1-5, reconstructing the human duplication causing craniosynostosis Philadelphia type[196,197]. Dup(syn) contained the *Ihh* gene as well as its upstream regulatory region including the enhancers i1-5, equivalent to the human duplication causing syndactyly type 1[196]. With the generation of these duplications, all three phenotypes (craniosynostosis, syndactyly and polydactyly) were successfully reconstructed in mice.

In human, craniosynostis is characterized by the premature fusion of the cranial sutures during infant skull development. A common feature of craniosynostosis is the abnormal shaping of the head as a result of compensation for the spatial limitation forced by the expanding brain[201,202]. The skulls of the duplication lines appeared to be shaped normally. However, Micro-CT analysis of adult animals revealed the fusion of the metopic sutures (Figure 23). This was observed in all homozygous duplication lines but not in heterozygous animals, indicating a dosage effect. In contrast, syndactyly and polydactyly were recapitulated only in one duplication line: Dup(syn). Dup(syn) animals showed fully penetrant cutaneous syntactyly of the digits 2-5 in hands and of variable degree in feet,

accompanied by a 50% penetrance of preaxial polydactyly that was detected only in the feet (Figure 24). Skeletal staining of E17.5 hands revealed a broadening of the terminal phalanges (digits 2-5) but no fusion of bones and joints or any additional hand phenotype, thus fully recapitulating the human phenotype that was previously associated with this duplication[196]. Interestingly, these phenotypes were observed in both heterozygous and homozygous animals of this line to a comparable extent, but not in Dup(int) and Dup(csp) mice. This implies that an increased enhancer dosage might not be sufficient to induce these phenotypes.

To investigate the effect of the duplications on *Ihh* expression and if this can be linked to the observed phenotypes, *Ihh* expression was quantified in the previously defined tissues (Figure 22). In all duplication lines, an upregulation of *Ihh* expression was detected in the skull cap, i.e. 1.8-fold in Dup(csp) and Dup(syn) embryos, which showed a mild fusion of the metopic sutures and up to 5-fold in Dup(int) mutants, in which suture fusion was most pronounced. In consistency with opposing effects that were observed in the deletion lines with decreasing enhancer dosage, this suggests that the craniosynostosis phenotype is the result of local *Ihh* upregulation, induced by an increased dosage of skull-specific enhancers. However, this scenario did not explain the syndactyly and polydactyly phenotypes, as in Dup(syn)/+ hands *Ihh* expression was upregulated only mildly, but was excessively high in Dup(csp) mutants (increase of 1.4-fold and 4-fold, respectively). In both lines, WISH confirmed the upregulation of *Ihh* as well as of the hedgehog downstream targets *Ptc1* and *Gli1*. Interestingly, in the proximal digits expression signals were restricted to the condensations, but were broadened in the distal fingertips. This broadening was most prominent in Dup(syn)/+ forelimbs, in which the signals protruded into the distal interdigital space (Figure 25).

Digit separation is facilitated by the controlled induction of apoptosis in the interdigital mesenchyme, a process called interdigital cell death (ICD). In Dup(syn)/+ forelimbs, a local loss of apoptotic signal, which precisely overlaps with the expanded signaling of the IHH pathway factors, was observed in the distal interdigital mesenchyme (Figure 25). Thus, indicating that the ectopic expression of *Ihh* might interfere with ICD pathways. In ICD, apoptosis is induced by BMP2/4 and BMP7 that reduce the expression of the distal cell survival factor *Fgf8* and enhance the processing of the proximal apoptosis factor RA, respectively. The simultaneous downregulation of both BMP2 and BMP4 was shown to result in prolonged FGF8 signaling across the AER, which interferes with ICD and thus prevents digit separation[189,192,193]. WISH revealed no abnormalities in the expression of ICD-associated genes, with the exception of *BMP4*, suggesting that IHH signaling alone was sufficient to suppress apoptosis in the distal interdigital mesenchyme (Figure 26). During digit formation, BMP signaling controls the directed outgrowth at the phalanx-forming region by recruiting mesenchymal progenitor cells to the cartilage condensations[239]. In Dup(syn)/+ mutants, *Bmp4* expression was upregulated in the distal mesenchyme, most likely induced by the increased IHH signaling as *Ihh* was previously reported to positively regulate *Bmp* expression[253–255]. Since deficiencies in IHH and BMP signaling during digit development have been reported to be associated with Brachydactyly (the shortening of phalanges)[113,120,143,256,257], the increased reach of short-range IHH signaling might interfere with the signaling pathways at the phalanx-forming

region, and thus result in the broadening of the terminal phalanges in Dup(syn)/+ mutants.

Polydactyly is commonly caused by the loss of digit identity, induced by disruptions of the anterior-posterior SHH gradient[158,217]. Recently, a severe form of polydactyly (doublefoot mutant) has been described to be associated with perturbations of the long-range regulation of *Ihh* expression, resulting in a strong ectopic overexpression of *Ihh* and the subsequent disruption of the SHH gradient[258,259]. Interestingly, a mild increase of *Ihh* expression was observed in the distal zeugopod of Dup(syn)/+ hindlimbs (Figure 24). As IHH is a diffusible morphogen, it is possible that the increased expression of *Ihh* at this position might interfere with the anterior-posterior hedgehog gradient, thus inducing preaxial polydactyly. In the doublefoot mutant, a 600kb deletion starting 50kb upstream of *Ihh* was hypothesized to permit the activation of *Ihh* expression by a distant enhancer that might be located beyond the deleted region and normally would not interact with the *Ihh* promoter[254]. In consideration with this hypothesis, the ectopic expression of *Ihh* in Dup(syn)/+ mutants might result from specific enhancer-promoter interactions induced by alterations in cluster composition.

To investigate the effects of the duplications on enhancer-promoter interactions, 4C-Seq was performed in E14.5 hands (digits 2-5 only) of Dup(int)/+, Dup(syn)/+ and wt controls, with the *Ihh* promoter as the viewpoint. In Dup(int)/+ mutants, an increased interaction of the complete duplicated region (including the majority of enhancers: i2'-9') with the *Ihh* promoter was observed, suggesting that the *Ihh* promoter was able to contact the entire regulatory region of the mutated allele. In contrast, Dup(syn)/+ mutants showed increased contacts only for the duplicated region containing i1'-5', but not for the complete *Nhej1*-intron (Figure 28). This indicates that the genomic copy of *Ihh* (*Ihh*') created an isolated regulatory domain with these enhancers (*Ihh*' and i1'-5').

Recently, it was shown that gene expression can be influenced by the relative orientation of CTCF binding sites that are situated at enhancers and their cognate promoters. Guo *et al* showed that head-to-head oriented CTCF binding sites facilitate enhancer-promoter contact by looping, while tail-to-tail orientations suppress interactions between two neighboring domains[260]. At the duplication breakpoint of Dup(syn)/+ mutants, a divergently oriented pair of CTCF binding sites was formed, which might explain the separation of enhancer-promoter contacts of the duplicated from the native domain. However, the duplicated major fingertip enhancer i5' was found to be located in-between the divergently oriented CTCF binding sites, placing it in close proximity to the *Ihh* promoter of the native domain. This close proximity of i5' to the *Ihh* promoter might induce the formation of a unique short-range interaction (Figure 29B), which acts independent of the established long-range interactions of the two separated domains. The pathogenic potential of this regulatory configuration was highlighted by the generation of two double heterozygous mutants: Dup(syn)/Ihh k.o. and Dup(syn)/Del(i2-9). Both mutants contain one allele that facilitates *Ihh* expression. This allele is consistent with the genomic configuration of the duplication Dup(syn). *Ihh* expression from the second allele is inhibited by providing either a defective copy of *Ihh* [Dup(syn)/Ihh k.o.] or by inactivating its regulatory landscape [Dup(syn)/Del(i2-9)].

These mutants fully recapitulated both limb phenotypes (Figure 27), suggesting that the syndactyly and polydactyly resulted from the loss of precision in the spatio-temporal expression levels of *Ihh* during limb development rather than from gene dosage effects. Furthermore, this pathogenic misexpression was caused by the formation of a novel tissue-specific enhancer-promoter interaction, which was induced by the genomic rearrangements of the *Ihh* enhancer cluster.

In summary, the human phenotypes – craniosynostosis, syndactyly and polydactyly – were accurately recapitulated in mice. The investigation of the duplication lines at various developmental stages revealed that alterations of enhancer number result in the disturbance of quantity and precision of gene expression. Interestingly, the observed phenotypes did not generally correlate with increased enhancer dosage, revealing the complexity of the disease pathomechanisms. Moreover, the spatio-temporal precision of *Ihh* expression that is crucial for normal skeletal development was shown to be controlled by specific enhancer-promoter interactions. The disruption of these interactions, induced by alterations in cluster composition resulted in ectopic expression patterns, and consequently in highly localized phenotypes that are not normally linked to IHH function (syndactyly and polydactyly). This shows that alterations in enhancer dosage and cluster composition can induce site-specific rather than general effects and hence highlights the complexity and importance of enhancer clusters to confer spatio-temporal precision of gene expression.

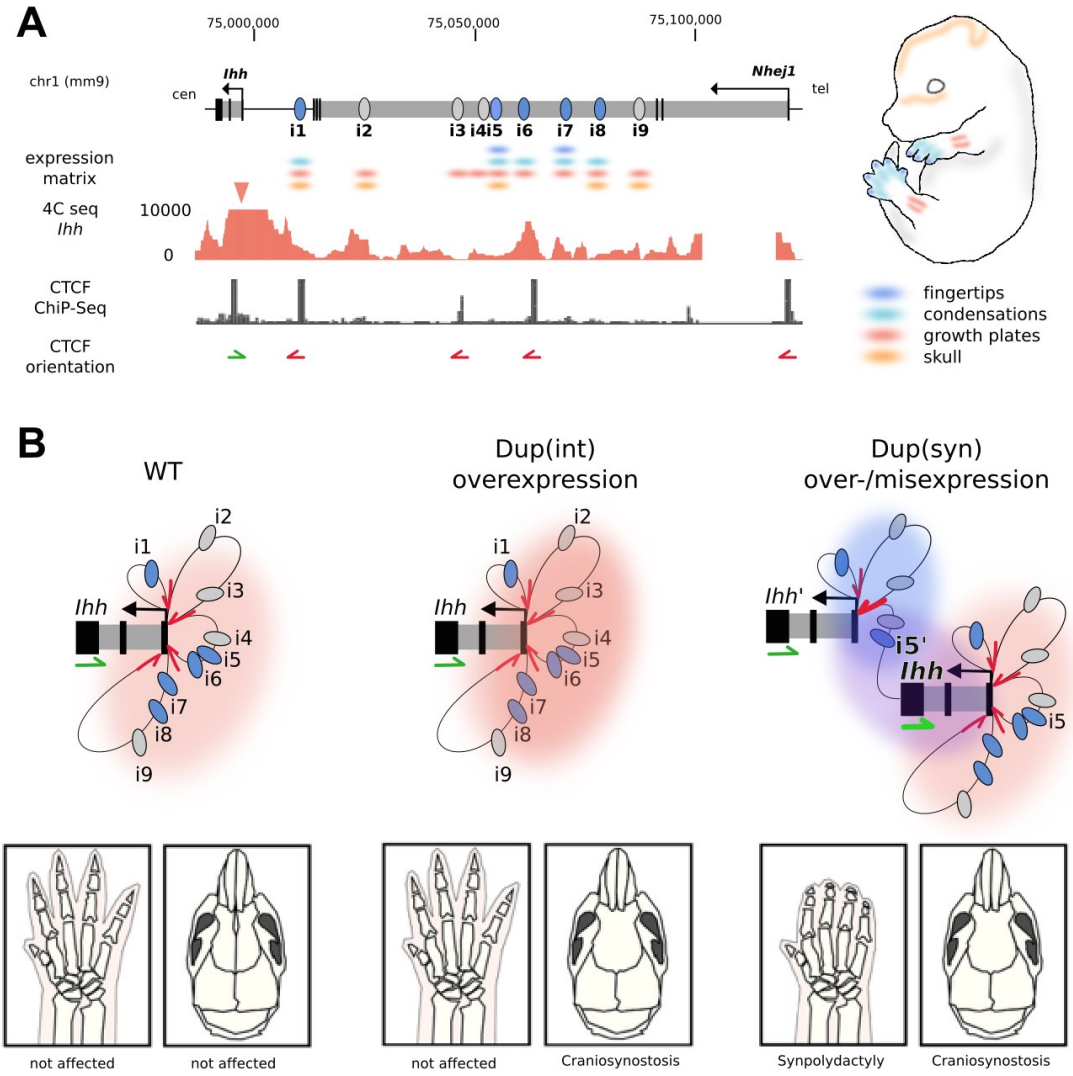

Figure 29. Functional model of the *Ihh* enhancer cluster and relation of cluster modifications with skeletal disease

(A) *Ihh* expression is controlled by a cluster of nine enhancers (i1-9) that are mainly located in the third intron of the upstream neighboring gene *Nhej1*. The most relevant enhancers are labeled in blue. Enhancer-reporter assays and tissue-based activity scoring suggest a modularity of the cluster elements, that function in both a partially redundant as well as additive manner, ensuring the spatio-temporal precision of *Ihh* expression. Color coding: dark blue – fingertips, light blue – digit condensations, red – growth plates, orange – skull sutures. The alignment of E14.5 4C-Seq and CTCF motif orientation[260] indicate enhancer-promoter contacts and suggest a model of chromatin configuration (B). (B) In wt configuration the *Ihh* promoter contacts all enhancers of the clusters (orange: *Ihh* and i1-9), resulting in normal development of skeletal features. In Dup(int) mutants, the *Ihh* promoter contacts all enhancers of the native domain (i1-9) as well as of the duplicated copy (dark orange: i2-9'), resulting in *Ihh* overexpression in all scored tissues and fusion of skull sutures, but no hand phenotype. In Dup(syn) mutants, only a part of the enhancer clusters is duplicated along with *Ihh* itself, forming a pair of divergently oriented CTCF binding sites at the duplication breakpoint (arrows are highlighted). The opposing orientation of this CTCF pair insulates interactions of the duplicated domain (blue: *Ihh*' and i1-5') with the native domain (orange: *Ihh* and i1-9). The positioning of the main fingertip enhancer (i5') beyond this boarder generates a unique enhancer-promoter configuration bridging the gap between the duplicated and the native domain (purple: i5' and *Ihh*). This unique interaction results in ectopic *Ihh* expression during limb development and in the highly localized phenotypes syndactyly and polydactyly.

8 References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
2. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7,** 29–59 (2006).
3. Butler, J. E. F. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16,** 2583–2592 (2002).
4. Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34,** 77–137 (2000).
5. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72,** 449–479 (2003).
6. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).
7. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol. CB* **20,** R754-763 (2010).
8. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12,** 1725–1735 (2003).
9. Ogbourne, S. & Antalis, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* **331 (Pt 1),** 1–14 (1998).
10. Privalsky, M. L. The role of corepressors in transcriptional regulation by nuclear hormone receptors. *Annu. Rev. Physiol.* **66,** 315–360 (2004).
11. Sertil, O., Kapoor, R., Cohen, B. D., Abramova, N. & Lowry, C. V. Synergistic repression of anaerobic genes by Mot3 and Rox1 in Saccharomyces cerevisiae. *Nucleic Acids Res.* **31,** 5831–5837 (2003).
12. Kulkarni, M. M. & Arnosti, D. N. cis-regulatory logic of short-range transcriptional repression in Drosophila melanogaster. *Mol. Cell. Biol.* **25,** 3411–3420 (2005).
13. Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6883–6888 (2002).
14. Capelson, M. & Corces, V. G. Boundary elements and nuclear organization. *Biol. Cell Auspices Eur. Cell Biol. Organ.* **96,** 617–629 (2004).
15. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405,** 482–485 (2000).
16. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98,** 387–396 (1999).
17. West, A. G. & Fraser, P. Remote control of gene transcription. *Hum. Mol. Genet.* **14 Spec No 1,** R101-111 (2005).
18. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461,** 186–192 (2009).
19. Weake, V. M. & Workman, J. L. Inducible gene expression: diverse regulatory mechanisms. *Nat. Rev. Genet.* **11,** 426–437 (2010).
20. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).
21. He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42,** 343–347 (2010).
22. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).
23. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41,** 941–945 (2009).
24. Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34,** 77–137 (2000).
25. Näär, A. M., Lemon, B. D. & Tjian, R. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* **70,** 475–501 (2001).

26. Vilar, J. M. G. & Saiz, L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.* **15,** 136–144 (2005).
27. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132,** 422–433 (2008).
28. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451,** 796–801 (2008).
29. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).
30. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20,** 578–588 (2010).
31. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23,** 297–305 (2006).
32. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11,** 394–403 (2004).
33. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.* **26,** 38–51 (2016).
34. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321,** 1314 (2008).
35. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr. Biol. CB* **20,** 1562–1567 (2010).
36. Perry, M. W., Bothma, J. P., Luu, R. D. & Levine, M. Precision of hunchback expression in the Drosophila embryo. *Curr. Biol. CB* **22,** 2247–2252 (2012).
37. Jiménez-Delgado, S., Pascual-Anaya, J. & Garcia-Fernàndez, J. Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. *Brief. Funct. Genomic. Proteomic.* **8,** 266–275 (2009).
38. Cande, J., Goltsev, Y. & Levine, M. S. Conservation of enhancer location in divergent insects. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14414–14419 (2009).
39. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol. CB* **26,** 38–51 (2016).
40. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466,** 490–493 (2010).
41. Dunipace, L., Ozdemir, A. & Stathopoulos, A. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Development* **138,** 4075–4084 (2011).
42. Dunipace, L., Saunders, A., Ashe, H. L. & Stathopoulos, A. Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev. Cell* **26,** 536–543 (2013).
43. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147,** 1132–1145 (2011).
44. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).
45. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. *Dev. Cell* **24,** 530–542 (2013).
46. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35,** 190–194 (2003).
47. Tark-Dame, M. *et al.* Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput. Biol.* **10,** e1003877 (2014).
48. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
49. van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. *Nat. Biotechnol.* **28,** 1089–1095 (2010).
50. de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods San Diego Calif* **58,** 189–191 (2012).

51. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485,** 376–380 (2012).
52. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48,** 471–484 (2012).
53. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
54. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–472 (2012).
55. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
56. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48,** 471–484 (2012).
57. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–472 (2012).
58. Rao, S. S. P. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680 (2014).
59. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **487,** 116–120 (2012).
60. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161,** 1012–1025 (2015).
61. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59–64 (2013).
62. Rao *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).
63. Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* **15,** R82 (2014).
64. Sandmann, T. *et al.* A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev.* **21,** 436–449 (2007).
65. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
66. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **advance online publication,** (2014).
67. Chiocchetti, A., Tolosano, E., Hirsch, E., Silengo, L. & Altruda, F. Green fluorescent protein as a reporter of gene expression in transgenic mice. *Biochim. Biophys. Acta* **1352,** 193–202 (1997).
68. Kothary, R. *et al.* Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Dev. Camb. Engl.* **105,** 707–714 (1989).
69. O'Kane, C. J. & Gehring, W. J. Detection in situ of genomic regulatory elements in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 9123–9127 (1987).
70. Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244,** 463–465 (1989).
71. Ruf, S. *et al.* Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43,** 379–386 (2011).
72. Visel, A. *et al.* Functional autonomy of distant-acting human enhancers. *Genomics* **93,** 509–513 (2009).
73. Schwarzer, W. & Spitz, F. The architecture of gene expression: integrating dispersed cis-regulatory modules into coherent regulatory domains. *Curr. Opin. Genet. Dev.* **27,** 74–82 (2014).
74. Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444,** 499–502 (2006).
75. Rossant, J., Zirngibl, R., Cado, D., Shago, M. & Giguère, V. Expression of a retinoic acid response element-hsplacZ transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes Dev.* **5,** 1333–1344 (1991).

76. Doyle, H. J., Kraut, R. & Levine, M. Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. *Genes Dev.* **3,** 1518–1533 (1989).
77. Goto, T., Macdonald, P. & Maniatis, T. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57,** 413–422 (1989).
78. Yang, X. W. & Gong, S. An overview on the generation of BAC transgenic mice for neuroscience research. *Curr. Protoc. Neurosci.* **Chapter 5,** Unit 5.20 (2005).
79. St-Jacques, B., Hammerschmidt, M. & McMahon, A. P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13,** 2072–2086 (1999).
80. Dassule, H. R., Lewis, P., Bei, M., Maas, R. & McMahon, A. P. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Dev. Camb. Engl.* **127,** 4775–4785 (2000).
81. Moon, A. M., Boulet, A. M. & Capecchi, M. R. Normal limb development in conditional mutants of Fgf4. *Dev. Camb. Engl.* **127,** 989–996 (2000).
82. Demetrius, L. Of mice and men. When it comes to studying ageing and the means to slow it down, mice are not just small humans. *EMBO Rep.* **6 Spec No,** S39-44 (2005).
83. Lewis, S. E. A consideration of the advantages and potential difficulties of the use of transgenic mice for the study of germinal mutations. *Mutat. Res.* **307,** 509–515 (1994).
84. Peng, Y. *et al.* Making designer mutants in model organisms. *Dev. Camb. Engl.* **141,** 4042–4054 (2014).
85. Orlando, S. J. *et al.* Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* **38,** e152–e152 (2010).
86. Kowalczykowski, S. C. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb. Perspect. Biol.* **7,** a016410 (2015).
87. Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y. & Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair* **5,** 1021–1029 (2006).
88. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu. Rev. Biochem.* **79,** 181–211 (2010).
89. Moore, J. K. & Haber, J. E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **16,** 2164–2173 (1996).
90. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161,** 1169–1175 (2002).
91. Sternberg, N. & Hamilton, D. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J. Mol. Biol.* **150,** 467–486 (1981).
92. Ramírez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378,** 720–724 (1995).
93. Smith, A. J. *et al.* A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat. Genet.* **9,** 376–385 (1995).
94. Van Deursen, J., Fornerod, M., Van Rees, B. & Grosveld, G. Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7376–7380 (1995).
95. Hérault, Y., Rassoulzadegan, M., Cuzin, F. & Duboule, D. Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat. Genet.* **20,** 381–384 (1998).
96. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
97. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
98. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
99. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497,** 254–257 (2013).
100. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

101. Inaki, K. & Liu, E. T. Structural mutations in cancer: mechanistic and functional insights. *Trends Genet. TIG* **28,** 550–559 (2012).
102. Kraft, K. *et al.* Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.01.016
103. Ecker, J. R. *et al.* Genomics: ENCODE explained. *Nature* **489,** 52–55 (2012).
104. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306,** 636–640 (2004).
105. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
106. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
107. Girirajan, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med.* **367,** 1321–1331 (2012).
108. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470,** 187–197 (2011).
109. Swaminathan, G. J. *et al.* DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum. Mol. Genet.* **21,** R37-44 (2012).
110. Volkmann, B. A. *et al.* Potential novel mechanism for Axenfeld-Rieger syndrome: deletion of a distant region containing regulatory elements of PITX2. *Invest. Ophthalmol. Vis. Sci.* **52,** 1450–1459 (2011).
111. D'haene, B. *et al.* Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening. *PLoS Genet.* **5,** e1000522 (2009).
112. Allou, L. *et al.* 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory elements. *Eur. J. Hum. Genet. EJHG* **20,** 1216–1223 (2012).
113. Dathe, K. *et al.* Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am. J. Hum. Genet.* **84,** 483–492 (2009).
114. Klopocki, E. *et al.* A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J. Med. Genet.* **45,** 370–375 (2008).
115. Sun, M. *et al.* Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J. Med. Genet.* **45,** 589–595 (2008).
116. Wieczorek, D. *et al.* A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum. Mutat.* **31,** 81–89 (2010).
117. Benko, S. *et al.* Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J. Med. Genet.* **48,** 825–830 (2011).
118. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538,** 265–269 (2016).
119. Kurth, I. *et al.* Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nat. Genet.* **41,** 862–863 (2009).
120. Gao, B. Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat. Genet.* **28,** 386–388 (2001).
121. Urban, M. & Krüger, S. Alice Vance ('Das Bärenweib'): a historical case of Nievergelt syndrome. *Am. J. Med. Genet.* **76,** 145–149 (1998).
122. Gilbert, S. F. Osteogenesis: The Development of Bones. (2000).
123. Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89,** 747–754 (1997).
124. Komori, T. *et al.* Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89,** 755–764 (1997).
125. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108,** 17–29 (2002).
126. Otto, F. *et al.* Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89,** 765–771 (1997).

127. Nagata, M. The primary site of the acrocephalic feature in Apert syndrome is a dwarf cranial base with accelerated chondrocytic differentiation due to aberrant activation of the FGFR2 signaling. *Bone* **48,** 847–856 (2011).
128. Opperman, L. A. Cranial sutures as intramembranous bone growth sites. *Dev. Dyn.* **219,** 472–485 (2000).
129. Wilkie, A. O. M. Craniosynostosis: Genes and Mechanisms. *Hum. Mol. Genet.* **6,** 1647–1656 (1997).
130. Akiyama, H., Chaboissier, M.-C., Martin, J. F., Schedl, A. & de Crombrugghe, B. 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 Dev.* **16,** 2813–2828 (2002).
131. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrugghe, B. Sox9 is required for cartilage formation. *Nat. Genet.* **22,** 85–89 (1999).
132. Wright, E. *et al.* The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nat. Genet.* **9,** 15–20 (1995).
133. Kozhemyakina, E. A pathway to bone: signaling molecules and transcription factors involved in chondrocyte development and maturation. *Development* **142,** 817–831 (2015).
134. Ruiz i Altaba, A. Gli and hedgehog in cancer: tumours, embryos and stem cells. *Nat. Rev. Cancer* **2,** 361–372 (2002).
135. Vortkamp, A. *et al.* Regulation of Rate of Cartilage Differentiation by Indian Hedgehog and PTH-Related Protein. *Science* **273,** 613–622 (1996).
136. Bitgood, M. J. & McMahon, A. P. HedgehogandBmpGenes Are Coexpressed at Many Diverse Sites of Cell–Cell Interaction in the Mouse Embryo. *Dev. Biol.* **172,** 126–138 (1995).
137. Lee, K. Expression of parathyroid hormone-related peptide and its receptor messenger ribonucleic acids during fetal development of rats. *Endocrinology* **136,** 453–463 (1995).
138. Kindblom, J. M. Expression and localization of Indian hedgehog (Ihh) and parathyroid hormone related protein (PTHrP) in the human growth plate during pubertal development. *J. Endocrinol.* **174,** R1–R6 (2002).
139. Lai, L. P. & Mitchell, J. Indian hedgehog: Its roles and regulation in endochondral bone development. *J. Cell. Biochem.* **96,** 1163–1173 (2005).
140. St-Jacques, B. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13,** 2072–2086 (1999).
141. Minina, E., Kreschel, C., Naski, M. C., Ornitz, D. M. & Vortkamp, A. Interaction of FGF, Ihh/Pthlh, and BMP Signaling Integrates Chondrocyte Proliferation and Hypertrophic Differentiation. *Dev. Cell* **3,** 439–449 (2002).
142. Yoshida, C. A. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes Dev.* **18,** 952–963 (2004).
143. Gao, B. *et al.* A mutation in Ihh that causes digit abnormalities alters its signalling capacity and range. *Nature* **458,** 1196–1200 (2009).
144. McMahon, A. P. Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53,** 1–114 (2003).
145. Hayashi, K. & Ozawa, E. Myogenic cell migration from somites is induced by tissue contact with medial region of the presumptive limb mesoderm in chick embryos. *Development* **121,** 661–669 (1995).
146. Johnson, R. L. & Tabin, C. J. Molecular Models for Vertebrate Limb Development. *Cell* **90,** 979–990 (1997).
147. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25,** 1–47 (1969).
148. Bénazet, J.-D. & Zeller, R. Vertebrate Limb Development: Moving from Classical Morphogen Gradients to an Integrated 4-Dimensional Patterning System. *Cold Spring Harb. Perspect. Biol.* **1,** a001339 (2009).
149. Zeller, R. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat. Rev. Genet.* **10,** 845–858 (2009).
150. Bénazet, J.-D. & Zeller, R. Vertebrate Limb Development: Moving from Classical Morphogen Gradients to an Integrated 4-Dimensional Patterning System. *Cold Spring Harb. Perspect. Biol.* **1,** a001339 (2009).

151. Niswander, L. Pattern formation: old models out on a limb. *Nat. Rev. Genet.* **4,** 133–143 (2003).
152. Rodriguez-Esteban, C. Radical fringe positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. *Nature* **386,** 360–366 (1997).
153. Niswander, L. Pattern formation: old models out on a limb. *Nat. Rev. Genet.* **4,** 133–143 (2003).
154. Muragaki, Y., Mundlos, S., Upton, J. & Olsen, B. R. Altered Growth and Branching Patterns in Synpolydactyly Caused by Mutations in HOXD13. *Science* **272,** 548–551 (1996).
155. Summerbell, D. & Lewis, J. H. Time, place and positional value in the chick limb-bud. *Development* **33,** 621–643 (1975).
156. Gilbert, S. F. Formation of the Limb Bud. (2000).
157. Lewandoski, M. Fgf8 signalling from the AER is essential for normal limb development. *Nat. Genet.* **26,** 460–463 (2000).
158. Riddle, R. D. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75,** 1401–1416 (1993).
159. Rallis, C. Tbx3 can alter limb position along the rostrocaudal axis of the developing embryo. *Development* **132,** 1961–1970 (2005).
160. Heikinheimo, M., Lawshé, A., Shackleford, G. M., Wilson, D. B. & MacArthur, C. A. Fgf-8 expression in the post-gastrulation mouse suggests roles in the development of the face, limbs and central nervous system. *Mech. Dev.* **48,** 129–138 (1994).
161. Crossley, P. H. & Martin, G. R. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121,** 439–451 (1995).
162. Charité, J. Ectopic expression of Hoxb-8 causes duplication of the ZPA in the forelimb and homeotic transformation of axial structures. *Cell* **78,** 589–601 (1994).
163. Welscher, P. te, Fernandez-Teran, M., Ros, M. A. & Zeller, R. Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes Dev.* **16,** 421–426 (2002).
164. Riddle, R. D. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75,** 1401–1416 (1993).
165. Ahn, S. & Joyner, A. L. Dynamic Changes in the Response of Cells to Positive Hedgehog Signaling during Mouse Limb Patterning. *Cell* **118,** 505–516 (2004).
166. Harfe, B. D. Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities. *Cell* **118,** 517–528 (2004).
167. Olsen, B. R. Bone Development. *Annu. Rev. Cell Dev. Biol.* **16,** 191–220 (2000).
168. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383,** 407–413 (1996).
169. Riddle, R. D. Induction of the LIM homeobox gene Lmx1 by WNT6a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83,** 631–640 (1995).
170. Parr, B. A., Shea, M. J., Vassileva, G. & McMahon, A. P. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119,** 247–261 (1993).
171. Niswander, L. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371,** 609–612 (1994).
172. Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79,** 993–1003 (1994).
173. Nissim, S. Regulation of Gremlin expression in the posterior limb bud. *Dev. Biol.* **299,** 12–21 (2006).
174. Bénazet, J.-D. A Self-Regulatory System of Interlinked Signaling Feedback Loops Controls Mouse Limb Patterning. *Science* **323,** 1050–1053 (2009).
175. Verheyden, J. M. & Sun, X. An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth. *Nature* **454,** 638–641 (2008).
176. Scherz, P. J. The limb bud Shh-Fgf feedback loop is terminated by expansion of former ZPA cells. *Science* **305,** 396–399 (2004).

177. Clarke, P. G. Developmental cell death: morphological diversity and multiple mechanisms. *Anat. Embryol. (Berl.)* **181,** 195–213 (1990).
178. Fernández-Terán, M. a. Birth and death of cells in limb development: A mapping study. *Dev. Dyn.* **235,** 2521–2537 (2006).
179. Chen, Y. & Zhao, X. Shaping limbs by apoptosis. *J. Exp. Zool.* **282,** 691–702 (1998).
180. Hernández-Martínez, R. & Covarrubias, L. Interdigital cell death function and regulation: New insights on an old programmed cell death model. *Dev. Growth Differ.* **53,** 245–258 (2011).
181. Kochhar, D. M. Evidence that retinoic acid-induced apoptosis in the mouse limb bud core mesenchymal cells is gene-mediated. *Prog. Clin. Biol. Res.* **383B,** 815–825 (1993).
182. Alles, A. J. & Sulik, K. K. Retinoic-acid-induced limb-reduction defects: perturbation of zones of programmed cell death as a pathogenetic mechanism. *Teratology* **40,** 163–171 (1989).
183. Niederreither, K. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat. Genet.* **21,** 444–448 (1999).
184. Salas-Vidal, E. Differential tissue growth and patterns of cell death in mouse limb autopod morphogenesis. *Dev. Dyn.* **220,** 295–306 (2001).
185. Abu-Abed, S. The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15,** 226–240 (2001).
186. Diez del Corral, R. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40,** 65–79 (2003).
187. Mercader, N. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Dev. Camb. Engl.* **127,** 3961–3970 (2000).
188. Hernández-Martínez, R. Progressive interdigital cell death: regulation by the antagonistic interaction between fibroblast growth factor 8 and retinoic acid. *Development* **136,** 3669–3678 (2009).
189. Guha, U. In Vivo Evidence That BMP Signaling Is Necessary for Apoptosis in the Mouse Limb. *Dev. Biol.* **249,** 108–120 (2002).
190. Lallemand, Y. Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb development. *Dev. Camb. Engl.* **132,** 3003–3014 (2005).
191. Wang, C.-K. L. Function of BMPs in the apical ectoderm of the developing mouse limb. *Dev. Biol.* **269,** 109–122 (2004).
192. Maatouk, D. M. In the limb AER Bmp2 and Bmp4 are required for dorsal–ventral patterning and interdigital cell death but not limb outgrowth. *Dev. Biol.* **327,** 516–523 (2009).
193. Pajni-Underwood, S., Wilson, C. P., Elder, C., Mishina, Y. & Lewandoski, M. BMP signals control limb bud interdigital programmed cell death by regulating FGF signaling. *Development* **134,** 2359–2368 (2007).
194. Zou, H. & Niswander, L. Requirement for BMP signaling in interdigital apoptosis and scale formation. *Science* **272,** 738–741 (1996).
195. Yokouchi, Y. BMP-2/-4 mediate programmed cell death in chicken limb buds. *Dev. Camb. Engl.* **122,** 3725–3734 (1996).
196. Klopocki, E. *et al.* Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* **88,** 70–75 (2011).
197. Barroso, E. *et al.* Identification of the fourth duplication of upstream IHH regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am. J. Med. Genet. A.* **167A,** 902–906 (2015).
198. Bosse, K. Localization of a Gene for Syndactyly Type 1 to Chromosome 2q34-q36. *Am. J. Hum. Genet.* **67,** 492–497 (2000).
199. Cohen, M. M. Craniosynostosis and syndromes with craniosynostosis: incidence, genetics, penetrance, variability, and new syndrome updating. *Birth Defects Orig. Artic. Ser.* **15,** 13–63 (1979).
200. Lenton, K. A., Nacamuli, R. P., Wan, D. C., Helms, J. A. & Longaker, M. T. in (ed. Biology, B.-C. T. in D.) **66,** 287–328 (Academic Press, 2005).

201. Kimonis, V., Gold, J.-A., Hoffman, T. L., Panchal, J. & Boyadjiev, S. A. Genetics of craniosynostosis. *Semin. Pediatr. Neurol.* **14,** 150–161 (2007).
202. Slater, B. J. *et al.* Cranial Sutures: A Brief Review: *Plast. Reconstr. Surg.* **121,** 170e–178e (2008).
203. Robin, N. H. Craniosynostosis, Philadelphia type: A new autosomal dominant syndrome with sagittal craniosynostosis and syndactyly of the fingers and toes. *Am. J. Med. Genet.* **62,** 184–191 (1996).
204. Panchal, J. & Uttchin, V. Management of craniosynostosis. *Plast. Reconstr. Surg.* **111,** 2032–2048; quiz 2049 (2003).
205. Gault, D. T., Renier, D., Marchac, D. & Jones, B. M. Intracranial pressure and intracranial volume in children with craniosynostosis. *Plast. Reconstr. Surg.* **90,** 377–381 (1992).
206. Bonaventure, J. & El Ghouzzi, V. Molecular and cellular bases of syndromic craniosynostoses. *Expert Rev. Mol. Med.* **5,** 1–17 (2003).
207. Cunningham, M., Seto, M., Ratisoontorn, C., Heike, C. & Hing, A. Syndromic craniosynostosis: from history to hydrogen bonds. *Orthod. Craniofac. Res.* **10,** 67–81 (2007).
208. Lajeunie, E., Le Merrer, M., Bonaïti-Pellie, C., Marchac, D. & Renier, D. Genetic study of nonsyndromic coronal craniosynostosis. *Am. J. Med. Genet.* **55,** 500–504 (1995).
209. Morriss-Kay, G. M. & Wilkie, A. O. Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. *J. Anat.* **207,** 637–653 (2005).
210. Wang, Y. *et al.* Abnormalities in cartilage and bone development in the Apert syndrome FGFR2+/S252W mouse. *Development* **132,** 3537–3548 (2005).
211. Amano, K. *et al.* Msx2 Stimulates Chondrocyte Maturation by Controlling Ihh Expression. *J. Biol. Chem.* **283,** 29513–29521 (2008).
212. Jabs, E. W. *et al.* A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell* **75,** 443–450 (1993).
213. Eggenschwiler, J. T., Bulgakov, O. V., Qin, J., Li, T. & Anderson, K. V. Mouse Rab23 regulates Hedgehog signaling from Smoothened to Gli proteins. *Dev. Biol.* **290,** 1–12 (2006).
214. McDonald-McGinn, D. M. *et al.* Metopic craniosynostosis due to mutations in GLI3: A novel association. *Am. J. Med. Genet. A.* **152A,** 1654–1660 (2010).
215. Jenkins, D. *et al.* RAB23 Mutations in Carpenter Syndrome Imply an Unexpected Role for Hedgehog Signaling in Cranial-Suture Development and Obesity. *Am. J. Hum. Genet.* **80,** 1162–1170 (2007).
216. Kornak, U. & Mundlos, S. Genetic Disorders of the Skeleton: A Developmental Approach. *Am. J. Hum. Genet.* **73,** 447–474 (2003).
217. Anderson, E. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet.* **28,** 364–373 (2012).
218. Malik, S. Syndactyly: phenotypes, genetics and current classification. *Eur. J. Hum. Genet.* **20,** 817–824 (2012).
219. Jordan, D., Hindocha, S., Dhital, M., Saleh, M. & Khan, W. The Epidemiology, Genetics and Future Management of Syndactyly. *Open Orthop. J.* **6,** 14–27 (2012).
220. Sanz-Ezquerro, J. J. & Tickle, C. Autoregulation of Shh expression and Shh induction of cell death suggest a mechanism for modulating polarising activity during chick limb development. *Development* **127,** 4811–4823 (2000).
221. Klopocki, E. *et al.* Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* **88,** 70–75 (2011).
222. Barroso, E. *et al.* Identification of the fourth duplication of upstream IHH regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am. J. Med. Genet. A.* **167,** 902–906 (2015).
223. Duprez, D. Sonic Hedgehog induces proliferation of committed skeletal muscle cells in the chick limb. *Development* **125,** 495–505 (1998).
224. Lohan, S. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin. Genet.* **86,** 318–325 (2014).
225. Biesecker, L. G. Polydactyly: How many disorders and how many genes? 2010 update. *Dev. Dyn.* **240,** 931–942 (2011).

226. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
227. Mouse Enhancer Screen Handbook. Available at: https://enhancer.lbl.gov/aboutproject_n.html#candidate_identity. (Accessed: 8th March 2017)
228. Artus, J. & Hadjantonakis, A.-K. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol. Biol. Clifton NJ* **693,** 37–56 (2011).
229. van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513,** 89–112 (2012).
230. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9,** 969–72 (2012).
231. Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods San Diego Calif* **58,** 221–30 (2012).
232. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25,** 1754–60 (2009).
233. Barroso, E. Identification of the fourth duplication of upstream IHH regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am. J. Med. Genet. A.* **167,** 902–906 (2015).
234. Klopocki, E. *et al.* Copy-Number Variations Involving the IHH Locus Are Associated with Syndactyly and Craniosynostosis. *Am. J. Hum. Genet.* **88,** 70–75 (2011).
235. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).
236. Meyer, M. B., Benkusky, N. A. & Pike, J. W. The RUNX2 cistrome in osteoblasts: characterization, down-regulation following differentiation, and relationship to gene expression. *J. Biol. Chem.* **289,** 16016–16031 (2014).
237. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
238. Klopocki, E. & Mundlos, S. Copy-number variations, noncoding sequences, and human phenotypes. *Annu. Rev. Genomics Hum. Genet.* **12,** 53–72 (2011).
239. Robert, B. Bone morphogenetic protein signaling in limb outgrowth and patterning. *Dev. Growth Differ.* **49,** 455–468 (2007).
240. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48,** 895–903 (2016).
241. Vortkamp, A. *et al.* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273,** 613–622 (1996).
242. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35,** D88-92 (2007).
243. Perry, M. W., Boettiger, A. N. & Levine, M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13570–13575 (2011).
244. Wu, S., Ying, G., Wu, Q. & Capecchi, M. R. Toward simpler and faster genome-wide mutagenesis in mice. *Nat. Genet.* **39,** 922–930 (2007).
245. Lee, H. J., Kweon, J., Kim, E., Kim, S. & Kim, J.-S. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* **22,** 539–548 (2012).
246. Gupta, A. *et al.* Targeted chromosomal deletions and inversions in zebrafish. *Genome Res.* **23,** 1008–1017 (2013).
247. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **292,** 2556 (2017).
248. Li, J. *et al.* Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell Biol.* mjv016 (2015). doi:10.1093/jmcb/mjv016

249. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342,** 253–257 (2013).
250. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
251. El-Sherif, E. & Levine, M. Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Curr. Biol. CB* **26,** 1164–1169 (2016).
252. Stricker, S. & Mundlos, S. Mechanisms of digit formation: Human malformation syndromes tell the story. *Dev. Dyn.* **240,** 990–1004 (2011).
253. Minina, E. *et al.* BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Dev. Camb. Engl.* **128,** 4523–4534 (2001).
254. Babbs, C., Furniss, D., Morriss-Kay, G. M. & Wilkie, A. O. M. Polydactyly in the mouse mutant Doublefoot involves altered Gli3 processing and is caused by a large deletion in cis to Indian hedgehog. *Mech. Dev.* **125,** 517–526 (2008).
255. Witte, F. Receptor tyrosine kinase-like orphan receptor 2 (ROR2) and Indian hedgehog regulate digit outgrowth mediated by the phalanx-forming region. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14211–14216 (2010).
256. Baur, S. T., Mai, J. J. & Dymecki, S. M. Combinatorial signaling through BMP receptor IB and GDF5: shaping of the distal mouse limb and the genetics of distal limb diversity. *Dev. Camb. Engl.* **127,** 605–619 (2000).
257. Yi, S. E., Daluiski, A., Pederson, R., Rosen, V. & Lyons, K. M. The type I BMP receptor BMPRIB is required for chondrogenesis in the mouse limb. *Dev. Camb. Engl.* **127,** 621–630 (2000).
258. Hayes, C., Brown, J. M., Lyon, M. F. & Morriss-Kay, G. M. Sonic hedgehog is not required for polarising activity in the Doublefoot mutant mouse limb bud. *Dev. Camb. Engl.* **125,** 351–357 (1998).
259. Yang, Y., Guillot, P., Boyd, Y., Lyon, M. F. & McMahon, A. P. Evidence that preaxial polydactyly in the Doublefoot mutant is due to ectopic Indian Hedgehog signaling. *Dev. Camb. Engl.* **125,** 3123–3132 (1998).
260. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162,** 900–910 (2015).

9 Supplementary Figures

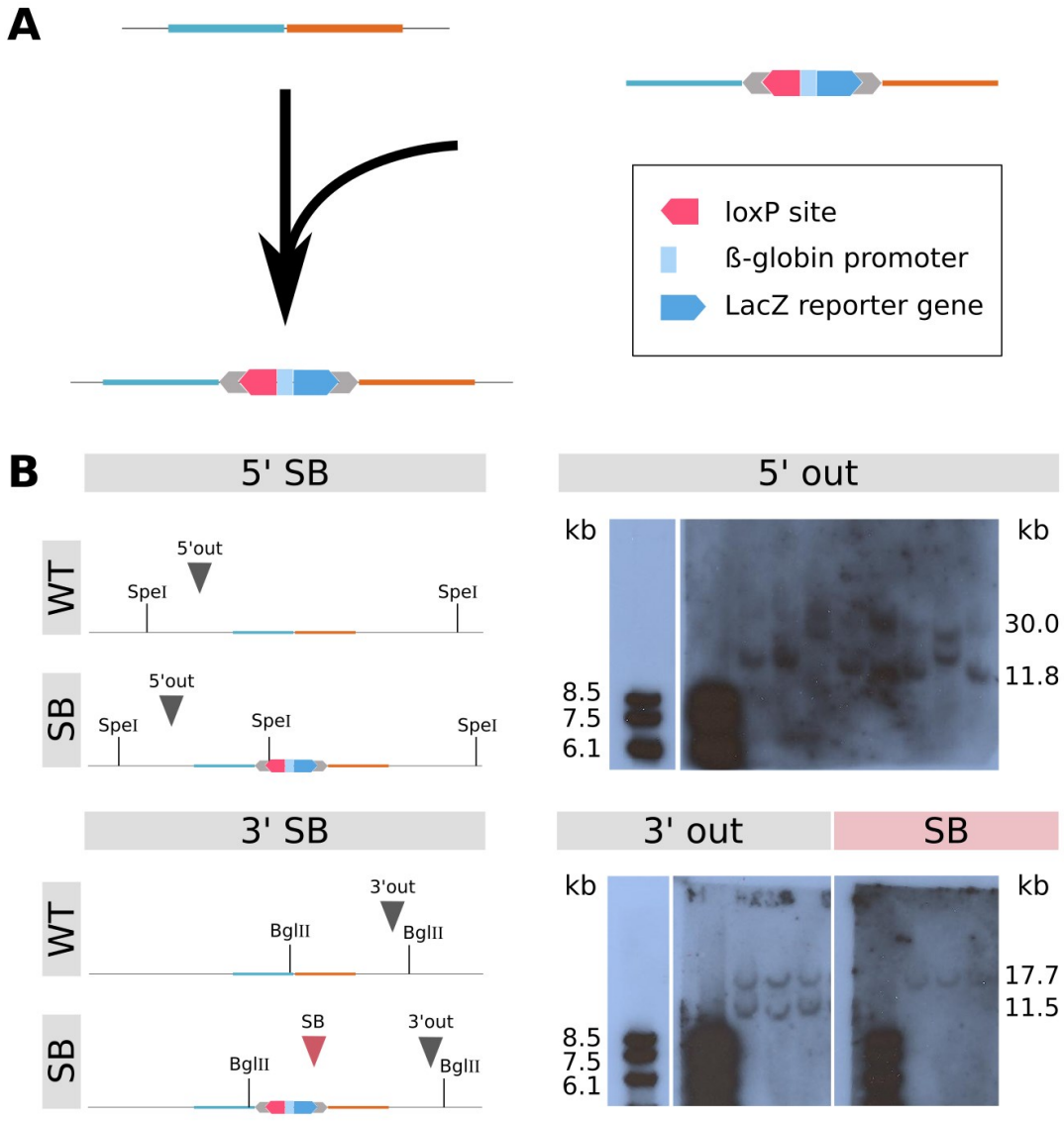

Figure 30. SB reporter insertion strategy and detection with Southern Blot

(A) The SB reporter cassette (red/blue) contains: a loxP site (red), a minimal ß-globin promoter (light blue) and a LacZ reporter (blue). Homologous regions that were used for SB insertion are indicated (blue and orange). (B) Detection of SB insertion via Southern Blot: the left panel shows the digestion strategies (5' and 3' of SB) and the position of the probes (5'out, 3'out and SB, arrowhead) that were used to detect the insertion. The right panel shows the digested fragments detected with Southern Blot (5'out, 3'out and SB). Clones with SB insertion: 30.0kb, 17.7kb and 17.7kb, wt clones: 11.8kb, 11.5kb and no signal for the probes 5'out, 3'out and SB, respectively.

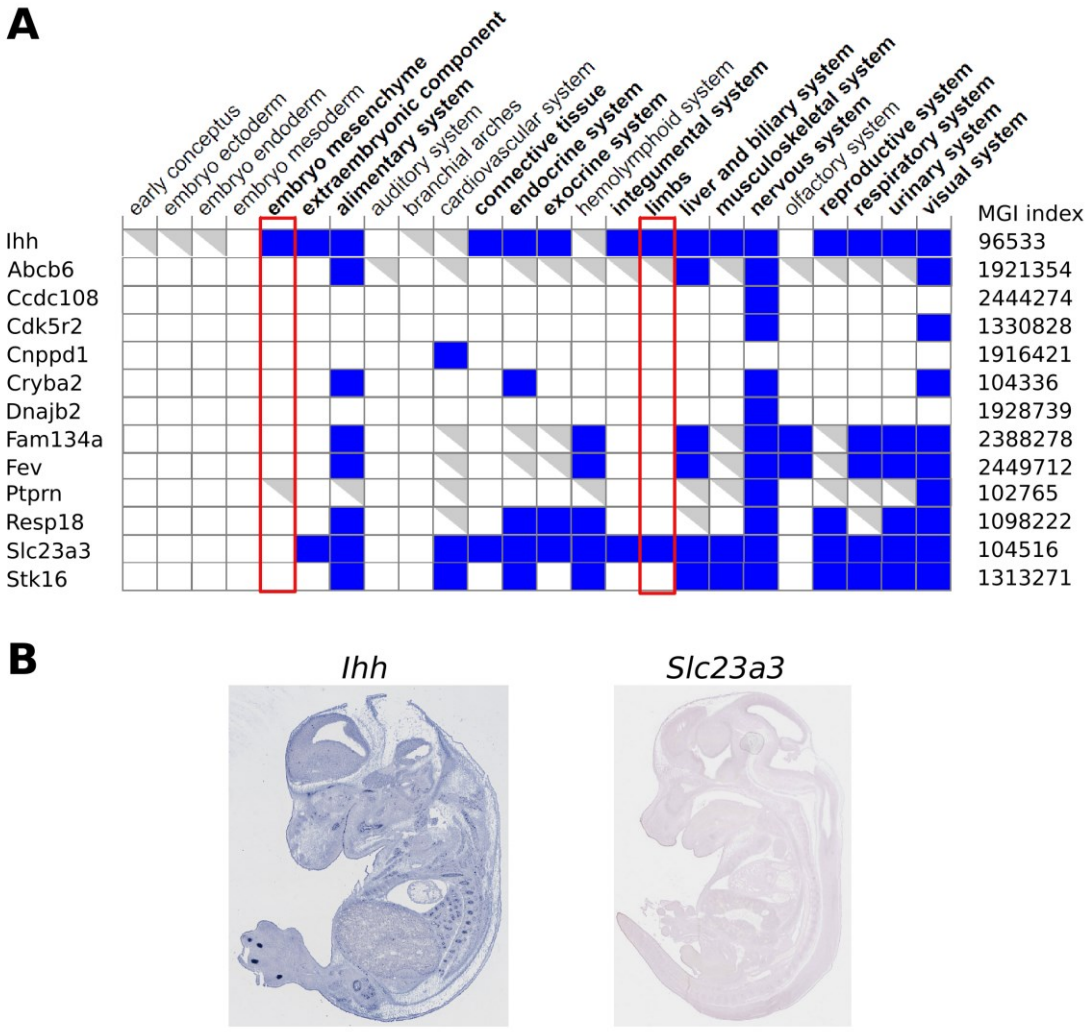

Figure 31. Expression overview of *Ihh* neighboring genes by MGI gene Expression Query

(A) Expression Matrix. Blue rectangles label cells/tissues that show gene expression in wt. Grey triangles represent other expression annotation (e.g. data of mutants). White rectangles indicate no expression. Only Slc23a3 was reported to be expressed in limb, however the published pattern does not correlate with the Ihh expression pattern. Data received from http://www.informatics.jax.org/marker/. (B) Sectioned WISH of E13.5 embryos; top Ihh (euxassay_000385_19), bottom Slc23a3 (euxassay_002900_1). Data received from http://www.informatics.jax.org/.

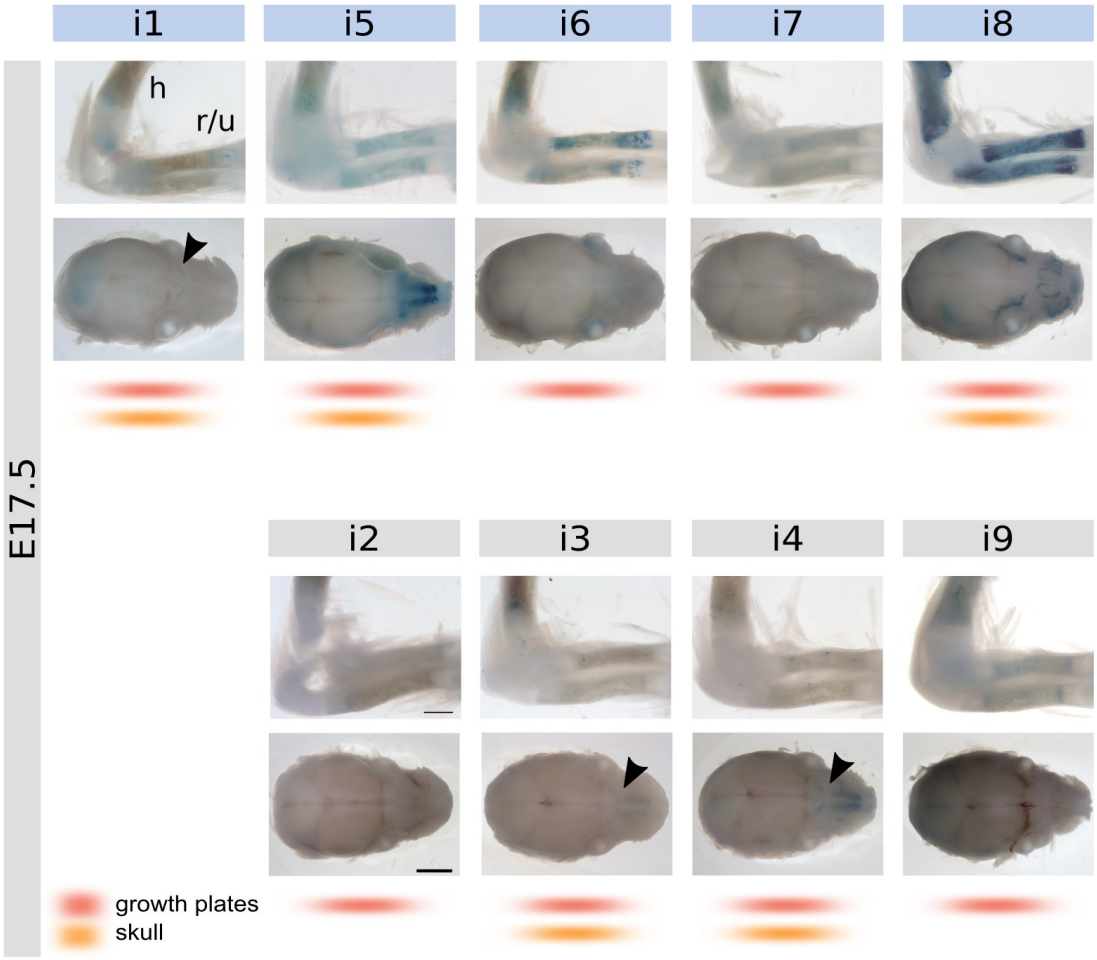

Figure 32. LacZ reporter assays of E17.5 embryos

LacZ reporter assay of Sleeping Beauty (SB) insertion and individual elements (blue: E14.5 positive, grey: E14.5 negative). Positive expression domains are indicated by color-coding: growth plates (red) and skull (orange) at E17.5. h: humerus, r/u: radius/ulna. Bars: E17.5 1000µm growth plates and 2000µm head.

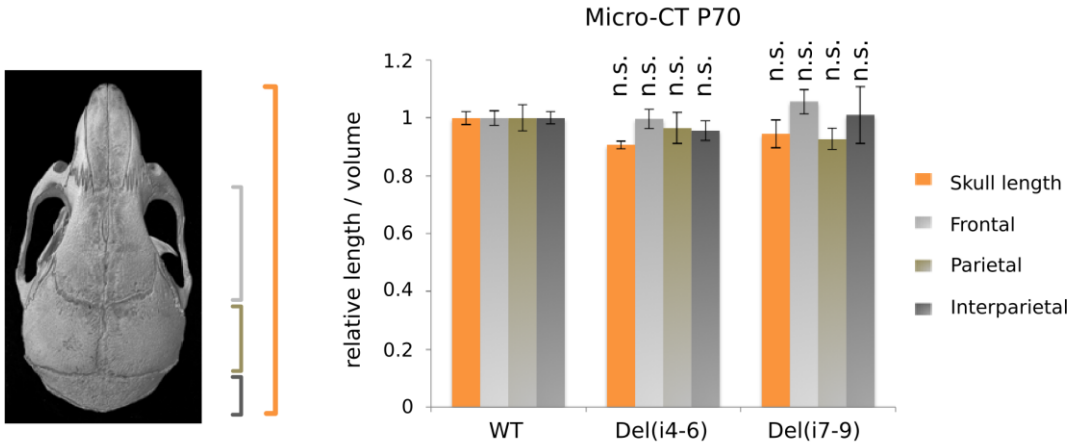

Figure 33. Skull Micro-CT of deletion constructs

Micro-CT analysis and length measurements, normalized to wt. The measured region is indicated in the schematic on the left. *P* values were calculated with two-sided Student's t-test (n≥3).

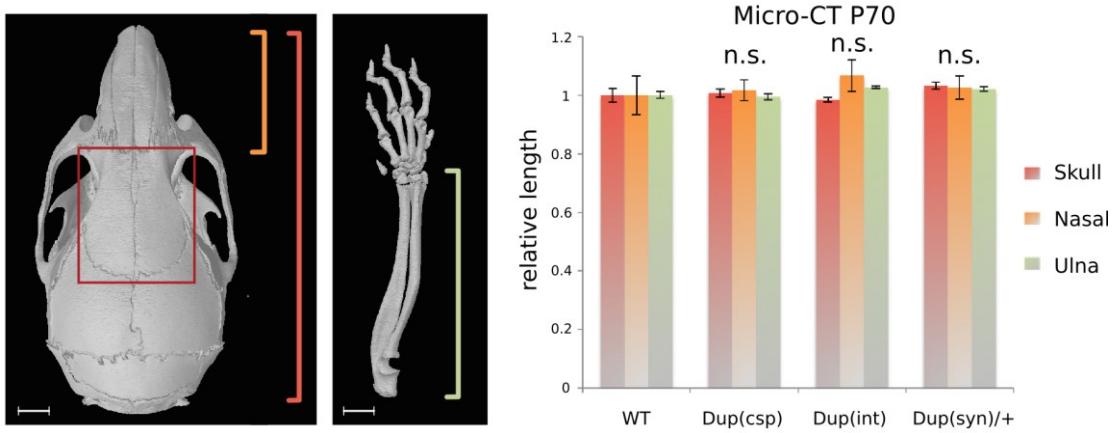

Figure 34. Micro-CT measurements of duplication lines

Micro-CT analysis of P70 mutants. The schematic of the analyzed region is shown on the left. Length measurements of skull, nasal suture and ulna revealed no significant changes. *P* values were calculated with two-sided Student's t-test (n≥3).

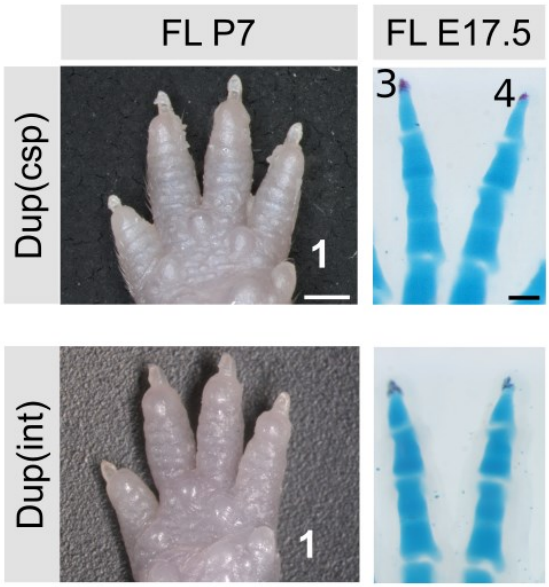

Figure 35. Forelimb phenotype of Dup(csp) and Dup(int) mutants

Phenotype analysis of P7 and Skeletal staining of E17.5 mutants shows normal forelimb development. Bars: 1000µm P7, 200µm E17.5 skeletal staining.

10 Scientific Productions

10.1 Publications

A manuscript gathering all the results presented in this thesis is currently under revision. In addition, another publication, which I signed as co-author, emerged from an internal collaborative contribution.

Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, Heinrich V, Vingron M, Klopocki E, Visel A, Lupiáñez DG, Mundlos S. Composition and dosage of a multipartite enhancer cluster control developmental gene expression at the *Indian hedgehog* locus. Currently under second revision at Nature Genetics.

Kraft K*, Geuer S*, **Will AJ**[1], Chan WL, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, Kragesteen BK, Spielmann M, Mundlos S, Lupiáñez DG, Andrey G. (2015) Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Reports 10†, 833839. *co-first author, [1]second author

10.2 Scientific communications

Oral presentation at ESHG: The European Human Genetics Conference, Copenhagen, Denmark, 2017. Nominated for Young Investigator Award. Excellence Travel Grant holder.

Poster presentations at the PhD retreat of the Max-Planck-Institute for Molecular Genetics, Krippen, Germany, 2016.

Poster presentations at Cell Symposia: Transcriptional Regulation in Development and Disease, Chicago IL, USA, 2016.

Poster presentations at the PhD retreat of the Max-Planck-Institute for Molecular Genetics, Potsdam, Germany, 2015.

Poster presentations at ESHG: 3rd Course of Next Generation Sequencing, Bertinoro, Italy, 2014. Scholarship holder.

Poster presentations at ESHG: 26th Course of Medical Genetics, Bertinoro, Italy, 2013. Scholarship holder.

Oral presentation at the PhD retreat of the Max-Planck-Institute for Molecular Genetics, Brandenburg an der Havel, Germany, 2012.

11 List of Figures

12 List of Tables

13 Abbreviations

Genome Organization and Transcription	
bp	basepair
kb	kilobase
Mb	megabase
3C	chromosome conformation capture
4C	circular chromosome conformation capture
Chip-Seq	Chromatin Immunoprecipitation Sequencing
CTCF	CCCTC-binding factors
DNA	Deoxyribunucleic acid
cDNA	complementary DNA
k.o.	knock out
PIC	preinitiation complex
RNA	ribunucleic acid
RNA-Pol	RNA-Polymerase II
TAD	topological associated domains
TF	transcription factor
wt	wildtype
X-Gal	bromochloroindoxyl-Galactosidase

Genome Modification	
DSB	double strand break
HDR	homology directed repair
NHEJ	non-homologous end joining
TAMERE	trans allelic targeted meiotic recombination
Cre	cyclization recombination
CRISPR	clustered regulatory interspaced palindromic repeats
Cas	CRISPR associated
crRNA	CRISPR RNA
tracrRNA	trans-activating crRNA
sgRNA	single guide RNA
PAM	protospacer adjacent motif
CRISVar	CRISPR/Cas-induced structural variants

Disease	
ENCODE	Encyclopedia of DNA Elements
DECIPHER	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
CNV	copy number variation
csp	Craniosynostosis Philadelphia type
syn / SD	Syndactyly

Limb Development	
Sox9	SRY-box9
BMP	bone morphogenic protein
SHH	sonic hedgehog
ZRS	ZPA regulatory seuence
Ihh	indian hedgehog
Ptc	Patched
Gli	Glioma associated oncogene homolog 1
PTHrP	parathyroid hormone-related peptide
Runx	runt-related transcription factor
RA	Retinoic acid
GREM	Gremlin
e-m	epithelial-mesenchymal
ECM	extracellular matrix
Col	collagen
AER	apical ectodermal ridge
PZ	progress zone
ZPA	zone of polarizing activity
ICD	interdigital cell death
ANZ, PNZ	anterior, posterior necrotic zone
Raldh	retinaldehyde dehydrogenase 2
Mapk	mitogen activated protein kinase

Material/Methods	
ON	over night
RT	room temperature
p	pico
n	nano
µ	micro
m	mili
l	liter
m	meter
M	molar
h	hour
min	min
sec	sec
H2O	water
dd	bidest
PCR	Polymerase chain reaction
qPCR	quantitative real-time PCR
WISH	whole mount in situ hybridization
SB	Sleeping Beauty
cDNA	complemenatry DANN
°C	grad celcius
ESC	embryonic stem cells
MEF	murine embryonic fibroblasts
E	embryonic stage