Regulation of Transcription and 3D Chromatin Architecture in Instructed Cell Differentiation

Karol Nowosad

Uitgegeven in eigen beheer - Publication privately printed – Karol Nowosad

Cover-design: Aneta Malesa

The work presented in this joint doctorate thesis was performed in the Department of Biomedical Sciences, Medical University of Lublin (MUL), Poland, and the Department of Cell Biology at Erasmus University Medical Center (Erasmus MC), Rotterdam, The Netherlands. In the Polish department, the PhD training was coordinated by the Postgraduate School of Molecular Medicine. PhD education and training at the Erasmus University Rotterdam (EUR) was coordinated through the Erasmus MC Graduate School. The Department of Cell Biology is also member of the Erasmus MC – Leiden University joint research school *Medisch Genetisch Centrum (MGC) Zuid-West Nederland*. This booklet is the dissertation presented in partial fulfilment of the requirements for the doctorate degree at the MUL and EUR, respectively.

Printed by

RIDDERPRINT

Ohmweg 17, 2952 Alblasserdam, The Netherlands

www.ridderprint.nl

Regulation of Transcription and 3D Chromatin Architecture in Instructed Cell Differentiation

Regulatie van transcriptie en 3D chromatine
architectuur in geïnstrueerde celdifferentiatie

Regulacja transkrypcji oraz 3D architektura chromatyny
w ukierunkowanym różnicowaniu komórek

Thesis

to obtain the joint doctorate degree from the Erasmus University Rotterdam
and the Medical University of Lublin by command of the
rector magnificus Prof. Dr. A.L. Bredenoord
and rektor Prof. Dr. Wojciech Załuska
and in accordance with the decisions of the respective Doctorate Boards

The public defenses shall be held on
Wednesday, February 8, 2023, at 1:00 p.m. in Rotterdam, The Netherlands, and
(2023, date/time t.b.d.) in Lublin, Poland
by

Karol Nowosad

born in Włodawa (Poland)

Erasmus University Rotterdam

Medical University of Lublin

Doctoral promotion committees at the respective universities

Erasmus University Rotterdam

Inner doctorate committee

Promoter:	Prof. Dr. D.F.E. Huylebroeck
Other members:	Assoc. Prof. Dr. G. Jansen
	Prof. Dr. M.K. Koblowska
	Prof. Dr. G.J.V.M. van Osch
Co-Promoter:	Prof. Dr. P. Tylżanowski

Medical University of Lublin

Inner doctorate committee

Promoter:	Prof. Dr. P. Tylżanowski
Other members:	Assoc. Prof. Dr. G. Jansen
	Prof. Dr. M.K. Koblowska
	Prof. Dr. E. Płuciennik
Co-Promoter:	Prof. Dr. D.F.E. Huylebroeck

Table of contents

List of used abbreviations

List of contents

List of used abbreviations

+:	positive
3C:	chromatin conformation capture
3D:	3-dimensional
5caC:	5-carboxylcytosine
5fC:	5-formylcytosine
5hmc:	5-hydroxymethylcytosine
5mC:	5-methylcytosine
ac:	acetylation
aCGH:	array comparative genome hybridization
AER:	apical ectodermal ridge
AGO:	argonaute
A-P:	anterior-posterior
AS:	anti-sense
ASC:	adipose-derived stem cell
ATAC-seq:	assay for transposase-accessible chromatin using sequencing
BD:	brachydactyly
BER:	base excision repair
bHLH:	basic helix-loop-helix
BMP:	bone morphogenetic protein
BMPR:	BMP receptor
bp:	basepair
CAGE:	cap analysis of gene expression
CD:	campomelic dysplasia
cDNA:	complementary DNA
CE:	candidate enhancer
CHH:	cartilage-hair hypoplasia
ChIA-PET:	chromatin interaction analysis by paired-end tag sequencing
ChIP:	chromatin immunoprecipitation
cKO:	conditional knockout
CNS:	central nervous system
CPSF:	Cleavage Polyadenylation Specificity Factor
CRISPR:	clustered regularly interspaced short palindromic repeats
CT:	chromosome territory
CTRL:	control
D:	day
Δ:	deletion
DacR:	differentially acetylated region
DEA:	differential gene-expression analysis
DEG:	differentially expressed gene (or deregulated expressed gene, ch7)
DM:	dermomyotome
DNase-seq:	DNase-I hypersensitive sites sequencing
DNMT:	DNA-methyltransferase
DPE:	downstream core promoter element
dpp:	decapentaplegic
dsDNA:	double-stranded DNA
D-V:	dorsal-ventral
E:	embryonic day
EB:	embryoid body
EMT:	epithelial-to-mesenchymal transformation
E-P:	enhancer-promoter
EpiLSC:	epiblast-like stem cell
eRNA:	enhancer RNA
ESC:	embryonic stem cell
FACS:	fluorescence activated cell sorting
FBS:	fetal bovine serum
F-syndrome:	Feingold-syndrome
GDF:	growth and differentiation factor
GO:	gene ontology
GRN:	gene regulatory network

GTF:	general transcription factor
GWAS:	genome-wide association study
GWBS:	genome-wide binding sites
hpf:	hours post-fertilization
ICM:	inner cell mass
IGF:	insulin-like growth factor
Inr:	initiator element
iPSC:	induced pluripotent stem cell
ISS:	idiopathic short stature
KD:	knockdown
KO:	knockout
LAD:	lamina-associated domain
lncRNA:	long non-coding RNA
LPM:	lateral plate mesoderm
LWD:	Leri-Weill dyschondrosteosis
Mb:	megabase
MDiGS:	multiplexed direct genomic selection
me:	methylation
MEL:	mouse erythroleukemia
mESC:	mouse embryonic stem cell
miRNA:	microRNA
MOWS:	Mowat-Wilson syndrome
MPRA:	massive parallel reporter assay
mRNA:	messenger RNA
MTE:	motif ten element
ND:	neural differentiation
NGS:	next-generation sequencing
NL:	nuclear lamina
NPC:	neural progenitor cell
NPC:	neuroprogenitor cell
nt:	nucleotide
O/N:	overnight
OA:	osteoarthritis
OPC:	oligodendrocyte precursor cell
PBS:	phosphate-buffered saline
PCA:	principal component analysis
PcG:	polycomb group
P-D:	proximal-distal
PE:	putative enhancer
PIC:	pre-initiation complex
PNS:	peripheral nervous system
PRC:	Polycomb Repressor Complex
PRO-cap:	precision nuclear run-on sequencing variant
PSC:	pluripotent stem cell
p-Smad:	phospho-Smad
PTM:	post-translational modification
PWM:	position weight matrix
qPCR:	quantitative polymerase chain reaction
RA:	retinoic acid
RE:	regulatory element
RISC:	RNA-induced silencing complex
RNAPol2:	(DNA-dependent) RNA-Polymerase II
RNase-A:	ribonuclease-A
RNA-seq:	RNA-sequencing
ROI:	region of interest
RPKM:	reads per kilobase per million mapped reads
rpm:	rounds per minute
RT:	reverse transcription
sc:	single-cell
SDS:	sodium dodecylsulphate
SHFM:	split hand/foot malformation
shRNA:	short-hairpin RNA

SNP:	single-nucleotide polymorphism
SNV:	single-nucleotide variation
SV:	structural variant
T2C:	targeted chromatin capture
TAD:	topologically associating domain
TE:	trophectoderm
TET:	ten-eleven translocation
TF:	transcription factor
TFBS:	transcription factor binding site
TGFβ:	transforming growth factor type β
TSS:	transcription start site
U:	unit
UMAP:	uniform manifold approximation and projection
UTR:	untranslated region
WT:	wild-type
ZPA:	zone of polarizing activity
ZRS:	ZPA regulatory sequence

Chapter 1

Part a

Introduction

1. Implicating the code of life: from genomic information to RNA

1.1. Regulations at different phases of the gene transcription cycle

Cellular responses to external signals (e.g., secreted growth/differentiation factors that act in embryogenesis and in adult tissues or organs) involve biochemical cascades that propagate the signal from the liganded receptor to the cell nucleus. Here, they result in appropriate changes in levels of mRNA expression of signaling pathway specific target genes, which affect cell proliferation, differentiation (lineage decisions, progression of differentiation), maturation and migration, but also e.g., cell survival vs. cell death (Selvamurugan *et al.*, 2007; Wang *et al.*, 2011; Stryjewska *et al.*, 2017; Liu *et al.*, 2018). In the embryo, this collectively assures cell allocation to the early embryonic germ layers and subsequent creation of remarkable cell diversity in post-gastrulation stages, including early and late organogenesis.

Transcription of most genes studied in eukaryotic cell differentiation is executed by DNA-dependent RNA-Polymerase II (RNAPol2) together with general transcription factors (GTFs), and occurs grossly in three steps, i.e. initiation, elongation, and termination of transcription (**Fig. 1a.1**; for reviews and other references therein, *see* Cramer, 2019; Eaton and West, 2020). Within the eukaryotic cell nucleus, such mRNAs have to be co-transcriptionally capped at their 5'-end, displaying ^{m7}GpppN (where N is any nucleotide), a ubiquitous feature of these mRNAs. Also, splicing of initial transcripts leading to removal of non-coding information (introns) from the primary transcripts takes place in nucleus, as well as processing/maturation and polyadenylation at the 3'-end of the nascent mRNA (for reviews, *see* Ramanathan *et al.*, 2016; Nicholson and Pasquinelli, 2019; Kitamura *et al.,* 2021). All this happens prior to translation of the mRNA into protein, which furthermore happens outside of the nucleus. The mRNA can also be subject to post-transcriptional controls, while the encoded proteins can also be post-translationally controlled by e.g., modifications such as phosphorylation, acetylation, and glycosylation.

In order to initiate transcription, the gene promoter has to be activated by binding the aforementioned GTFs. The GTFs are crucial for recruitment of RNAPol2 to the promoter region, formation of the transcription pre-initiation complex (PIC), and the transition between initiation and productive elongation. The documented changes in RNAPol2 accumulation on a gene following a stimulus of the cell correlate well with observed steady-state level changes in mRNA. Elongating RNAPol2, which serves progression of the polymerase through the locus for lengthening the transcript, is most often biochemically identified by phosphorylation at mapped Ser residues of the carboxyl-terminal domain (Pol2 CTD) of its largest subunit Rbp1 (for reviews and other references therein, *see* Egloff and Murphy, 2008; Zhou *et al.*, 2012; Bowman and Kelly, 2014). A well-conserved characteristic of this CTD is that it is composed of multiple repeats of a 7 amino-acid long sequence, including Ser at its 2 and 5 position. For most of the protein-encoding genes, RNAPol2 is recruited with hypo-phosphorylated CTD, and Ser5 and Ser2 modification by phosphorylation then occurs and is typical for initiation and elongation, respectively. Some of the involved kinases have been well-studied, e.g., Cdk9 and Cdk12, for predominantly Ser2 phosphorylation. Cdk9 acts more at the 5'-end of genes, enhancing Cdk12 action further downstream (for a recent review on Cdk9 in transcriptional control in various

eukaryotes, *see* Bacon and D'Orso, 2019). This RNAPol2 phosphorylation on itself recruits transcription-associated proteins in each of the two phases.

Another post-recruitment regulatory mechanism is RNAPol2 pausing, which takes place between early elongation and productive elongation, in cells of higher eukaryotes (and less in yeast). This pausing, initially biochemically detected by highest RNAPol2 localization, occurs in many genes about 25-60 basepairs (bp) downstream of the transcription start site (TSS). RNAPol2 pausing is characterized by quick transition from the PIC to the paused state of RNAPol2, and can last for several minutes. There are many opinions and hypotheses about the purposes of RNAPol2 pausing, including its contribution to fine-tuning and quality control of expression (for reviews and other references therein, *see* Core and Adelman, 2019; Abuhashem *et al.,* 2022). Importantly, pausing defects caused by mutation in genes encoding pausing factors (including those shown to act in release of pausing) have been identified as a mechanistic cause of intellectual disability (ID) syndromes (*see* van den Berg *et al.*, 2017).

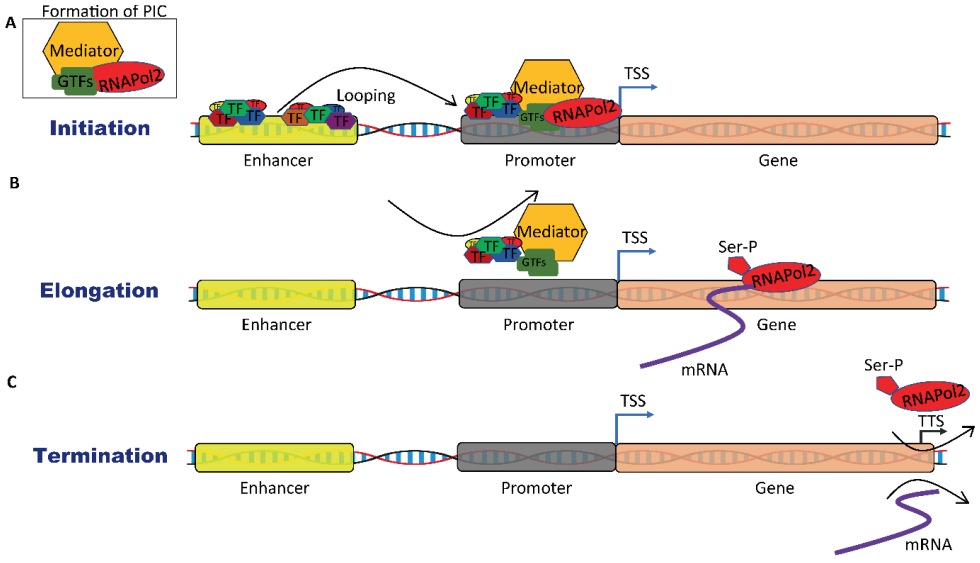


Figure 1a.1. Summary of the 3-step transcription cycle.

A. Prior to transcription the pre-initiation complex (PIC) is formed by GTFs, the Mediator, and RNAPol2. To initiate transcription, this PIC has to recognize and bind to the promoter region. Also, bound TFs are bridged from enhancer to promoter and facilitate transcription. **B.** Further, the RNAPol2 leaves the promoter and starts mRNA synthesis, while proceeding thourgh the gene body towards the transcription termination site (TTS). **C.** RNAPol2 is released from the DNA and RNA when reaching the TTS.

Elongation is terminated when RNAPol2 reaches the transcription termination site, and the polymerase is released. Specialized proteins with high affinity to this site thus stop RNAPol2, but the transcript still has to undergo maturation. Cleavage Polyadenylation Specificity Factor (CPSF) complex cleaves the RNA transcript, which further enables poly(A) addition by poly(A)-polymerase at the 3'-end of the nascent RNA (for more details or reviews on canonical pre-mRNA 3'-end processing, and its relevance to alternative poly-adenylation and splicing as well, *see* Barabino *et al.*, 1997; Preker *et al.*, 1997; Kyburz *et al.*, 2006; Misra and Green, 2016; Sun *et al.*, 2020). Importantly, identified key subunits of CPSF travel with the RNAPol2 already

from the site of initiation of transcription, thereby mechanistically providing means to couple initiation to processing/maturation at the 3'-end of the mRNA (for a recent comment, in particular on the reconstitution *in vitro* of CPSF activity, *see* Yoon and Shi, 2022).

The transcription process is also regulated by the Integrator complex (often referred to more briefly as Integrator), a 14-subunit complex that is multi-functional and altogether controls and is crucial for the fate of nascent RNAs transcribed by RNAPol2, including enhancer RNAs (eRNAs) (for a review, with focus on functions in transcription and development, *see* Mendoza-Figueroa *et al.*, 2020). Mediator complex is a master regulator of RNAPol2-mediated transcription closely linked to enhancer-based gene expression (for a recent review, *see* Richter *et al.*, 2022) via enhancer-to-promoter (E-P) DNA-looping (for a recent discussion of the DNA-looping model, *see* Popay and Dixon, 2022). In addition, gene transcription is controlled by *cis* and/or *trans* regulatory elements (REs, e.g., promoter-proximal as well as distant enhancers, *see* section 1.2 below), epigenetic modifications, and rearrangements in 3D chromatin structure (*see* section 1.4).

1.2. *Cis*-regulatory elements control gene transcription

Protein-encoding genes represent a tiny fraction only (i.e. <2%) of the mammalian genome. The remaining genomic DNA in the nucleus contains also a few hundred thousand *cis*-REs, including not only promoter, but also enhancer, silencer and insulator sequences, respectively (Gao and Qian, 2020; Pang and Snyder, 2020). Putative *cis*-REs thus outnumber genes in genomes, underscoring the complexity of gene regulation. These REs have the ability to positively (i.e. drive, upregulate) or negatively (prevent, downregulate) regulate gene transcription. They can be located in close proximity to the gene or act from more distant locations, in many cases passing multiple genes (for review, *see* Schoenfelder and Fraser, 2019).

Cis-REs commonly contain binding motifs for more than one transcription factor (TF) (Quang *et al.,* 2015). These motifs are often organized into clusters, enabling specific interaction with TFs (and co-factors in their respective complexes), thereby orchestrating spatio-temporal expression of the cognate target genes with the required high precision, including at e.g., different cell states and developmental stages (for review, *see* Spitz and Furlong, 2012).

The presence and use of multiple REs per gene increase the combinatorial complexity of gene expression and are needed for accurate and precise regulation of spatio-temporal gene expression. e.g., during embryogenesis (Perry *et al.,* 2011). Often this specific gene control is then part of a gene regulatory network (GRN, with multiple genes), with the latter often marked by detectable characteristics of synexpression, autoregulation, and positive or negative feedback of sometimes low numbers of genes (Devkota and Wuchty, 2022).

Importantly, genomic variants affecting the activity of individual REs may result in detectable alteration of expression of steady-state mRNA level(s) and hence be disease-causing or lead to developmental defects. Such defects include limb malformations and neurodevelopmental syndromes, the focus of this PhD. For examples of REs and their association with congenital disorders, we refer to **Chapter 1b**. In the following sub-sections different types of *cis*-RE will be described.

1.2.1. Promoters

Promoter sequences direct transcription initiation from one (for many genes more than one) TSS by binding several components of the basal/general transcriptional machinery (the GTFs mentioned above) and DNA-dependent RNA-Polymerase (RNAPol), e.g., RNAPol2 (Sloutskin *et al.,* 2021). Promoters can differ in their core elements and sequence. In addition, the presence of methylation-sensitive CpG dinucleotides sometimes arranged as CpG islands near the promoter dictate transcriptional repression vs. activation. Further promoter diversity comes from histone modifications, specific nucleosome positioning, and transcription initiation itself (*see* next paragraph; Haberle and Lenhard, 2015). Promoters are often divided into two groups, the "focused" and "dispersed" promoters, respectively, based on mechanistic differences in transcriptional initiation.

The "focused" promoters, also termed "sharp peak" promoters, have been connected to control of developmental genes, which characteristically display and require precise spatio-temporal gene expression (Frith *et al.,* 2014). They consist of a single TSS or multiple TSSs located within a short region, with the core promoter located from -40 to +40 bp from the unique or main TSS. Focused promoters often contain core promoter elements such as the TATA box, motif ten element (MTE), initiator element (Inr), and downstream core promoter element (DPE) (Carninci *et al.,* 2006; Danino *et al.,* 2015; Sloutskin *et al.,* 2021). The "dispersed" promoters, also termed "broad" promoters, contain multiple weakly-defined TSSs spread over a 50-100 bp-long DNA-segment, and have been found in over 70% of promoters in vertebrates. They are mainly linked to housekeeping or near-constitutively expressed genes, and often contain CpG islands and binding sites for the TFs Sp1 and NF-Y (Carninci *et al.,* 2006; Sloutskin *et al.,* 2021).

A universal subset of core promoter elements allowing to precisely define the promoter type has not been defined as yet, and novel core promoter motifs are being proposed or identified. Recent studies showed that transcription initiation may be more complex and also involve a combination of "focused" and "dispersed" characteristics (Ni *et al.,* 2010; Frith *et al.,* 2014; Sloutskin *et al.,* 2021). Furthermore, new technologies have prompted the field to revise the classification of promoter regions. For example, application of cap analysis of gene expression (CAGE) tends to divide gene promoters into four classes: single dominant peak, broad peak, bimodal/multimodal peak, and broad with dominant peak, respectively (Carninci *et al.,* 2006). Another technology used for promoter classification is chromatin immunoprecipitation (ChIP), which led to extra characterization of active/inactive promoters (Kim *et al.*, 2005), regardless of the aforementioned sharp vs. broad peak initiation mode. This classification was performed based on the investigation of the binding regions of the GTF TFIID and RNAPol2, and also histone modification such as H3-acetylation (H3ac) and H3K4-methylation (H3K4me).

Importantly, promoters do not contain all of the information necessary for the precise spatio-temporal regulation of gene expression: physical proximity of the promoter to other *cis*-regulatory elements such as enhancers (often referred to as "interaction", *see* section 1.2.2) facilitates transcription, including by physically bridging additional enhancer-bound TFs to the promoter region.

1.2.2. Enhancers

Enhancers typically are relatively short (up to 1,000 bp) DNA-sequences with *cis*-RE activity. They can however also modulate transcription of a gene located on a different chromosome (hence act in *trans*; Monahan *et al.*, 2019). Enhancers orchestrate place (i.e. cell type), timing and/or level of transcription of their target genes either from close proximity or large distances (for a recent review, *see* Ibragimov *et al.*, 2020; Nair *et al.*, 2002). Further, enhancers are distributed genome-wide, mostly within non-protein-coding regions, including introns (and often the first intron of a gene), but also in intergenic regions sometimes presenting as gene desert over a significantly longer section of the chromosome. Enhancers can however also be found in exons of protein-coding genes (Birnbaum *et al.,* 2012).

How enhancers identify and activate their cognate target genes is a fundamental question of genome biology and remains an active area of research. Enhancers together with multiple co-expressed genes are often located within the borders of regulatory domains, such as topologically associating domains (TADs), which promote enhancer-promoter (E-P) physical proximity (**Fig. 1a.2;** Dixon *et al.,* 2012; Dixon *et al.,* 2015; Galouzis and Furlong, 2022; for a discussion of TADs and their role in gene regulation, *see* section 1.3.2). This physical proximity of enhancer and its cognate promoter bursts transcription, whereas disruption of such proximity can result in gene downregulation (Lupiáñez *et al.,* 2015). Observed co-expression of genes within the same genomic regulatory domain suggests that enhancers can also interact with multiple promoters (Nora *et al.,* 2012). Indeed, such "promiscuity" of enhancers was for example shown by CRISPR-based enhancer removal, which caused downregulation of multiple genes (Guerrero-Martínez *et al.,* 2020).

Many TADs contain genes characterized by a distinct spatio-temporal expression domain (including cell-state and cell-type specificity), co-supporting the notion of enhancers' gene-specificity. Chromosome conformation capture (3C) studies revealed that enhancers located multiple genes away from the target gene can "skip" non-cognate genes in exclusively activating their target promoter(s) (Williamson *et al.,* 2016). Also, multiple enhancers for a given gene often act additively or synergistically (Birkhoff *et al.*, 2020; Choi *et al.,* 2021).

Enhancers are a functionally more heterogeneous collection of REs than initially expected. This can experimentally be reflected by differences in activity of enhancers, e.g., as documented in massive parallel reporter assays (MPRAs; Arnold *et al.,* 2013; Vanhille *et al.,* 2015). This functional heterogeneity might very well be partially explained by the enhancer sequence. Enhancers contain multiple TF binding sites (TFBSs) organized into clusters located in close proximity to each other, and are organized in a specific fashion (Grossman *et al.,* 2018). Such "syntax" of TFBSs predetermines the set of TFs interacting with the enhancer body, which brings the needed TFs into spatial proximity of target promoter(s). Conversely, the core promoter structure can also (co-)mediate the E-P connection and (co-)determine the specificity of such interactions. For instance, experimental changes in a genomic locus encompassing three genes in the fruit fly, i.e. *dpp, slh, oaf*, and also the enhancers of *dpp*, showed that replacement of *oaf* promoter with the *dpp* promoter results then in activation of *oaf* through the *dpp* enhancers (Merli *et al.,* 1996). This indeed strongly suggested that specific features of promoters can guide enhancers.

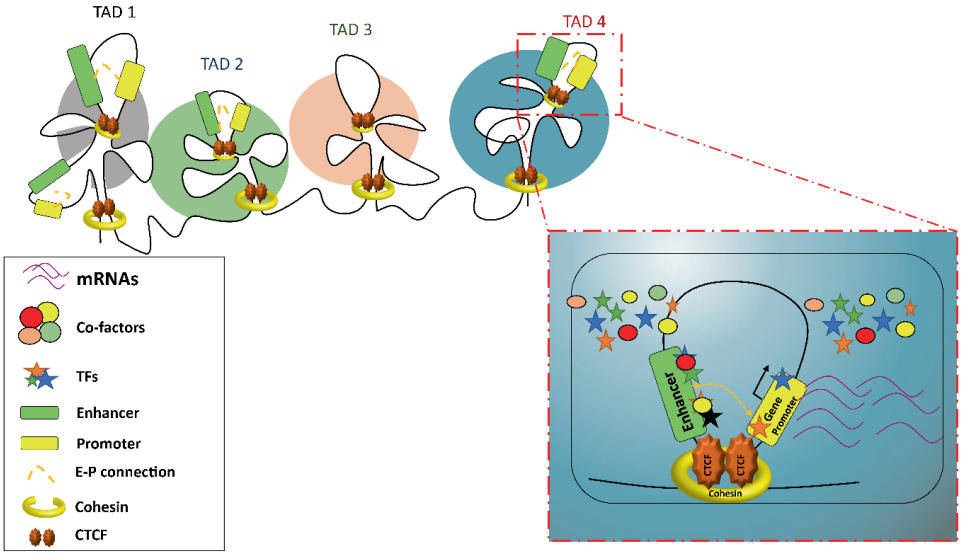


Figure 1a.2. Enhancers regulate target genes within TADs.

Topologically associating domains (TADs) acts as regulatory units within which enhancers and promoters loop to each other, facilitated by cohesin-CTCF dependent loop extrusion mechanims. In contrast, the inter-TAD enhancer promoter (E-P) connections are insulated by boundaries located between TADs that are enriched in CTCF binding regions.

One of such specific features can be the presence of motifs for TF-binding in or around the promoter region. Therefore, the combination of both enhancer-specific and promoter-specific TF-binding can attract and then orchestrate the transcriptional machinery, leading to gene activation. Indeed, the control of expression of TF-encoding genes in a spatio-temporal manner (Lefebvre *et al.,* 1997; Xie *et al.,* 2017), as well as changes in TF protein overall concentration (Chopra and Levine 2009) or local concentration (e.g., in transcription foci or factories; Furlong and Levine, 2018; Kimura and Sato, 2022), can be directly linked to higher gene expression and activation of enhancers.

However, simple correlation of these two variables (presence of TF motif and expression of the TF-encoding gene) is not sufficient to build prediction models of global enhancer-driven gene expression regulation (Zeitlinger, 2020). This suggests that other factors also play a role in enhancer control, creating a multi-level and thus complex pattern of regulation. This complexity can include many cause-effect sequences of action, and different mechanistic crosstalk. For instance, binding of TFs to an enhancer core region can lead to co-operative binding of co-factor(s) with acetyltransferase activity (e.g., P300), which acetylate K27 on histone H3 (Narita *et al.*, 2021; for description of histone modifications and their role in gene regulation, *see* section 1.3.1). Adding extra acetyl groups to the histone tail reduces the positive charge of histone octamer towards more neutral, consequently changing the affinity between histones and DNA (Onufriev and Schiessel, 2019). Such reduction of affinity leads to increased nucleosomal eviction from the enhancer region, elevating the degree of chromatin accessibility, thus removing a potential "barrier" for binding of another key TF (Guerrero-Martínez *et al.,* 2020). Therefore, the presence of high numbers of TFBSs at regulatory regions correlates with enhancer activity,

but this activity varies between enhancers, as demonstrated with endogenous as well as synthetic enhancers (Burz *et al.,* 1998; Erceg *et al.,* 2014).

Enhancers can also be activated by extrinsic signaling pathways coupled to intrinsic signal transduction cascades, which enables cells to appropriately and dynamically respond to their changing micro-environment (*see* also section 1.1). For instance, the treatment of epithelial cells (such as Nme cells, from normal mouse mammary gland) with TGFβ increases chromatin accessibility, which is correlated with enhancer activation typical for the induced epithelial-to-mesenchymal transition (EMT). Surprisingly, in this case the changes in chromatin accessibility are global, for they occur at repressed enhancer regions and enhancers which are not directly regulated by TGFβ (Guerrero-Martínez *et al.,* 2020). This suggests that the mechanism regulating the "open state" of chromatin is promiscuous. Genomic context is another factor that regulates the activity of an enhancer. The same mutation within a well-characterized enhancer region can have a different effect on its activity when the mutant enhancer is placed outside of its endogenous locus (and/or is tested in an ectopic reporter assay) (López-Rivera *et al.,* 2020).

Enhancers can be divided into different states based on many different (and the list is still growing) histone post-translational modifications (PTMs; as documented by biochemistry of histones, and as mapped in the genome by ChIP-seq, *see* below in this section). These are a part of epigenetic mechanisms that contribute to gene expression regulation (*see* also section 1.4.1). Active enhancers are characterized by acetylation of K27 on histone H3 (H3K27ac) and mono-methylation of K4 on H3 (H3K4me1, especially at distal enhancers, while the tri-methylation H3K4me3 is found more at promoters) (Creyghton *et al.*, 2010); poised enhancers are marked by H3K27me3 (often relating to Polycomb Complex (PCR) repressed gene loci, and also characterized by absence of DNase-I hypersensitive sites) and the absence of H3K27ac (Rada-Iglesias *et al.,* 2010). These most commonly analyzed histone marks are broadly utilized for identifying candidate active enhancers, and their genomic mapping is mostly obtained by ChIP followed by massive parallel sequencing (ChIP-seq). Notably, ChIP-seq often routinely uses formaldehyde and specific antibodies (e.g., anti-H3K27ac) to crosslink proteins with DNA and then immunoprecipitate the histones with the respective specific modification.

ChIP-seq has facilitated the genome-wide discovery of few hundred thousand enhancers and the accompanying development of atlases containing REs specific for widely used cell lines or FACS-enriched/purified primary cells (Monti *et al.*, 2017; Gao and Qian, 2019). However, this technique does not allow to investigate the E-P proximities/contacts and identify true target gene(s) controlled by these enhancers. For this, one uses chormatin conformation capture (3C) technology, and 3C derivatives (4C, Hi-C, Capture-C, T2C etc), which map such interaction between enhancers and promoters in the context of 3D chromatin architecture (Hughes *et al.*, 2014; Kolovos *et al.*, 2018). 3C-based methods have been used to describe the changes in chromatin architecture which affect E-P contacts and can lead to enhancer acquisition or adoption (Lupiáñez *et al.,* 2015). This latter mechanism involves genomic rearrangements, leading to genomic relocation of enhancers into the already existing *cis*-regulatory landscape of genes previously not controlled by this novel RE. Additionally, 3C-based techniques have been used to characterize the correlation between loss of E-P interaction and reduction of steady-state mRNA expression levels, which can also be directly linked to congenital malformations (for details, *see* **Chapter 1b**).

Enhancers share similar features with different *cis*-REs. They can act as promoters, hence not only bind RNAPol2, but also lead to transcription of often unstable RNA, called enhancer (e)RNA (Santa *et al.,* 2010). Transcription of eRNA is associated with enhancer activity, however

not all active enhancers produce eRNA (Chen and Liang, 2020). The production of eRNA (and its eventual stability) can depend on the direction of eRNA transcription (Chen *et al.*, 2013; Mikhaylichenko *et al.*, 2017; Ibrahim *et al.*, 2018). However, the directionality of transcription and the levels vary also among enhancers (Mikhaylichenko *et al.*, 2017). Interestingly, when an active enhancer is located in one of the gene's introns (often the first intron, *see* above), it can produce alternative transcript isoforms of such gene (Kowalczyk *et al.*, 2012). Conversely, bi-directional promoters can also act as enhancers, unlike uni-directional promoters (Mikhaylichenko *et al.*, 2017). Such promoter-promoter (P-P) loops are considered similar to E-P ones, and are for example reorganized during different states of embryonic stem cell (ESC) pluripotency (Joshi *et al.,* 2016). This finding suggests that P-P interactions are also dynamic, play a role in organization of chromatin architecture and may be involved in gene transcription regulation.

It has to be noted that the operational definition of a promoter region is not consistent in the literature, for some studies define it for example as the region -/+ 1kb from the TSS, -/+ 1.5kb or -/+ 2kb from it (Li *et al.,* 2015; Dao and Spicuglia, 2018; Grabowicz *et al.,* 2021). Therefore, there is always a chance that in some cases the defined promoters encompass both the core promoter and promoter-proximal enhancers. Last, but not least, enhancers can also act as silencers (*see* section 1.2.3 below).

1.2.3. Silencers

Silencers are *cis*-REs that negatively regulate activity of target promoters, leading to reduction or prevention of gene expression (Segert *et al.,* 2021). Like enhancers, silencers are distributed genome-wide, within intergenic regions or protein-coding gene introns (Pand and Snyder, 2020), acting from close proximity to the gene body or from a longer range. Importantly, the majority of silencers operate within the TADs, similarly to enhancers (for more details about TADs, *see* section 1.4; Pang and Snyder, 2020; for recent reviews, *see* Liu *et al.*, 2022; Zhang *et al.*, 2022). They can act independently of genomic position and orientation, and maintain silencer activity when taken out of their native genomic context (Qi *et al.,* 2015; Segert *et al.,* 2021).

Recent genome-wide screens have shown that silencers are mostly cell-type and/or developmental-stage specific (Pand and Snyder, 2020). Surprisingly, they can also function as enhancers, either within one (Bandara *et al.,* 2020) or more different cell lines (Huang and Ovcharenko, 2022). Therefore, their dual function highlights the complexity of silencers and also partially explains why they have been relatively under studied in comparison to other *cis*-regulatory elements.

How silencers regulate gene expression is not yet fully understood, and many mechanisms have been proposed. One mechanism suggests that silencers bind TFs with intrinsic transcription-repressive function and also bridge this to promoters (**Fig. 1a.3, panel A;** Segert *et al.,* 2021). Indeed, motif enrichment analysis of silencers shows enrichment of TFBSs within silencers, especially for such TFs known to negatively regulate gene transcription (Pand and Snyder, 2020). Also, investigation of the dual-function silencers revealed that these REs are enriched for binding motifs for both transcriptional repressors and activators (Huang and Ovcharenko, 2022). Therefore, it is possible that such dual-function REs act in one way, depending on the concentration of TFs in the cell, or that the TFs with high affinity to particular silencers compete with activators to bind such regions. Furthermore, many TFs have both repressive and activation properties, depending as to which co-factors bind to the TF and hence form a complex together at a given genomic locus (for reviews and other references therein, *see* Ma, 2005).

Another mechanism suggests that silencers can act as 'anti-looping factors', which block physical E-P connection (**Fig. 1a.3, panel B**; Hopra *et al.,* 2012). Importantly, neither of these mechanisms contradict each other, and together they provide an elaborate view of silencer-driven regulation of gene expression.

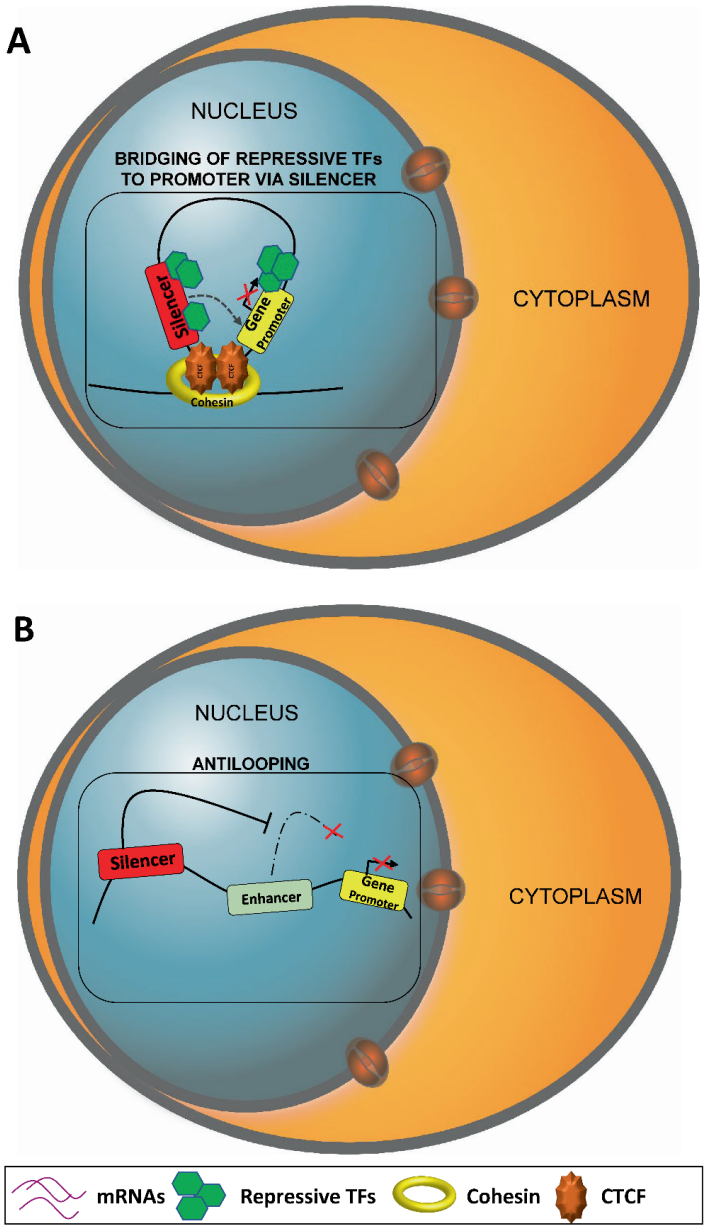


Figure 1a.3. Mechanisms of gene silencing.

A. Silencers can recognize and interact with the gene promoter, subsequently bridging transcription factors (TFs) with repressive function, which leads to gene silencing. **B.** Silencers can act as anti-looping factor, which blocks formation of the enhancer-promoter connection.

1.2.4. Insulators

Insulators reduce promiscuous gene expression by limiting the physical contact(s) of enhancer/silencer with target gene(s) (Scott *et al.,* 1999; Recillas-Targa *et al.,* 2002; West *et al.,* 2002; Yao *et al.,* 2003). To restrict the enhancer/silencer action, the insulator has to be located between such RE(s) and target promoter(s) (**Fig. 1a.4**). Therefore, insulators can be involved in establishment of independent domains of transcriptional activity, and promote E-P contacts within the region separated by insulators (Ali *et al.,* 2016). Indeed, such regulatory domains have been identified using techniques investigating frequencies of chromatin contacts, such as Hi-C (Wang *et al.,* 2016) (*see* section 1.4). Insulators may also prevent spreading of repressive chromatin, creating a "barrier" of active genomic regions (Prioleau *et al.,* 1999).

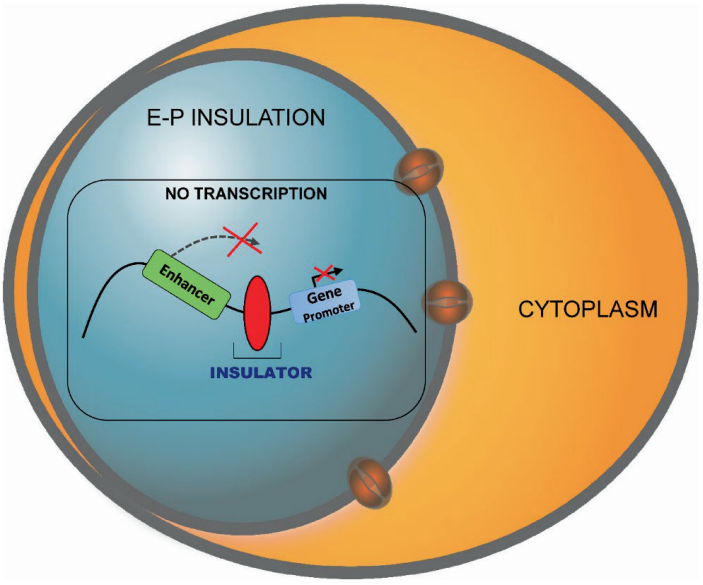


Figure 1a.4. Insulators block interactions between *cis*-regulatory elements.

Location of an insulator element between enhancer (E) and promoter (P) can block formation of the E-P chromatin loop and affect gene expression.

1.3. *Trans*-regulatory elements are also master regulators of gene expression

1.3.1. Introduction

Trans-regulatory elements (*trans*-REs) are typically factors such as TFs and co-factors, and also long non-coding RNA (lncRNA) and microRNA (miRNA) that interact with *cis*-REs. TFs regulate gene expression in a specific fashion, presenting a fixed syntax (defined as the number and order of e.g., their binding motifs), which predefine the set of TFs controlling the activity of the *cis*-REs (Spitz and Furlong, 2012). TFs often either co-operate with each other or compete for binding to the same genomic region providing a complex regulatory network involved in many biological processes, including cell (de)differentiation and even trans-differentiation (Xie *et al.,* 2004; Takahashi and Yamanaka, 2006; Stryjewska *et al.,* 2017). Despite decades of research and the discovery of about 1,500 TFs (Lambert *et al.,* 2018) and thousands of different lncRNAs and miRNAs (Lorenzi *et al.,* 2021) in the human genome, the comprehensive characterization of the gene regulatory network (GRN) orchestrated by the *trans*-REs is challenging and remains an active area research (Mattioli *et al.,* 2020). In the following sub-section, TFs will be discussed more in detail. For lncRNAs and miRNAs, and their role in development and disease, *see* **Chapter 1b**.

1.3.2. Transcription factors

TFs are classically defined as proteins that are involved in direct and selective regulation of gene/mRNA expression. TFs able to bind to specific, fairly sometimes degenerated (i.e. displaying a certain level of intrinsic variability), short DNA sequences (varying in length from 5 to 15 bp, depending on the TF family), called TF DNA-binding motifs. These motifs can be defined by position weight matrixes (PWMs), presenting an array of nucleotides organized into a specific order, with a defined frequency of each nucleotide. Importantly, TF DNA-binding motifs co-determine the TF binding pattern over the entire genome (Maerkl and Quake, 2007; Geertz *et al.,* 2012), but does not automatically equal a functionally active TF binding site, increasing the complexity of studying TFs. The syntax of the TF binding-motifs appears to be consistent across cell types, suggesting that TF binding can be associated with the structural or functional features of REs. For instance, TFs that interact with histones or with chromatin remodelers present enriched binding at the edges of chromatin-accessible sites, in contrast to TFs with a tendency to interact with other TFs, which bind to chromatin-accessible regions more centrally (Grossman *et al.,* 2018). This also suggests the presence of more complex mechanism(s) involved in the regulation of TF occupancy at functional elements, including *cis*-REs.

Enhancers and promoters are often enriched in DNA-binding motifs for different TFs, suggesting that TFs that occupy such REs in a combinatorial manner may co-operate with each other. Together with the overlapping expression in given cell types of TF mRNAs during e.g., embryonic development, this results in improved control of gene expression (Halfon *et al.,* 2000; Lettice *et al.,* 2012). TF binding may lead to either activation/upregulation or inactivation/downregulation of *cis*-REs, and depends on the nature of the TF, but also its co-factors with which it can form a complex (*see* above). Genetic interaction of TF repressors produced in more restricted regions with broadly expressed TF activators can lead to "stripe" gene expression patterns, as shown for multiple *gap* genes (also encoding TFs) during fly development (Stanojevic *et al.,* 1991).

Many TFs can also act as both activators or repressors, as in the case of Zeb2 studied in this PhD research (Conidi *et al.*, 2011; Birkhoff *et al.,* 2021) and Gli3, a DNA-binding TF acting downstream of the sonic hedgehog (*Shh*) signal (Buttitta *et al.*, 2003). Gli3 is also subject to proteasome degradation from which it can be actively protected by stabilizing factors (Sufu, an intracellular component of hedgehog signaling, Makino *et al.*, 2015; Zhang *et al.*, 2017; the TF Zic2, which binds to Gli3; Sun *et al.*, 2022). Importantly, such dual-function of TFs can thus be explained by their ability to co-operate with other DNA-binding TFs that bind to adjacent sites on DNA. Also, the dual-function of TFs can be obtained by direct protein–protein interactions of the TF with one or more co-factors, which do not necessarily bind to DNA (**Fig. 1a.5;** Lin *et al.,* 1990; Merika *et al.,* 1998). Another important element in the regulation of TF stability and/or activity is also its post-translational modification (PTM), such as phosphorylation and SUMOylation (e.g., Sox9, Liu *et al.*, 2013; for a review on Sox9 in chondrogenesis, see Lefebvre *et al.*, 2019; for a review on Zeb2, Birkhoff *et al.*, 2021).

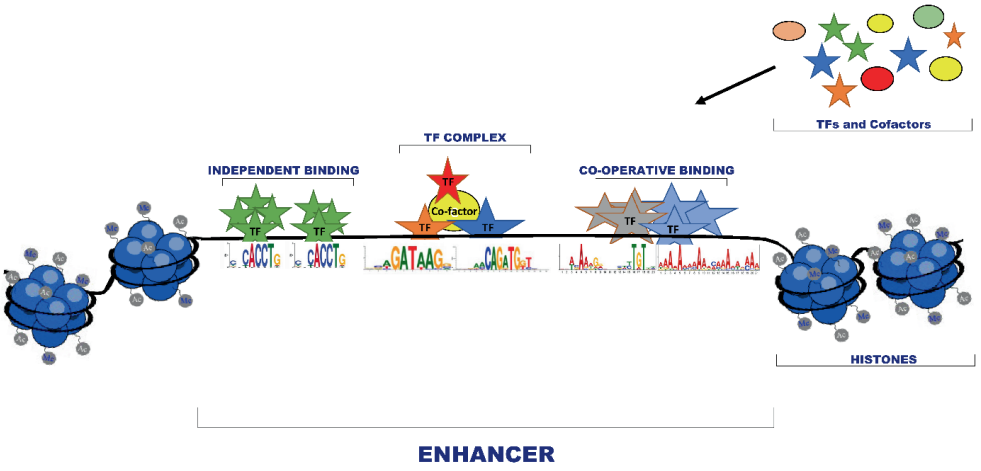


Figure 1a.5. Transcription factors bind to enhancers.

Transcription factors (TFs) bind to the TF binding motifs defined by the position weight matrices (PWMs). TFs mostly bind to the 'open state' chromatin regions within the core of enhancer regions. TFs can bind either independently from other proteins or do so in co-operation with TFs binding in close proximity. TFs can also bind to DNA as a part of a TF complex. In this case, not all TFs or co-factors have to bind directly to DNA. They can also be involved on other aspects, e.g., in stabilization of complex structures.

Additional explanation of such dual-function can be the competitive binding of two TFs to one (identical or overlapping) binding site on DNA. However, overlapping expression of TF mRNAs encoding TFs with identical DNA-binding motif not always leads to competitive displacement of one of the TFs as intuitively often assumed: in fact, it can even result in increased binding of each TF, for example achieved by the increased chromatin accessibility gained by elevated occupancy by the TFs at the binding site (Voss *et al.,* 2011). TF concentration also plays a role in the regulation of TF binding and resulting gene expression. For instance, increasing steady-state levels of the Rel family TF nuclear factor-κB (NF-κB) gradualy change transcription level of its target gene(s) (Giorgetti *et al.,* 2010). In contrast to non-cooperative binding (e.g., of NF-κB), the relationship between concentration and cooperative binding of TFs is not always linear. Rather, it may have a binary effect, so is leading to activation/inactivation of *cis*-REs (reviewed in Courey, 2001; Lebrecht *et al.,* 2005).

TFs are often presented as "master regulators" of development, playing a pivotal role in cell differentiation processes, with myogenic TFs such as *MyoD* being classical historical examples (for reviews, *see* Buckingham, 1992; Buckingham and Rigby, 2014). Zeb2, studied in this PhD research, is one of such factors for neuronal and glial differentiation of cells, for example derived from ESCs (Stryjewska *et al.,* 2017). However, our understanding of TF function in e.g., cell differentiation processes, and certainly in target gene selectivity, is far from complete. Simple correlation between two variables such as expression pattern(s) of TF mRNAs and occurrence of TBFSs at REs (promoters and e.g., enhancers) is not efficient to build effective prediction models of global regulation of GRNs, which furthermore could be applied across different cell lineages (Zeitlinger, 2020). This means that despite the REs of two genes are occupied by the same TF, e.g., by a repressor, it can be that one of the two genes indeed becomes actively repressed, whereas the other is not. Furthermore, investigation of TF-binding has also shown that TFs can occupy distinct subsets of REs or bind with different frequencies specific for individual differentiation stages, or cell transition states during differentiation. For instance, studies of myoblast differentiation in the fly revealed a group of *cis*-REs characterized by continuous binding of the concerned TFs, in contrast to the subset of *cis*-REs which are bound by the same TFs, but restrictively at specific differentiation stage. Importantly, such TF-encoding mRNAs were expressed all the time (Sandmann *et al.,* 2006; Cao *et al.,* 2010). Therefore, this points out that not only temporal expression of TF mRNAs, but also the timing of TF binding is critical to proper regulation of GRNs.

1.4. Higher-order chromatin organization and epigenetic modifications that regulate gene transcription

The estimated length of 2m of human DNA per cell in total is stored in the 6-µm cell nucleus, which requires massive DNA-folding. At the same time, the higher-order 3D chromatin structure is involved in gene expression control. Thus, strict gene expression control is orchestrated at multiple levels, using mechanisms such as dynamic regulation of 3D chromatin organization, DNA-(de)methylation and modification of histones. The crosstalk between the epigenetic modifications and changes in chromatin architecture activate/inactivate *cis*-REs, and together with the TFs and co-factors, control complex gene expression patterns in a correct spatio-temporal manner. In the following sections, major aspects of 3D genome organization and epigenetic modifications will be summarized.

1.4.1. Histone modifications

Histones are a family of small, positively charged proteins that bind tightly to the negatively charged DNA. As such they create a fundamental unit of chromatin, the nucleosome. Each nucleosome consists of DNA (~150bp, length varying between species) wrapped twice around the core histone octamer composed of two copies each of H2A, H2B, H3 and H4. This nucleosome core particle is 1-4 nm in height and 13-22 nm in width (Zhao *et al.,* 1999).

Importantly, PTMs of histone tails influence inter-nucleosomal interactions as well as frequency of contacts between DNA and histones, causing either nucleosomal depletion or concentration. Nucleosomal depletion is associated with "open state" chromatin (euchromatin), whereas regions with concentrated nucleosomes are mostly inaccessible and hence interpreted as "closed state" chromatin (heterochromatin; Schübeler *et al.,* 2004; Ebert *et al.,* 2006). As mentioned (*see* section 1.2.2), one of the histone PTMs linked to active

genomic regions is acetylation (ac) of specific Lysine residues (e.g., H3K27ac or H3K9ac). So, such a histone mark leads to neutralization of the Lysine positive charge, thereby weakening of the interaction of the histone with DNA. It occurs at the sites of active *cis*-REs (e.g., enhancers and promoters; *see* section 1.2.2 above), which is also reflected in reduced nucleosome occupancy of such region (Creyghton *et al.,* 2010; Karmodiya *et al.,* 2012; Herrera-Uribe *et al.,* 2020). In contrast, methylation of Lysine does not affect the histone charge; depending on the position of the Lysine, and also type of methylation (me1, me2 or me3) it is linked to active or repressive regions. H3K4me1 maps to enhancers and promoters, whereas H3K4me3 exclusively marks promoter regions (Creyghton *et al.,* 2010; Bae and Lesch, 2020). H3K9me3 and H4K27me3 occur in inactive heterochromatin and are associated with gene repression (**Fig. 1a.6;** for reviews and other references therein, *see* Nicetto and Zaret, 2019; Guo *et al.,* 2021).

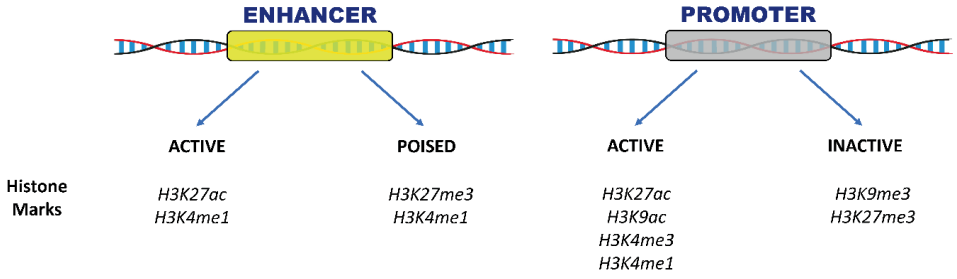


Figure 1a.5. Most commonly studied histone marks at enhancer (left panel) **and promoter regions** (right panel)**.**

1.4.2. Higher-order chromatin organization

Development of molecular methods based on next-generation sequencing (NGS e.g., 3C and Hi-C) that capture 3D chromatin structure (Kempfer and Pombo, 2020), but also technological progress in microscopy (Lakadamyali and Pia Cosma, 2015; Fraser *et al.,* 2015) provided new insight into the spatial organization of the genome and its dynamics. In interphase nuclei, chromosomes are localized in discrete regions, called chromosome territories (CTs) (Szczepinska *et al.,* 2019). At large scale, chromosomes are organized into two compartments, the gene-enriched compartment A and the gene-poor compartment B (Fortin and Hansen, 2015; Lieberman-Aiden *et al.,* 2009; Wang *et al.,* 2016; Beagrie *et al.,* 2017). The A compartment tends to localize at the center of the nucleus and is associated with "open state" chromatin (Fortin and Hansen, 2015; Solovei *et al.,* 2016). In contrast, compartment B occupies more the nuclear periphery, overlapping with lamina-associated domains (LADs) (van Steensel and Belmont, 2018), and contains mostly "closed state" chromatin regions (Chang *et al.,* 2020). Both compartments are associated with cell-specific transcriptional activity, however they vary across different ESC- derived cell types (Dixon *et al.,* 2015). A/B compartments are also rearranged during cell reprogramming and in cell differentiation (Dixon *et al.,* 2015; Krijge *et al.,* 2016).

At a (sub)megabase scale, chromosomes are segregated into TADs, which are characterized by a higher frequency of chromatin contacts within their borders as compared to the inter-TAD interactions (Dixon *et al.,* 2012; Beagrie *et al.,* 2017; Chen *et al.*, 2019).

TAD borders are enriched in motifs of insulator CCCTC-binding factor (CTCF), which are often overlapping with the promoter regions, and active chromatin (Dixon *et al.,* 2012; Rao *et al.,* 2014; Ramírez *et al.,* 2018). CTCF together with cohesin, a versatile multi-functional and multi-subunit protein able to form a ring-shaped DNA-entrapping complex that acts like a mobile DNA clamp, physically interact with each other to form TADs, likely via loop extrusion mechanism (Fudenberg *et al.,* 2016; Hansen *et al.,* 2017; Davidson *et al.,* 2019; Horsfield, 2022; Popay and Dixon, 2022). The formation of chromatin loops, including structural loops establishing TADs, is determined by orientation of insulator binding sites. For instance, a single CTCF motif can be "skipped" by the cohesin ring, in contrast to convergent pairs of CTCF motifs, which stabilize chromatin loop (de Wit *et al.,* 2015). This is also reflected by the number of CTCF binding sites, which substantially outnumbers the chromatin loops identified in the cell (Rao *et al.,* 2014).

Chromatin loops facilitate E-P contacts, which is strongly associated with gene transcription activation (Rao *et al.,* 2014; Bertolini *et al.,* 2019; Giammartino *et al.,* 2019; for a recent discussion of the DNA-looping model, *see* Popay and Dixon, 2022). Consequently, their disruption can lead to gene misexpression, and as a result may alter cell progression and identity (Dowen *et al.,* 2014). Multiple E-P pairs reside in the same TAD, where the E and P are brought into spatial proximity. Contrary, the vast majority of interaction between enhancers and promoters located in neighboring TADs are blocked, which "protects" genes from ectopic activation. Consequently, when the TAD border is disrupted, the enhancer can activate ectopic promoter activity, leading to gene misexpression (Lupiáñez *et al.,* 2015). Therefore, TADs are often considered as stable structures, which control gene expression via much more dynamic E-P intra-TAD looping (Dixon *et al.,* 2012; Lupiáñez *et al.,* 2015; Krefting *et al.,* 2018; McArthur and Capra, 2021).

However, depletion of CTCF (applying e.g., the plant auxin-inducible degron system in nonplant cells; Nishimura *et al.*, 2009; for a review on more ligand-induced degron systems, *see* Kanemaki, 2022; Cummings and Rowley, 2022) has a relatively modest effect on global gene expression, in comparison to the reported strong reorganization of TADs (Nora *et al.,* 2017; Hyle *et al.,* 2019). Other studies, depleting either cohesin or CTCF from interphase chromosomes, have documented different effects on the partitioning into self-associating topological domains in the genome. In the case of cohesin disruption (using an inducible cleavage approach as well as a RNAi- approach targeting its subunit RAD21), the local, i.e. shorter-range interactions within topological domains are generally lost, but the domains themselves remain intact. In the case of CTCF depletion (by RNAi) the intradomain interactions were found reduced, but the interdomain interactions increased (Zuin *et al.*, 2014). Furthermore, the resulting transcriptome also differed between CTCF and cohesin depletion. Thus, CTCF and cohesion were found to act non-redundant in shaping the topological domain organization.

The strong TAD reorganization upon CTCF depletion as documented by Nora *et al.* (2017) and Hyle *et al.* (2019) can be partially explained by micro-C data, which showed that acute (3-hour) depletion of CTCF has modest effect on the majority of E-P interactions. However, approximately 20% of the E-P connections were affected upon CTCF depletion, and this was also associated with gene transcription (Heiseh *et al.,* 2021). This therefore shows that a CTCF-dependent mechanism is crucial for regulation of a significant number of genes. Surprisingly, removal of CTCF does not affect regulation of genes crucial for trans-differentiation of B cells to macrophages (Stik *et al.*, 2020), but CTCF-mediated loops are pivotal for regulation of developmental genes during vertebrate embryogenesis (Franke *et al.,* 2021). Also, application of highly rearranged chromosomes (balancers) in investigation of the role of structural variants

(SVs) in regulation of gene expression showed that disruption of TADs borders often does not drastically alter gene expression (Ghavi-Helm *et al.,* 2019). Altogether, these observations suggest the presence of additional mechanism(s) involved in regulation of E-P contacts, or that activation of a promoter is more dependent on other factors than just spatial E-P proximity (e.g., enhancer specificity), or that ectopic E-P activation is a relatively rare event. Such alternative mechanisms controlling E-P connections are also supported by the fact that during early development such interactions arise before these TADs that are CTCF-cohesin dependent (Espinola *et al.,* 2021).

TADs have been described as largely invariant when compared across different cell types or species, which created a common opinion that they are highly conserved (Dixon *et al.,* 2012; Dixon *et al.*, 2015; Vietri Rudan *et al.,* 2016). However, a recent study showed that TADs are significantly variable across human cell lineages, with even ~20–80% unshared TAD boundaries (Sauerwald *et al.,* 2019). This suggests that TADs are much more dynamic as previously proposed, which is also in line with recent studies using high-resolution microscopy (Hansen *et al.,* 2017; Gabriele *et al.,* 2022). The dynamic nature of TADs is also partialy reflected by the differences in strength of boundary insulation (so, the number of inter-TADs interactions) as well as the frequencies of the intra-TAD interactions, which vary across cell types and change during differentiation (Dixon *et al.*, 2015; Nagano *et al.*, 2017; Gong *et al.*, 2018; Pekowska *et al.,* 2018; Stadhouders *et al.*, 2018; Birkhoff *et al.,* 2020). These distinct patterns of chromatin interaction between cell types, or their reorganization during cell differentiation, are also associated with changes in transcriptional activity (Dixon *et al.,* 2015; Schoenfelder *et al.,* 2015). Importantly, when considering organization of TADs based on Hi-C data, it is important to remember that most of these experiments are at cell population level, presenting a collective view of thousands or millions of cells (Lafontaine *et al.,* 2021). However, microfluidics devices led for example to the development of high-throughput single-cell methods, including single-cell (sc)Hi-C (Nagano *et al.,* 2013). This allowed to readdress the questions regarding TADs structure and its variability. Application of scHi-C revealed that TADs do vary between individual cells (Flyamer *et al.,* 2017), which was further confirmed by microscopy studies showing that TAD-like structures with sharp domain boundaries also differ between cells (Bintu *et al.,* 2018).

Later studies showed substantial differences between mechanisms involved in formation of TADs and A/B compartments. For instance, degradation of CTCF in mouse ESCs leads to disruption of TAD structures, and changes in local insulation of E-P contacts, but does not affect the higher-order organization of A/B compartments (Nora *et al.,* 2017). The A/B compartments are also independent of cohesin absence, in contrast to TADs. Surprisingly, cohesin removal reinforces genome A/B compartmentalization (Schwarzer *et al.,* 2017). Therefore, growing evidence about CTCF/cohesin independent genome organization suggests the presence of additional mechanism(s). The potential mechanism(s) can act independly from loop extrusion or crastalk during regulation of chromatin architecture and gene activity.

Recent insights into phase separation mechanisms (Palikyras and Papantonis, 2019) showed that membraneless, liquid/liquid-phase separated nuclear condensates are also involved in modulation of chromatin structure and transcription. The nuclear condensates are formed by a number of TFs and co-factors promoting gene transcription (Boehning *et al.,* 2018; Boija *et al.,* 2018). The phase-separated nuclear units may also be comprised of Polycomb group (PcG) proteins, or heterochromatin protein-1 (HP1) and be associated with inactive chromatin regions (Larson *et al.,* 2017; Tatavosian *et al.,* 2019; Kent *et al.,* 2020). Interestingly, PRC2 mediates long-range repressive/poising contacts, which cross multiple TADs and regulate gene

silencing (Joshi *et al.,* 2015; Bonev *et al.,* 2017). Collectively, it suggests that nuclear condensates might be involved in formation of A/B compartments, and also transcriptional control.

1.4.3. DNA (de)-methylation

Methylation of the fifth carbon atom of cytosine (5mC) is associated with gene/enhancer silencing. It is mainly, but not exclusively restricted to CpG dinucleotides (hence can sometimes be found at non-CpG sites such as CpA, CpT and CpC). CpG is distributed genome-wide within intragenic/intergenic regions and also found as CpG islands (Nazor *et al.,* 2012; Meissner *et al.,* 2008; Song *et al.*, 2019; Sheaffer *et al.*, 2014, Hattori *et al.,* 2004). Importantly, CPG islands co-localize with the promoters/enhancers of a vast majority of genes, including genes for stem cell identity (Park *et al.,* 2013; Deaton and Bird 2011; Tropel *et al.,* 2017; Shen *et al.,* 2006). Subsequently, genes associated with the differentiation process that are silenced/repressed in stem cells, are often methylated at their promoter regions. In contrast, key pluripotency genes remain unmethylated and thus active during acquisition and maintenance of pluripotency (Fouse *et al.,* 2008). Also, DNA-methylation patterns differ between naïve (hypo-methylated) and primed pluripotent stem cells (hyper-methylated) (Kalkan *et al.*, 2017), and also are reorganized during stem cell reprogramming and differentiation (Lister *et al.,* 2009; Meissner *et al.*, 2008).

Global DNA-methylation is orchestrated by the dynamic interplay between DNA-methyltransferases (DNMTs) and demethylation enzymes, i.e. the ten–eleven translocation (TET) family proteins (Ooi *et al.,* 2007; Ito *et al.,* 2010; Hermann *et al.,* 2004). DNMTs, in particular DNMT3a/3b/3l, are responsible for *de novo* methylation associated with establishment of new DNA methylation sites (Epsztejn-Litman *et al.,* 2008; Ooi *et al.,* 2007; Jia *et al.,* 2007), while the more prominent methylation of hemi-methylated DNA, which allows the transmission of the methylation patterns during DNA replication (Hermann *et al.,* 2004; Probst *et al.,* 2009), is executed by DNMT1. TET enzymes de-methylate DNA via conversion of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Ito *et al.,* 2010). Further, the converted nucleotides are removed from DNA and replaced by cytosine during base excision repair (BER) (Maiti and Drohat 2011). Thus, the steady-state mRNA level of TET enzymes and their activity is a crucial factor in DNA-methylation status. Additionally, the cell-specific methylation pattern may still change by several rounds of replication, leading to progressive dilution of DNA-methylation through cell division not compensated for by the maintenance methylase DNMT1 (Kim and Costello, 2017).

2. The developing limb: an *in vivo* model to investigate *cis*-REs in 3D chromatin

The developing limb remains an attractive model to investigate embryonic patterning and comcommitant cell fate determination. This organ is not only widely used to study principles of embryonic organogenesis and cellular processes and communications therein (such as tissue patterning and axis determination; Newman *et al.*, 2018), but also used as an *in vivo* model to study mechanisms of gene regulation and *cis*-regulatory programs, which upon perturbation can also lead to congenital malformation (Spielmann *et al.,* 2012; VanderMeer *et al.*, 2014; Gherke and Shubin, 2016; Monti *et al.,* 2017; Hörnblad *et al.,* 2021). A fundamental feature of limb formation, initiated as a limb bud in the flank of tetrapod embryos and then characterized by proximal-distal

(P-D) outgrowth, is the dynamic and at the same time highly specific spatio-temporal gene expression, which orchestrates tissue patterning and cell fate in these embryonic extremities (Petit *et al.,* 2017).

2.1. Limb development in brief

At the onset of limb formation undifferentiated mesenchymal cells condensate to form in both flanks of the embryo a structure of densely packed cells surrounded by ectoderm, called the limb bud. These limb buds protrude at thoracic and lumbar levels to form forelimb and hindlimb, respectively. The cells in these buds originate from the lateral plate mesoderm (LPM), which contributes to the formation of skeletal elements and connective tissue of the limb (Zeller *et al.*, 2009). Additionally, the limb bud mesenchyme becomes populated by myogenic cells that migrate from the dermomyotome (DM) part of the segmented somites (Tickle *et al.*, 2015). Limb patterning and growth are controlled by signaling centers with respect to three axes (**Fig. 1a.7**):

- The anterior–posterior (AP) axis (in the human hand, running from thumb to little finger) is controlled by the zone of polarizing activity (ZPA), a region located at the posterior margin of the limb bud, which consists of mesenchymal cells that express Sonic Hedgehog (*Shh*) (Riddle *et al.,* 1993; Hill 2007; for a review on *Shh* in limb development, *see* Tickle and Towers, 2017). The ZPA will be responsible for establishing the right numbers of digits and for their patterning, which was shown by a historical experiment in which ZPA cells were grafted to the ZPA-free, *Shh*-negative anterior region of the limb bud, causing ultimately mirror image digit duplications if sufficient ZAP tissue was transplanted (Saunders, 1948; Tickle, 1981). *Shh* expression is crucial for correct digit patterning and establishment of A-P polarity of the extremities (Chiang *et al.*, 2001). Interestingly, the enhancer regulating *Shh* is also associated with a spectrum of limb malformations (for details, *see* **Chapter 1b**), including dactyly-type of malformation. This is a classic example of *cis*-REs being crucial for proper development, in this case limb formation. Recently, the genome-wide investigation of putative enhancers (PEs) revealed the *cis*-regulatory landscape of active enhancers in ZPA cells. This work identified 1233 PEs, suggesting that this region undergoes complex (epi)genetic regulation via crucial REs that act in limb development and can result in disease (VanderMeer *et al.*, 2014).

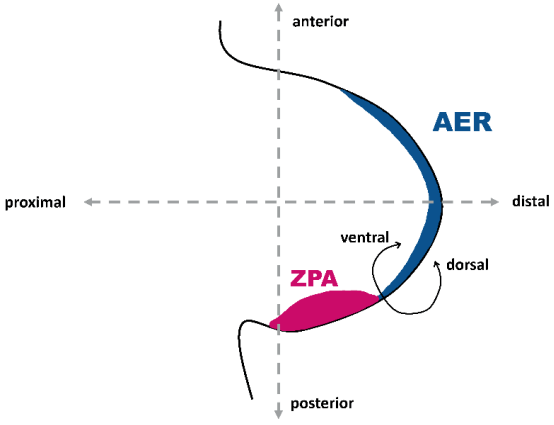


Figure 1a.7. The three axes of the early formed limb bud.

The development of the limb bud is controlled by different signaling centers. The zone of polarizing activity (ZPA), a region of mesenchymal cells expressing *SHH* in the posterior part of each limb bud, controls anterior–posterior patterning, including digit patterning. The apical ectodermal ridge (AER), thickened ectodermal cells expressing *Fgf4* and *Fgf8* at the dorso-ventral edge of the limb bud, orchestrate transiently the proximal-distal outgrowth and patterning of the limb by acting on the mesenchymal cells of the underlying progress zone.

- Outgrowth along the P-D axis (running from the shoulder to ultimately the metacarpal bones and finger tips) is regulated by the apical ectodermal ridge (AER) located at the dorso-ventral (D-V) edge of the limb bub. The AER contains cells expressing the genes encoding Fibroblast Growth Factors-4 and -8 (*Fgf4, Fgf8*) (Boulet *et al.,* 2004). Interestingly, *Fgf8*, which is expressed over the entire length of the AER in the limb (*Fgf4* only in part of it), has a very complex *cis*-regulatory landscape with dozens of enhancers shown to be active in various tissues. Six of these *Fgf8* enhancers are also active in the AER (Marinić *et al.,* 2013; Hörnblad *et al.,* 2021). *Fgf8* is also stimulated by *Ffg10* signaling originating from the LPM derived mesenchymal cells still located more into the flank region of the embryo. Expression of these two *Fgf* genes facilitates positive feedback controlled by the AER and is at the same time responsible for mesenchymal cell proliferation and thus limb growth. Importantly, removal of the AER from the developing limb bud at the appropriate time in development results in subsequent loss of distal skeletal elements (Saunders, 1948; Summerbell 1974). Also, inactivation of *Fgf8* in AER causes reduction in limb bud size, and leads to hypoplasia or aplasia of skeletal elements (Lewandoski *et al.,* 2000). The genome-wide study investigating REs in the AER identified 715 specific PEs active in this region (VanderMeer *et al.*, 2014). It is tempting to suggest that some of these PEs are pivotal for formation of the AER. For example, removal of *Fgf8*-annotated enhancers active in this AER leads to limb defects (Marinić *et al.,* 2013). Interestingly removal of individual enhancers does neither significantly affect the expression of *Fgf8* nor resulted in limb malformation, suggesting that this aspect of the AER presents regulatory redundancy, that individual enhancers have partial contribution to the total expression, or that some of these enhancers can compensate the loss of others (Hörnblad *et al.,* 2021).
- The D-V axis (running in humans from the back of the hand to the palm) is controlled by a genetic network involving *Wnt7a* (expressed in the dorsal ectoderm), LIM homeodomain TFs (*Lmx1b*; expressed in the dorsal mesenchyme), and engrailed-1 (encoded by *En1*), a strong repressor in the (non-AER) ectoderm of *Wnt7a* (Chen and Johnson, 2002). The homozygous knock-out of *Wnt7a* leads to a D-to-V transition of the limb mesenchyme (Parr and McMahon, 1995). Wnt-7a is an inducer of *Lmx1b* in the dorsal mesenchyme. *Lmx1b* is essential for the specification of dorsal limb fate, as shown by homozygous knock-out, again leading to D-to-V transition of the limb (Chen *et al.*, 1998; Chen and Johnson, 2002). In contrast, disruption of a single allele of *LMX1B* in humans leads to Nail-Patella Syndrome characterized by dysplasia of nails, which are located at the dorsal part of the limb digit (McIntosh *et al.,* 1998). Hence, reduction of *Lmx1b* causes incomplete limb dorsalization. Recently, two enhancers (LARM1 and LARM2) involved in regulation of dorsal-identity limb patterning were found in the *cis*-regulatory landscape of *Lmx1b*. Both of them are active in developing limb and indeed overlap with the expression of *Lmx1b* when tested (Haro *et al.,* 2021). Importantly, removal of LARM2 has been associated with Nail-Patella syndrome again, in cases where no mutation/s in the protein-coding region of *Lmx1b* have been mapped (Haro *et al.*, 2021). This highlights the importance of enhancers in limb (D-V axis) development and also congenital abnormalities.

In contrast to *Wnt7a*, *En-1*'s genetic inactivation results in V-to-D limb identity (Loomis *et al.*, 1996).

The limb signaling centers mutually support each other's activity, ultimately controlling limb outgrowth and patterning in the three dimensions (Delgado *et al.*, 2017). Importantly, improper regulation of one of the signaling centers affects other signaling centers leading to crosstalk disbalance and limb abnormalities. For instance, inactivation of *Fgf8* in the AER leads to delay of *Shh* expression in the ZPA (Lewandoski *et al.,* 2000), whereas knock-out of *Wnt7a* results also in defects of posterior skeletal elements (Parr *et al.*, 1995).

The signaling pathways involved in establishment and maintanence of these three axes are highly conserved between hindlimb and forelimb and also among most tetrapod animals, yet limbs often present different shapes due to adaptation to specific types of locomotion (e.g., bats). This morphological diversity is exemplified by the differences in shapes and length of skeletal elements and also by the number of synovial joints connecting the long bones. For instance, each digit of the chicken hindlimb contains different number of phalanges (digit 1 has 2 phalanges; digit 2 has 3 of them; digit 3 has 4, and digit 4 has 5 of them). In contrast, the human leg contains four digits with three phalanges, while the big toe has two.

Limb skeletal elements originate from undifferentiated mesenchymal cells which undergo condensation and further differentiate into two distinct populations. The former contains chondrocytes involved in formation of long bones in the process called endochondral ossification. The latter concentrates at the prospective site of future synovial joints and contains progenitors of joint structures.

2.2. Endochondral ossification and joint formation

Endochondral ossification starts with condensation of limb mesenchymal cells and their differentiation towards chondrocytes. These subsequently form a cartilage template that will determine the shape of the future bones. This process involves expression of a member of Sry family of TFs (*Sox9*), which starts during mesenchymal cell condensation and persists through chondrogenic differentiation (Akiyama *et al.*, 2002). Importantly, *Sox9* binds and activates the chondrocyte-specific enhancers of collagen-αII (*Col2a1*) and collagen-αXI (*Col11a2*), genes encoding chondrocyte-specific expreacellular matrix proteins (Bi *et al.,* 1999). Mutation in one of the *SOX9* alleles in humans causes Campomelic Dysplasia (CD), a rare neonatal chondrodysplasia (Wagner *et al.,* 1994). A similar phenotype was also observed resulting from heterozygous chromosome rearrangements mapping at least 50 kb upstream of human *SOX9* in the region normally containing its enhancers, with phenotype-causal consequences, as shown in a transgenic *SOX9*-LacZ mouse model (Wunderle *et al.,* 1998).

Further steps of endochondral ossification require a series of sequential changes leading to conversion of proliferating chondrocytes expressing *Col2a1* and Matrilin1 (*Matn1*) to pre-hypertrophic chondrocytes, which express RUNX Family Transcription Factor 2 (*Runx2*) and *Col2a1*. Finally, pre-hypertrophic chondrocytes differentiate to hypertrophic chondrocytes expressing *Runx2* and collagen-αX (*Col10a1*) (Otto *et al.,* 1997; Takeda *et al.,* 2001; Hyde *et al.,* 2007; Ding *et al.,* 2012), enhancers of which have also been studied (Lefebvre *et al.,* 1997; Gu *et al.,* 2014; Ghayor *et al.,* 2000). Differentiation of proliferating chondrocytes to hypertrophic chondrocytes is also accompanied by changes in cell morphology, inhibition of cell proliferation, and production a mineralized extracellular cartilage matrix

(Wuelling and Vortkamp, 2011). Later, hypertrophic cartilage is gradually replaced by mineralized bone, which is associated with co-invasion of osteoprogenitor cells involved in establishment of the primary ossification center. The formation of mineralized bone is also alos associated with invasion of blood vessels, endothelial cells and hematopoietic cells from the perichondrium (layer of fibroblast-like cells surrounding the cartilage elements) into hypertropic cartilage (Berendsen and Olsen, 2015).

At the onset of synovial joint formation the mesenchymal cells start to migrate to the prospective site of future joint, called interzone (**Fig. 1a.8;** Ray *et al.*, 2015; Shwartz *et al.,* 2015). This migration increases cell density within the interzone region, and leads to changes in the cell morphology (from rounded to a more flat shape). How the location of the developing joint between future skeletal elements is determined is not yet clear. However, this process is suggested to require spatio-temporal co-expression of genes involved in establishment of interzone cell identity in the first place. Induction of interzone is associated with increased expression of Growth and differentiation factor-5 (*Gdf5*), Autotaxin (*Enpp2*), Wnt Family Member 9A (*Wnt9a,* also known as *Wnt14*) and ETS Transcription Factor ERG (*Erg*) (**Fig. 1a.8;** Hartmann and Tabin, 2001; Guo *et al.*, 2004; Ray *et al.,* 2015; Howard *et al.,* 2016; Bian *et al.,* 2020). Importantly, lineage tracing studies in prenatal and postnatal stages, using a *Gfd5*-Cre controlled Rosa26-LacZ reporter mouse, showed that *Gdf5*+ cells are progenitors of synovial joints (Koyama *et al.,* 2008). In adulthood, the cells derived from the *Gdf5*+ developing joint interzone cells persist as mesenchymal stem cell population located in synovium in the mouse, and are involved in articular cartilage regeneration (Roelofs *et al.,* 2017). During joint interzone formation, the upregulation of interzone markers is accompanied with Wnt9a-induced reduction of *Sox9*, *Col2a1* expression levels, which leads to suppression of the chondrogenic potential in interzone cells (Soeda *et al.,* 2010). It has been suggested that the proliferating chondrocytes adjacent to interzone region may contribute to development of the synovial joints, but their contribution to formation of interzone is not yet clear, and may rely on the precise restricted exposure to BMP vs. exposure to Wnt signaling activity (Ray *et al.*, 2015).

Crosstalk between Wnt and BMP signaling plays an essential role during formation of both the joint interzone and its adjacent cartilage template (Chijimatsu and Saito, 2019). Ablation of Noggin, encoding a secreted BMP inhibitor and that is produced at the border region located between interzone and cartilage anlagen, leads to ectopic activation of BMP signaling in the prospective site of the future joint, and also results in fusion of the distal ends of the future skeletal elements (Brunet et al., 1998; Tylzanowski et al., 2006; Ray et al., 2015). BMPs themselves inhibit Wnt signaling in the interzone (Ray et al., 2015). This suggests that BMP signaling is crucial for chondrocyte differentiation, and thus may control cell identity of mesenchymal cells. Indeed, functional inhibition of BMP activity in limb bud mesenchyme (and at the same time somatic mesoderm, for a replication-competent retrovirus was used in chick embryos) by ectopic production of Noggin, leads to complete inhibition of chondrogenesis (Capdevila and Johnson, 1998). Surprisingly, BMP genes such as Gdf5 and Bmp2 are expressed in interzone (Storm et al., 1994; Macias et al., 1997). GDF5-null mutations in humans (albeit not in such mutant mice) or genetic inactivation of Bmp2 in mice leads to joint defects (Storm and Kingsley, 1996). In contrast, overproduction of either Bmp2 (and Bmp4) or Gdf5 in developing limb affects joint development by leading to joint fusion (limbs of chick embryos, Duprez et al., 1996; Gdf5 transgenic mice, Tsumaki et al., 1999).

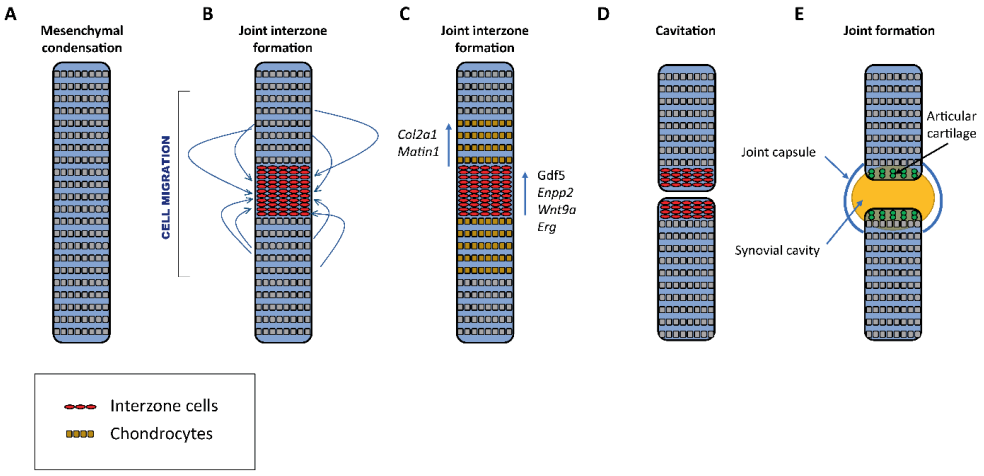


Figure 1a.8. Formation of the joint interzone and development of the synovial joint.

A. Condensation of limb mesenchymal cells to form the template that will determine the shape of the future bones precedes development of synovial joints. **B.** Formation of synovial joints starts with migration of the mesenchymal cells to the prospective site of the future joint, called interzone. **C.** The interzone cells are characterized by their more flattened shape compared with chondrocytes, and by expression of marker genes (e.g., *Gdf5, Enpp2, Wnt9a, Erg*). Adjacent chondrocytes express high levels of *Col2a1* and *Matn1*. **D.** Subsequent cavitation at the center of the joint interzone is leading to formation of the **E** joint capsule and synovial joint structures .

Wnt ligands (including Wnt4, Wnt9a (also known as Wnt14), Wnt16) may play a crucial role in induction of joint interzone, with Wnt9a(Wnt14)-βcatenin being demonstrated to be sufficient and necessary for synovial joint formation (Guo *et al.*, 2004; Hartmann and Tabin, 2001). Virus-based ectopic expression of *Wnt9a* in the prechondrogenic region of future cartilage anlagen leads to activation of β-catenin and repression of *Sox9*, effectively arresting the chondrogenic program (Hartmann and Tabin, 2001). This is correlated also change of cell shape form cubical to more flattened, and is accompanied by expression of *Gdf5*, characteristic of interzone cells (Hartmann and Tabin, 2001). Also, other loss-of-function experiments showed that Wnt signaling plays an important role in the arrest of the chondrogenic program in the interzone region, and is essential for the maintenance of interzone cell fate. However, and remarkably, β-catenin activity is not crucial for the induction of *Gdf5* in early joint interzone, but is required for joint integrity and regulation of *Ihh* during chondrogenesis, including at E12.5-13.5 (Später *et al.,* 2006). In contrast, genetic removal of *β-catenin* in *Col2a1*+ chondrocytes was also associated with joint fusion (Guo *et al.*, 2004). However, this could also be a secondary effect for the cartilage template as it may simply grow much larger and eventually fuse. Altogether, these experiments and results highlight the complexity of the precise regulation of both Wnt and BMP pathways in interzone, and also suggest that spatio-temporal regulation of gene expression allows to control establishment of interzone/chondrocyte cell identity.

Recently, new data regarding joint induction and its interaction with adjacent cartilage was decribed, including identification of *Lgr5* (also marking the onset of lineage divergence in joint formation) and *Lgr5/Col22a1* double+ committed articular chondrocyte progenitors (Feng *et al.*, 2019; for a review, *see* Marín-Llera *et al.*, 2019). However, the field needs also

increased insight in gene expression control in the joint interzone. Multiple publications started to report transcriptomes of the synovial joint, including at single-cell level (Jenner *et al.*, 2014; Pazin *et al.*, 2014; Feng *et al.*, 2019; Bian *et al.*, 2020). Future studies would ideally include also identification and functional characterization of promoter-proximal and distal enhancers controlling key genes in cell differentiation in articular and joint chondrocytes, similar to what is done in other fields, including neurodevelopment (Whyte *et al.*, 2012; Lee *et al.*, 2013; Zhou *et al.*, 2014; Choukrallah *et al.*, 2015; Cruz-Molina *et al.*, 2017; Chang *et al.*, 2019; Birkhoff *et al.*, 2020; Cai *et al.*, 2020).

3. Neural differentiation of stem cells and neuroprogenitors as an *in vitro* model to study the regulation of gene transcription

3.1. Stem cells in the context of embryogenesis and adult tissue repair

The fertilization of a mammalian oocyte by one spermatocyte results in the formation of a diploid zygote, from which embryonic development proceeds. The zygote goes through a series of mitotic cell divisions (cleavages) and forms the morula, which in mammals undergoes compaction (cells adhere more tightly than before) and cavitation. These cells have an equal developmental potential and can differentiate into the later embryonic cells as well as cells of extra-embryonic structures (for a recent discussion of the regulative nature and remarkable plasticity of early mammalian development, *see* Nichols *et al.*, 2022). In morula, these cells can be divided into internally and externally located cells, respectively (Barlow *et al.*, 1972).

Most of these external cells after the subsequent blastocyst stage will become trophectoderm (TE) cells, which are involved in development of the placenta, whereas the internal derived cells will participate in formation of the blastocyst's inner cell mass (ICM; Dyce *et al.*, 1987). This cell lineage segregation is co-controlled by Hippo signaling and the downstream TFs *Yap* and *Taz*, which are involved in the regulation of *Nanog* and *Sox2* pluripotency genes, encoding key factors in ICM development and subsequent embryogenesis. *Yap* and *Taz* also control expression of *Cdx2* and *Gata3*, which are crucial for TE formation (Zhu and Zernicka-Goetz, 2020).

Further, the second wave of cell fate decision takes place in the ICM, where cells form the hypoblast (or primitive endoderm) or the epiblast, composed of pluripotent stem cells (PSCs), which can differentiate into all cell types of the embryo. ICM cells represent the *in vivo* cells from which embryonic stem cells (ESCs) have been established, which can be grown in cell culture, whilst preserving their ability not only to expand and self-renew, but also maintaining their capacity to participate in the formation of embryonic lineages (Evans and Kaufman, 1981; Martin, 1981; Tam and Rossant, 2003; De Los Angeles *et al.*, 2015).

Historical research with pluripotent ESCs, and their instructed differentiation in cell culture and knowledge of the underlying signaling and executing TFs and GRNs, has at the same time significantly contributed to the stem cell research field as a whole. Such work nicely complemented emerging knowledge of multipotent hematopoietic stem cell/progenitor differentiation, giving rise to multiple specialized types of blood cell (Seita and Weissman, 2010). Many (but not all) stem cells that reside in different adult organs or tissues are also at least

bi- or multipotent (with the exception of stem cells in the epidermis, which are unipotent and can only form novel keratinocytes) and do support regenerative processes. Recent examples are adipose tissue-derived mesenchymal stem cells (ASCs), which can (like skin cells) be instructed to convert to cells of the neural lineage, one of potentially many examples of the promises of application of stem cell-based therapy in regenerative medicine (Prpar Mihevc *et al.*, 2020).

3.2. Transition from pluripotent stem cell state(s) prior to exit from pluripotency and cell differentiation

Formation of the blastocyst with its typical blastocoel cavity and its ICM differentiating into hypoblast and epiblast precedes the embryonic implantation of the mammalian embryo into the uterine wall (Zhu and Zernicka-Goetz, 2020). The IMC differentaitied epiblast consists of naive pluripotent cells and expreses *Nanog*, whereas hypoblast (primitive endoderm) is marked by high expression of *Gata6* (Schrode *et al.*, 2014). Nanog plays also an important role in self-renewal of ESCs, which is also controlled by (active) phospho-Stat3 (Stuart *et al.*, 2014). Shortly after implantation, the blastocyst rapidly changes its shape (in the mouse it will become an elongated, cylindric embryo) and size. The naive pluripotent cells of the epiblast undergo morphological, transcriptomic and epigenetic changes and further progress into the well-characterized and well-defined primed pluripotent state (Nichols and Smith, 2009), via an epiblast rosette stage.

The naive rosette and primed pluripotency states identified in embryos could also be "captured" during cell culture of ESCs, as well as the rosette-like state, the "formative" state and the epiblast-like stem cell (EpiLC-like) state (Nichols and Smith, 2009; Neagu *et al.*, 2020; Lackner *et al.*, 2021). However, cell heterogeneity within ICM cells has been demonstrated using scRNA-seq (Kolodziejczyk *et al.*, 2015). Core pluripotency TFs are for example Oct4, Sox2 and Nanog, but additional factors include Esrrb, Klf4 and Klf2, Tbx3 and Rex1 (Nichols and Smith, 2009; Hayashi *et al.*, 2011). Furthermore, the transition between naive and primed pluripotent state, and intermediate states captured in cell culture, is also controlled at the epigenetic level. For instance, following the exit from naive pluripotency, the cells start losing their genome-wide hypomethylation profile, which is associated with changes in gene expression (von Meyenn *et al.*, 2016; Gökbuget and Blelloch, 2019; Bell *et al.*, 2020). Also, the 3D chromatin architecture is reorganized during transition between the naive and primed state, leading to enrichment in chromatin looping as embryogenesis proceeds (Pękowska *et al.*, 2018).

At the exit from primed pluripotency, which co-incides with the onset of differentiation, gene expression profiles change rapidly, and this is mainly orchestrated by TFs. Zeb2 is one of the TFs that controls the exit from pluripotency state and general as well as neural differentiation. For example, in mouse ESCs Zeb2 binds to the *Nanog* promoter and downregulates *Nanog*, including via *Zeb2* upregulation (and *Zeb2* autoregulation is critical for this, *see* **Chapter 7**) (Stryjewska *et al.*, 2017; for human ESCs, *see* Chng *et al.*, 2010). Characterization of the homozygous *Zeb2*-KO mouse ESCs subjected to neural differentiation showed that such cells are impaired in neural and also general differentiation, and that the vast majority of them are characterized by a "blocked" differentiation program at the early epiblast-like state (i.e. the KO cells reach the EpiLC-like state). Lack of ability to exit form this pluripotency state was further linked to alteration in the DNA-methylation profile and changes in gene expression associated with developmental progression (Stryjewska *et al.*, 2017).

3.3. Stem cells and neural-neuronal differentiation in cell culture

The original description of protocols for mouse ICM-derived ESC cultures more than 40 years ago (Martin, 1981) provided a solid ground for cell culture based differentiation protocols towards several cell types, including neural progenitor cells (NPCs), and subsequent neurons (Doetschman *et al.*, 1985; Bibel *et al.*, 2001; Schuldiner *et al.*, 2001; Singec *et al.*, 2016; Kishimoto *et al.*, 2020). For instance, Bain and co-workers developed a retinoic acid (RA) based neural differentiation protocol, with the obtained cells expressing neural-associated genes such as *Tubulin β-III* and TF *Brn3* (Bain *et al.*, 1995). Another group described the differentiation protocol of ESCs into Nestin+ neural progenitor cells (NPCs), which can give rise to both neurons and glial cells, with most mouse ESCs yielding most easily GABAergic-type neuronal cells (e.g., Okabe *et al.*, 1996; Chatzi *et al.*, 2009; Maroof *et al.*, 2010). Further, establishment of culture condition for human stem cells (hESCs; Thomson *et al.*, 1998), led to the development of differentiation protocol of hESCs towards NPCs (Zhang *et al.*, 2001), which have the ability to differentiate into neurons, astrocytes and oligodendrocytes.

Reprogramming of somatic cells into PSCs (Takahashi and Yamanaka, 2006) using specific culture condition and, above all an appropriate cocktail of transduced heterologous TFs, was a great breakthrough in the field of stem cells research. These induced PSC (iPSCs) have the ability to differentiate into neural cells, including neurons (Hu *et al.*, 2010). This also allows to use somatic cells derived from patients with various congenital neurodevelopmental defects, and investigate the role of causal genes - if known - that are expressed during early embryogenesis and neurogenesis, and subsequent brain development. This approach led to the more detailed characterization of function(s) of multiple genes and allowed to better understand the molecular mechanisms involved in the etiology of many such rare disorders (reviewed in Sabitha *et al.*, 2021). Furthermore, human iPSCs have been broadly used to investigate the role of gene regulation in human neuronal developments (Xu *et al.*, 2009; Zhao *et al.*, 2017; di Giammartino *et al.*, 2019; Birkhoff *et al.*, 2020).

Literature references

Abuhashem A, Garg V, Hadjantonakis AK. RNA polymerase II pausing in development: orchestrating transcription. Open Biol. 2022 Jan;12(1):210220.

Akiyama H, Chaboissier MC, Martin JF, 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. 2002 Nov 1;16(21):2813-28.

Ali T, Renkawitz R, Bartkuhn M. Insulators and domains of gene expression. Curr Opin Genet Dev. 2016 Apr;37:17-26.

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013 Mar 1;339(6123):1074-7.

Bacon CW, D'Orso I. CDK9: a signaling hub for transcriptional control. Transcription. 2019 Apr;10(2):57-75.

Bae S, Lesch BJ. H3K4me1 Distribution Predicts Transcription State and Poising at Promoters. Front Cell Dev Biol. 2020 May 5;8:289.

Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. Embryonic stem cells express neuronal properties *in vitro*. Dev Biol. 1995 Apr;168(2):342-57.

Bandara TAMK, Otsuka K, Matsubara S, Shiraishi A, Satake H, Kimura AP. A dual enhancer-silencer element, DES-K16, in mouse spermatocyte-derived GC-2spd(ts) cells. Biochem Biophys Res Commun. 2021 Jan 1;534:1007-12.

Barabino SM, Hübner W, Jenny A, Minvielle-Sebastia L, Keller W. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev. 1997 Jul 1;11(13):1703-16.

Barlow P, Owen DA, Graham C. DNA synthesis in the preimplantation mouse embryo. J Embryol Exp Morphol. 1972 Apr;27(2):431-45.

Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, Fraser J, Dostie J, Game L, Dillon N, Edwards PA, Nicodemi M, Pombo A. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 2017 Mar 23;543(7646):519-24.

Bell E, Curry EW, Megchelenbrink W, Jouneau L, Brochard V, Tomaz RA, Mau KHT, Atlasi Y, de Souza RA, Marks H, Stunnenberg HG, Jouneau A, Azuara V. Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. Nat Commun. 2020 Feb 28;11(1):1112.

Berendsen AD, Olsen BR. Bone development. Bone. 2015 Nov;80:14-8.

Bertolini JA, Favaro R, Zhu Y, Pagin M, Ngan CY, Wong CH, Tjong H, Vermunt MW, Martynoga B, Barone C, Mariani J, Cardozo MJ, Tabanera N, Zambelli F, Mercurio S, Ottolenghi S, Robson P, Creyghton MP, Bovolenta P, Pavesi G, Guillemot F, Nicolis SK, Wei CL. Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. Cell Stem Cell. 2019 Mar 7;24(3):462-476.e6.

Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B. Sox9 is required for cartilage formation. Nat Genet. 1999 May;22(1):85-9.

Bian Q, Cheng YH, Wilson JP, Su EY, Kim DW, Wang H, Yoo S, Blackshaw S, Cahan P. A single cell transcriptional atlas of early synovial joint development. Development. 2020 Jul 20;147(14):dev185777.

Bibel M, Richter J, Lacroix E, Barde YA. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc. 2007;2(5):1034-43.

Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018 Oct 26;362(6413):eaau1783.

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, Nowosad K, van den Hout MCGN, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. Hum Mol Genet. 2020 Aug 29;29(15):2535-50.

Birkhoff JC, Huylebroeck D, Conidi A. ZEB2, the Mowat-Wilson Syndrome Transcription Factor: Confirmations, Novel Functions, and Continuing Surprises. Genes (Basel). 2021 Jul 3;12(7):1037.

Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, Gurrieri F, Everman DB, Schwartz CE, Birk OS, Bejerano G, Lomvardas S, Ahituv N. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012 Jun;22(6):1059-68.

Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, McSwiggen DT, Kokic G, Dailey GM, Cramer P, Darzacq X, Zweckstetter M. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol. 2018 Sep;25(9):833-40.

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, Abraham BJ, Afeyan LK, Guo YE, Rimel JK, Fant CB, Schuijers J, Lee TI, Taatjes DJ, Young RA. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018 Dec 13;175(7):1842-55.e16.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017 Oct 19;171(3):557-72.e24.

Boulet AM, Moon AM, Arenkiel BR, Capecchi MR. The roles of Fgf4 and Fgf8 in limb bud initiation and outgrowth. Dev Biol. 2004 Sep 15;273(2):361-72.

Bowman EA, Kelly WG. RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation. Nucleus. 2014;5(3):224-36.

Brunet LJ, McMahon JA, McMahon AP, Harland RM. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. Science. 1998 May 29;280(5368):1455-7.

Buckingham M. Making muscle in mammals. Trends Genet. 1992 Apr;8(4):144-8.

Buckingham M, Rigby PW. Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev Cell. 2014 Feb 10;28(3):225-38.

Burz DS, Rivera-Pomar R, Jäckle H, Hanes SD. Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo. EMBO J. 1998 Oct 15;17(20):5998-6009.

Buttitta L, Mo R, Hui CC, Fan CM. Interplays of Gli2 and Gli3 and their requirement in mediating Shh-dependent sclerotome induction. Development. 2003 Dec;130(25):6233-43.

Cai W, Huang J, Zhu Q, Li BE, Seruggia D, Zhou P, Nguyen M, Fujiwara Y, Xie H, Yang Z, Hong D, Ren P, Xu J, Pu WT, Yuan GC, Orkin SH. Enhancer dependence of cell-type-specific gene expression increases with developmental age. Proc Natl Acad Sci USA. 2020 Sep 1;117(35):21450-21458.

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, Tapscott SJ. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell. 2010 Apr 20;18(4):662-74.

Capdevila J, Johnson RL. Endogenous and ectopic expression of noggin suggests a conserved mechanism for regulation of BMP function during limb and somite patterning. Dev Biol. 1998 May 15;197(2):205-17.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet. 2006 Jun;38(6):626-35.

Chang L, Li M, Shao S, Li C, Ai S, Xue B, Hou Y, Zhang Y, Li R, Fan X, He A, Li C, Sun Y. Nuclear peripheral chromatin-lamin B1 interaction is required for global integrity of chromatin architecture and dynamics in human cells. Protein Cell. 2022 Apr;13(4):258-280.

Chatzi C, Scott RH, Pu J, Lang B, Nakamoto C, McCaig CD, Shen S. Derivation of homogeneous GABAergic neurons from mouse embryonic stem cells. Exp Neurol. 2009 Jun;217(2):407-16.

Chen H, Johnson RL. Interactions between dorsal-ventral patterning genes lmx1b, engrailed-1 and wnt-7a in the vertebrate limb. Int J Dev Biol. 2002;46(7):937-41.

Chen H, Liang H. A High-Resolution Map of Human Enhancer RNA Loci Characterizes Super-enhancer Activities in Cancer. Cancer Cell. 2020 Nov 9;38(5):701-15.e5.

Chen H, Lun Y, Ovchinnikov D, Kokubo H, Oberg KC, Pepicelli CV, Gan L, Lee B, Johnson RL. Limb and kidney defects in Lmx1b mutant mice suggest an involvement of LMX1B in human nail patella syndrome. Nat Genet. 1998 May;19(1):51-5.

Chen RA, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LW, Jeffers TE, Ahringer J. The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. Genome Res. 2013 Aug;23(8):1339-47.

Chen X, Ke Y, Wu K, Zhao H, Sun Y, Gao L, Liu Z, Zhang J, Tao W, Hou Z, Liu H, Liu J, Chen ZJ. Key role for CTCF in establishing chromatin structure in human embryos. Nature. 2019 Dec;576(7786):306-10.

Chiang C, Litingtung Y, Harris MP, Simandl BK, Li Y, Beachy PA, Fallon JF. Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function. Dev Biol. 2001 Aug 15;236(2):421-35.

Chijimatsu R, Saito T. Mechanisms of synovial joint and articular cartilage development. Cell Mol Life Sci. 2019 Oct;76(20):3939-52.

Chng Z, Teo A, Pedersen RA, Vallier L. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. Cell Stem Cell. 2010 Jan 8;6(1):59-70.

Choi J, Lysakovskaia K, Stik G, Demel C, Söding J, Tian TV, Graf T, Cramer P. Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. Elife. 2021 Mar 26;10:e65381.

Chopra VS, Levine M. Combinatorial patterning mechanisms in the Drosophila embryo. Brief Funct Genomic Proteomic. 2009 Jul;8(4):243-9.

Conidi A, Cazzola S, Beets K, Coddens K, Collart C, Cornelis F, Cox L, Joke D, Dobreva MP, Dries R, Esguerra C, Francis A, Ibrahimi A, Kroes R, Lesage F, Maas E, Moya I, Pereira PN, Stappers E, Stryjewska A, van den Berghe V, Vermeire L, Verstappen G, Seuntjens E, Umans L, Zwijsen A, Huylebroeck D. Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFβ/BMP signaling *in vivo*. Cytokine Growth Factor Rev. 2011 Oct-Dec;22(5-6):287-300.

Choukrallah MA, Song S, Rolink AG, Burger L, Matthias P. Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat Commun. 2015 Oct 19;6:8324.

Core L, Adelman K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. 2019 Aug 1;33(15-16):960-82.

Courey AJ. Cooperativity in transcriptional control. Curr Biol. 2001 Apr 3;11(7):R250-2.

Cramer P. Organization and regulation of gene transcription. Nature. 2019 Sep;573(7772):45-54.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA. 2010 Dec 14;107(50):21931-6.

Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, van Ijcken WFJ, Grosveld F, Frommolt P, Bazzi H, Rada-Iglesias A. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell. 2017 May 4;20(5):689-705.e9.

Cummings CT, Rowley MJ. Implications of Dosage Deficiencies in CTCF and Cohesin on Genome Organization, Gene Expression, and Human Neurodevelopment. Genes (Basel). 2022 Mar 25;13(4):583.

Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. Biochim Biophys Acta. 2015 Aug;1849(8):1116-31.

Dao LTM, Spicuglia S. Transcriptional regulateon by promoters with enhancer function. Transcription. 2018;9(5):307-14.

Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters JM. DNA loop extrusion by human cohesin. Science. 2019 Dec 13;366(6471):1338-45.

De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, Hochedlinger K, Jaenisch R, Lee S, Leitch HG, Lensch MW, Lujan E, Pei D, Rossant J, Wernig M, Park PJ, Daley GQ. Hallmarks of pluripotency. Nature. 2015 Sep 24;525(7570):469-78.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010 May 11;8(5):e1000384.

De Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF Binding Polarity Determines Chromatin Looping. Mol Cell. 2015 Nov 19;60(4):676-84.

Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011 May 15;25(10):1010-22.

Delgado I, Torres M. Coordination of limb development by crosstalk among axial patterning pathways. Dev Biol. 2017 Sep 15;429(2):382-86.

Devkota P, Wuchty S. Promoter/enhancer-based controllability of regulatory networks. Sci Rep. 2022 Mar 3;12(1):3528.

Di Giammartino DC, Kloetgen A, Polyzos A, Liu Y, Kim D, Murphy D, Abuhashem A, Cavaliere P, Aronson B, Shah V, Dephoure N, Stadtfeld M, Tsirigos A, Apostolou E. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. Nat Cell Biol. 2019 Oct;21(10):1179-90.

Ding M, Lu Y, Abbassi S, Li F, Li X, Song Y, Geoffroy V, Im HJ, Zheng Q. Targeting Runx2 expression in hypertrophic chondrocytes impairs endochondral ossification during early skeletal development. J Cell Physiol. 2012 Oct;227(10):3446-56.

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y, Liang J, Zhao H, Lobanenkov VV, Ecker JR, Thomson JA, Ren B. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015 Feb 19;518(7539):331-6.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 Apr 11;485(7398):376-80.

Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. The *in vitro* development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol. 1985 Jun;87:27-45.

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schujiers J, Lee TI, Zhao K, Young RA. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014 Oct 9;159(2):374-87.

Duprez D, Bell EJ, Richardson MK, Archer CW, Wolpert L, Brickell PM, Francis-West PH. Overexpression of BMP-2 and BMP-4 alters the size and shape of developing skeletal elements in the chick limb. Mech Dev. 1996 Jul;57(2):145-57.

Dyce J, George M, Goodall H, Fleming TP. Do trophectoderm and inner cell mass cells in the mouse blastocyst maintain discrete lineages? Development. 1987 Aug;100(4):685-98.

Eaton JD, West S. Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet. 2020 Sep;36(9):664-75.

Egloff S, Murphy S. Cracking the RNA polymerase II CTD code. Trends Genet. 2008;24:280–8.

Epsztejn-Litman S, Feldman N, Abu-Remaileh M, Shufaro Y, Gerson A, Ueda J, Deplus R, Fuks F, Shinkai Y, Cedar H, Bergman Y. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol. 2008 Nov;15(11):1176-83.

Ebert A, Lein S, Schotta G, Reuter G. Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res. 2006;14(4):377-92.

Erceg J, Saunders TE, Girardot C, Devos DP, Hufnagel L, Furlong EE. Subtle changes in motif positioning cause tissue-specific effects on robustness of an enhancer's activity. PLoS Genet. 2014 Jan;10(1):e1004060.

Espinola SM, Götz M, Bellec M, Messina O, Fiche JB, Houbron C, Dejean M, Reim I, Cardozo Gizzi AM, Lagha M, Nollmann M. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. Nat Genet. 2021 Apr;53(4):477-86.

Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981 Jul 9;292(5819):154-6.

Feng C, Chan WCW, Lam Y, Wang X, Chen P, Niu B, Ng VCW, Yeo JC, Stricker S, Cheah KSE, Koch M, Mundlos S, Ng HH, Chan D. Lgr5 and Col22a1 Mark Progenitor Cells in the Lineage toward Juvenile Articular Chondrocytes. Stem Cell Reports. 2019 Oct 8;13(4):713-29.

Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature. 2017 Apr 6;544(7648):110-4.

Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol. 2015 Aug 28;16(1):180.

Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell. 2008 Feb 7;2(2):160-9.

Franke M, De la Calle-Mustienes E, Neto A, Almuedo-Castillo M, Irastorza-Azcarate I, Acemel RD, Tena JJ, Santos-Pereira JM, Gómez-Skarmeta JL. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. Nat Commun. 2021 Sep 13;12(1):5415.

Fraser J, Williamson I, Bickmore WA, Dostie J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol Mol Biol Rev. 2015 Sep;79(3):347-72.

Frith MC; FANTOM consortium. Explaining the correlations among properties of mammalian promoters. Nucleic Acids Res. 2014 Apr;42(8):4823-32.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016 May 31;15(9):2038-49.

Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science. 2018 Sep 28;361(6409):1341-45.

Gabriele M, Brandão HB, Grosse-Holz S, Jha A, Dailey GM, Cattoglio C, Hsieh TS, Mirny L, Zechner C, Hansen AS. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. Science. 2022 Apr 29;376(6592):496-501.

Galouzis CC, Furlong EEM. Regulating specificity in enhancer-promoter communication. Curr Opin Cell Biol. 2022 Apr;75:102065.

Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. 2020 Jan 8;48(D1):D58-D64.

Geertz M, Shore D, Maerkl SJ. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. Proc Natl Acad Sci USA. 2012 Oct 9;109(41):16540-5.

Gehrke AR, Shubin NH. Cis-regulatory programs in the development and evolution of vertebrate paired appendages. Semin Cell Dev Biol. 2016 Sep;57:31-39.

Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet. 2019 Aug;51(8):1272-82.

Ghayor C, Herrouin JF, Chadjichristos C, Ala-Kokko L, Takigawa M, Pujol JP, Galéra P. Regulation of human COL2A1 gene expression in chondrocytes. Identification of C-Krox-responsive elements and modulation by phenotype alteration. J Biol Chem. 2000 Sep 1;275(35):27421-38

Gehrke AR, Shubin NH. Cis-regulatory programs in the development and evolution of vertebrate paired appendages. Semin Cell Dev Biol. 2016 Sep;57:31-9.

Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T, Pasparakis M, Milani P, Bulyk ML, Natoli G. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. Mol Cell. 2010 Feb 12;37(3):418-28.

Gökbuget D, Blelloch R. Epigenetic control of transcriptional regulation in pluripotency and early differentiation. Development. 2019 Sep 25;146(19):dev164772.

Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, Aifantis I, Tsirigos A. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun. 2018 Feb 7;9(1):542.

Grabowicz IE, Wilczyński B, Kamińska B, Roura AJ, Wojtaś B, Dąbrowski MJ. The role of epigenetic modifications, long-range contacts, enhancers and topologically associating domains in the regulation of glioma grade-specific genes. Sci Rep. 2021 Aug 2;11(1):15668

Grossman SR, Engreitz J, Ray JP, Nguyen TH, Hacohen N, Lander ES. Positional specificity of different transcription factor classes within enhancers. Proc Natl Acad Sci USA. 2018 Jul 24;115(30):E7222-E30.

Gu J, Lu Y, Li F, Qiao L, Wang Q, Li N, Borgia JA, Deng Y, Lei G, Zheng Q. Identification and characterization of the novel Col10a1 regulatory mechanism during chondrocyte hypertrophic differentiation. Cell Death Dis. 2014 Oct 16;5(10):e1469.

Guerrero-Martínez JA, Ceballos-Chávez M, Koehler F, Peiró S, Reyes JC. TGFβ promotes widespread enhancer chromatin opening and operates on genomic regulatory domains. Nat Commun. 2020 Dec 3;11(1):6196.

Guo X, Day TF, Jiang X, Garrett-Beal L, Topol L, Yang Y. Wnt/beta-catenin signaling is sufficient and necessary for synovial joint formation. Genes Dev. 2004 Oct 1;18(19):2404-17.

Guo Y, Zhao S, Wang GG. Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 'Readout', and Phase Separation-Based Compaction. Trends Genet. 2021 Jun;37(6):547-65.

Haberle V, Lenhard B. Promoter architectures and developmental gene regulation. Semin Cell Dev Biol. 2016 Sep;57:11-23.

Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jiménez F, Baylies MK, Michelson AM. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell. 2000 Sep 29;103(1):63-74.

Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife. 2017 May 3;6:e25776.

Haro E, Petit F, Pira CU, Spady CD, Lucas-Toca S, Yorozuya LI, Gray AL, Escande F, Jourdain AS, Nguyen A, Fellmann F, Good JM, Francannet C, Manouvrier-Hanu S, Ros MA, Oberg KC. Identification of limb-specific Lmx1b auto-regulatory modules with Nail-patella syndrome pathogenicity. Nat Commun. 2021 Sep 20;12(1):5533.

Haro E, Watson BA, Feenstra JM, Tegeler L, Pira CU, Mohan S, Oberg KC. Lmx1b-targeted *cis*-regulatory modules involved in limb dorsalization. Development. 2017 Jun 1;144(11):2009-20.

Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. Cell. 2001 Feb 9;104(3):341-51.

Hattori N, Nishino K, Ko YG, Hattori N, Ohgane J, Tanaka S, Shiota K. Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. J Biol Chem. 2004 Apr 23;279(17):17063-9.

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell. 2011 Aug 19;146(4):519-32.

Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004 Nov 12;279(46):48350-9.

Herrera-Uribe J, Liu H, Byrne KA, Bond ZF, Loving CL, Tuggle CK. Changes in H3K27ac at Gene Regulatory Regions in Porcine Alveolar Macrophages Following LPS or PolyIC Exposure. Front Genet. 2020 Aug 20;11:817.

Hill RE. How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. Dev Growth Differ. 2007 Aug;49(6):439-48.

Hörnblad A, Bastide S, Langenfeld K, Langa F, Spitz F. Dissection of the Fgf8 regulatory landscape by *in vivo* CRISPR-editing reveals extensive intra- and inter-enhancer redundancy. Nat Commun. 2021 Jan 19;12(1):439.

Horsfield JA. Full circle: a brief history of cohesin and the regulation of gene expression. FEBS J. 2022 Jan 20. doi: 10.1111/febs.16362. Online ahead of print.

Howard M, Tuan RS, Wallis GA. The function and interrelationship between GDF5 and ERG-010 during chondrogenesis *in vitro*. In Vitro Cell Dev Biol Anim. 2016 Feb;52(2):182-92.

Hsieh T-HS, Cattaglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhamcer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesion, WAPL, and YY1. bioRXiv 2021, July 14.

Hu BY, Weick JP, Yu J, Ma LX, Zhang XQ, Thomson JA, Zhang SC. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proc Natl Acad Sci USA. 2010 Mar 2;107(9):4335-40.

Huang D, Ovcharenko I. Enhancer-silencer transitions in the human genome. Genome Res. 2022 Mar;32(3):437-48.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet. 2014 Feb;46(2):205-12.

Hyde G, Dover S, Aszodi A, Wallis GA, Boot-Handford RP. Lineage tracing using matrilin-1 gene expression reveals that articular chondrocytes exist as the joint interzone forms. Dev Biol. 2007 Apr 15;304(2):825-33.

Hyle J, Zhang Y, Wright S, Xu B, Shao Y, Easton J, Tian L, Feng R, Xu P, Li C. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. Nucleic Acids Res. 2019 Jul 26;47(13):6699-713.

Ibragimov AN, Bylino OV, Shidlovskii YV. Molecular Basis of the Function of Transcriptional Enhancers. Cells. 2020 Jul 5;9(7):1620.

Ibrahim MM, Karabacak A, Glahs A, Kolundzic E, Hirsekorn A, Carda A, Tursun B, Zinzen RP, Lacadie SA, Ohler U. Determinants of promoter and enhancer transcription directionality in metazoans. Nat Commun. 2018 Oct 26;9(1):4472.

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010 Aug 26;466(7310):1129-33.

Jenner F, IJpma A, Cleary M, Heijsman D, Narcisi R, van der Spek PJ, Kremer A, van Weeren R, Brama P, van Osch GJ. Differential gene expression of the intermediate and outer interzone layers of developing articular cartilage in murine embryos. Stem Cells Dev. 2014 Aug 15;23(16):1883-98.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007 Sep 13;449(7159):248-51.

Joshi O, Wang SY, Kuznetsova T, Atlasi Y, Peng T, Fabre PJ, Habibi E, Shaik J, Saeed S, Handoko L, Richmond T, Spivakov M, Burgess D, Stunnenberg HG. Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency. Cell Stem Cell. 2015 Dec 3;17(6):748-57.

Kalkan T, Olova N, Roode M, Mulas C, Lee HJ, Nett I, Marks H, Walker R, Stunnenberg HG, Lilley KS, Nichols J, Reik W, Bertone P, Smith A. Tracking the embryonic stem cell transition from ground state pluripotency. Development. 2017 Apr 1;144(7):1221-34.

Kanemaki MT. Ligand-induced degrons for studying nuclear functions. Curr Opin Cell Biol. 2022 Feb;74:29-36.

Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics. 2012 Aug 24;13:424.

Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture. Nat Rev Genet. 2020 Apr;21(4):207-26.

Kent S, Brown K, Yang CH, Alsaihati N, Tian C, Wang H, Ren X. Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. Cell Rep. 2020 Oct 13;33(2):108248.

Kim M, Costello J. DNA methylation: an epigenetic mark of cellular memory. Exp Mol Med. 2017 Apr 28;49(4):e322.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. Nature. 2005 Aug 11;436(7052):876-80.

Kimura H, Sato Y. Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories. Curr Opin Cell Biol. 2022 Feb;74:71-9.

Krijger PH, Di Stefano B, de Wit E, Limone F, van Oevelen C, de Laat W, Graf T. Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. Cell Stem Cell. 2016 May 5;18(5):597-610.

Kishimoto K, Furukawa KT, Luz-Madrigal A, Yamaoka A, Matsuoka C, Habu M, Alev C, Zorn AM, Morimoto M. Bidirectional Wnt signaling between endoderm and mesoderm confers tracheal identity in mouse and human cells. Nat Commun. 2020 Aug 27;11(1):4159.

Kitamura K, Nimura K. Regulation of RNA Splicing: Aberrant Splicing Regulation and Therapeutic Targets in Cancer. Cells. 2021 Apr 16;10(4):923.

Kolodziejczyk AA, Kim JK, Tsang JC, Ilicic T, Henriksson J, Natarajan KN, Tuck AC, Gao X, Bühler M, Liu P, Marioni JC, Teichmann SA. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell. 2015 Oct 1;17(4):471-85.

Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, Imam AMA, van de Werken HJG, Wendt KS, Knoch TA, van IJcken WFJ, Grosveld F. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. Nat Protoc. 2018 Mar;13(3):459-77.

Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D, Brown JM, Gray NE, Collavin L, Gibbons RJ, Flint J, Taylor S, Buckle VJ, Milne TA, Wood WG, Higgs DR. Intragenic enhancers act as alternative promoters. Mol Cell. 2012 Feb 24;45(4):447-58.

Koyama E, Shibukawa Y, Nagayama M, Sugito H, Young B, Yuasa T, Okabe T, Ochiai T, Kamiya N, Rountree RB, Kingsley DM, Iwamoto M, Enomoto-Iwamoto M, Pacifici M. A distinct cohort of progenitor cells participates in synovial joint and articular cartilage formation during mouse limb skeletogenesis. Dev Biol. 2008 Apr 1;316(1):62-73.

Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. BMC Biol. 2018 Aug 7;16(1):87.

Kyburz A, Friedlein A, Langen H, Keller W. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Mol Cell. 2006 Jul 21;23(2):195-205.

Lackner A, Sehlke R, Garmhausen M, Giuseppe Stirparo G, Huth M, Titz-Teixeira F, van der Lelij P, Ramesmayer J, Thomas HF, Ralser M, Santini L, Galimberti E, Sarov M, Stewart AF, Smith A, Beyer A, Leeb M. Cooperative genetic networks drive embryonic stem cell transition from naïve to formative pluripotency. EMBO J. 2021 Apr 15;40(8):e105776.

Lafontaine DL, Yang L, Dekker J, Gibcus JH. Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture. Curr Protoc. 2021 Jul;1(7):e198.

Lakadamyali M, Cosma MP. Advanced microscopy methods for visualizing chromatin structure. FEBS Lett. 2015 Oct 7;589(20 Pt A):3023-30.

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The Human Transcription Factors. Cell. 2018 Feb 8;172(4):650-65.

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature. 2017 Jul 13;547(7662):236-40.

Lebrecht D, Foehr M, Smith E, Lopes FJ, Vanario-Alonso CE, Reinitz J, Burz DS, Hanes SD. Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila. Proc Natl Acad Sci USA. 2005 Sep 13;102(37):13176-81.

Lefebvre V, Angelozzi M, Haseeb A. SOX9 in cartilage development and disease. Curr Opin Cell Biol. 2019 Dec;61:39-47.

Lefebvre V, Huang W, Harley VR, Goodfellow PN, de Crombrugghe B. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol. 1997 Apr;17(4):2336-46.

Lettice LA, Williamson I, Wiltshire JH, Peluso S, Devenney PS, Hill AE, Essafi A, Hagman J, Mort R, Grimes G, DeAngelis CL, Hill RE. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. Dev Cell. 2012 Feb 14;22(2):459-67.

Lewandoski M, Sun X, Martin GR. Fgf8 signaling from the AER is essential for normal limb development. Nat Genet. 2000 Dec;26(4):460-3.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93.

Lin YS, Carey M, Ptashne M, Green MR. How different eukaryotic transcriptional activators can cooperate promiscuously. Nature. 1990 May 24;345(6273):359-61.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009 Nov 19;462(7271):315-22.

Liu JA, Wu MH, Yan CH, Chau BK, So H, Ng A, Chan A, Cheah KS, Briscoe J, Cheung M. Phosphorylation of Sox9 is required for neural crest delamination and is regulated downstream of BMP and canonical Wnt signaling. Proc Natl Acad Sci USA. 2013 Feb 19;110(8):2882-7.

Liu S , Liu Y , Jiang L , Li Z , Lee S , Liu C , Wang J , Zhang J . Recombinant human BMP-2 accelerates the migration of bone marrow mesenchymal stem cells via the CDC42/PAK1/LIMK1 pathway in vitro and in vivo. Biomater Sci. 2018 Dec 18;7(1):362-72.

Liu ZW, Simmons CH, Zhong X. Linking transcriptional silencing with chromatin remodeling, folding, and positioning in the nucleus. Curr Opin Plant Biol. 2022 Jul 13;69:102261.

Loomis CA, Harris E, Michaud J, Wurst W, Hanks M, Joyner AL. The mouse Engrailed-1 gene and ventral limb patterning. Nature. 1996 Jul 25;382(6589):360-3.

Lorenzi L, Chiu HS, Avila Cobos F, Gross S, Volders PJ, Cannoodt R, Nuytens J, Vanderheyden K, Anckaert J, Lefever S, Tay AP, de Bony EJ, Trypsteen W, Gysens F, Vromman M, Goovaerts T, Hansen TB, Kuersten S, Nijs N, Taghon T, Vermaelen K, Bracke KR, Saeys Y, De Meyer T, Deshpande NP, Anande G, Chen TW, Wilkins MR, Unnikrishnan A, De Preter K, Kjems J, Koster J, Schroth GP, Vandesompele J, Sumazin P, Mestdagh P. The RNA Atlas expands the catalog of human non-coding RNAs. Nat Biotechnol. 2021 Nov;39(11):1453-65.

López-Rivera F, Foster Rhoades OK, Vincent BJ, Pym ECG, Bragdon MDJ, Estrada J, DePace AH, Wunderlich Z. A Mutation in the *Drosophila melanogaster eve* Stripe 2 Minimal Enhancer Is Buffered by Flanking Sequences. G3 (Bethesda). 2020 Dec 3;10(12):4473-82.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 21;161(5):1012-25.

Macias D, Gañan Y, Sampath TK, Piedra ME, Ros MA, Hurle JM. Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. Development. 1997 Mar;124(6):1109-17.

Maerkl SJ, Quake SR. A systems approach to measuring the binding energy landscapes of transcription factors. Science. 2007 Jan 12;315(5809):233-7.

Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem. 2011 Oct 14;286(41):35334-8.

Ma J. Crossing the line between activation and repression. Trends Genet. 2005 Jan;21(1):54-9.

Makino S, Zhulyn O, Mo R, Puviindran V, Zhang X, Murata T, Fukumura R, Ishitsuka Y, Kotaki H, Matsumaru D, Ishii S, Hui CC, Gondo Y. T396I mutation of mouse Sufu reduces the stability and activity of Gli3 repressor. PLoS One. 2015 Mar 11;10(3):e0119455.

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. 2013 Mar 11;24(5):530-42.

Marín-Llera JC, Garciadiego-Cázares D, Chimal-Monroy J. Understanding the Cellular and Molecular Mechanisms That Control Early Cell Fate Decisions During Appendicular Skeletogenesis. Front Genet. 2019 Oct 11;10:977.

Maroof AM, Brown K, Shi SH, Studer L, Anderson SA. Prospective isolation of cortical interneuron precursors from mouse embryonic stem cells. J Neurosci. 2010 Mar 31;30(13):4667-75.

Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci USA. 1981 Dec;78(12):7634-8.

Mattioli K, Oliveros W, Gerhardinger C, Andergassen D, Maass PG, Rinn JL, Melé M. Cis and trans effects differentially contribute to the evolution of promoters and enhancers. Genome Biol. 2020 Aug 20;21(1):210.

McArthur E, Capra JA. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am J Hum Genet. 2021 Feb 4;108(2):269-83.

McIntosh I, Dreyer SD, Clough MV, Dunston JA, Eyaid W, Roig CM, Montgomery T, Ala-Mello S, Kaitila I, Winterpacht A, Zabel B, Frydman M, Cole WG, Francomano CA, Lee B. Mutation analysis of LMX1B gene in nail-patella syndrome patients. Am J Hum Genet. 1998 Dec;63(6):1651-8.

Mendoza-Figueroa MS, Tatomer DC, Wilusz JE. The Integrator Complex in Transcription and Development. Trends Biochem Sci. 2020 Nov;45(11):923-34.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008 Aug 7;454(7205):766-70.

Merika M, Williams AJ, Chen G, Collins T, Thanos D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol Cell. 1998 Jan;1(2):277-87.

Merli C, Bergstrom DE, Cygan JA, Blackman RK. Promoter specificity mediates the independent regulation of neighboring genes. Genes Dev. 1996 May 15;10(10):1260-70.

Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. 2018 Jan 1;32(1):42-57.

Misra A, Green MR. From polyadenylation to splicing: Dual role for mRNA 3' end formation factors. RNA Biol. 2016;13(3):259-64.

Monahan K, Horta A, Lomvardas S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature. 2019;565(7740):448-53.

Monti R, Barozzi I, Osterwalder M, Lee E, Kato M, Garvin TH, Plajzer-Frick I, Pickle CS, Akiyama JA, Afzal V, Beerenwinkel N, Dickel DE, Visel A, Pennacchio LA. Limb-Enhancer Genie: An accessible resource of accurate enhancer predictions in the developing limb. PLoS Comput Biol. 2017 Aug 21;13(8):e1005720.

Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, Mendelson Cohen N, Wingett S, Fraser P, Tanay A. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature. 2017 Jul 5;547(7661):61-7.

Nair SJ, Suter T, Wang S, Yang L, Yang F, Rosenfeld MG. Transcriptional enhancers at 40: evolution of a viral DNA element to nuclear architectural structures. Trends Genet. 2022 Jul 7:S0168-9525(22)00141-X.

Narita T, Ito S, Higashijima Y, Chu WK, Neumann K, Walter J, Satpathy S, Liebner T, Hamilton WB, Maskey E, Prus G, Shibata M, Iesmantavicius V, Brickman JM, Anastassiadis K, Koseki H, Choudhary C. Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Mol Cell. 2021 May 20;81(10):2166-82.e6.

Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Müller FJ, Wang YC, Boscolo FS, Fakunle E, Dumevska B, Lee S, Park HS, Olee T, D'Lima DD, Semechkin R, Parast MM, Galat V, Laslett AL, Schmidt U, Keirstead HS, Loring JF, Laurent LC. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell. 2012 May 4;10(5):620-34.

Neagu A, van Genderen E, Escudero I, Verwegen L, Kurek D, Lehmann J, Stel J, Dirks RAM, van Mierlo G, Maas A, Eleveld C, Ge Y, den Dekker AT, Brouwer RWW, van IJcken WFJ, Modic M, Drukker M, Jansen JH, Rivron NC, Baart EB, Marks H, Ten Berge D. *In vitro* capture and characterization of embryonic rosette-stage pluripotency between naive and primed states. Nat Cell Biol. 2020 May;22(5):534-45.

Newman SA, Glimm T, Bhat R. The vertebrate limb: An evolving complex of self-organizing systems. Prog Biophys Mol Biol. 2018 Sep;137:12-24.

Ni T, Corcoran DL, Rach EA, Song S, Spana EP, Gao Y, Ohler U, Zhu J. A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat Methods. 2010 Jul;7(7):521-7.

Nicetto D, Zaret KS. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr Opin Genet Dev. 2019 Apr;55:1-10.

Nichols J, Lima A, Rodríguez TA. Cell competition and the regulative nature of early mammalian development. Cell Stem Cell. 2022 Jul 7;29(7):1018-30.

Nichols J, Smith A. Naive and primed pluripotent states. Cell Stem Cell. 2009 Jun 5;4(6):487-92.

Nicholson AL, Pasquinelli AE. Tales of Detailed Poly(A) Tails. Trends Cell Biol. 2019 Mar;29(3):191-200.

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods. 2009 Dec;6(12):917-22.

Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017 May 18;169(5):930-44.e22.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 Apr 11;485(7398):381-5.

Okabe S, Forsberg-Nilsson K, Spiro AC, Segal M, McKay RD. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells *in vitro*. Mech Dev. 1996 Sep;59(1):89-102.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007 Aug 9;448(7154):714-7.

Onufriev AV, Schiessel H. The nucleosome: from structure to function through physics. Curr Opin Struct Biol. 2019 Jun;56:119-30.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GW, Beddington RS, Mundlos S, Olsen BR, Selby PB, Owen MJ. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell. 1997 May 30;89(5):765-71.

Palikyras S, Papantonis A. Modes of phase separation affecting chromatin regulation. Open Biol. 2019 Oct 31;9(10):190167.

Pang B, Snyder MP. Systematic identification of silencers in human cells. Nat Genet. 2020 Mar;52(3):254-63.

Park Y, Lee JM, Hwang MY, Son GH, Geum D. NonO binds to the CpG island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression. Mol Cells. 2013 Jan;35(1):61-9.

Parr BA, McMahon AP. Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. Nature. 1995 Mar 23;374(6520):350-3.

Pazin DE, Gamer LW, Capelo LP, Cox KA, Rosen V. Gene signature of the embryonic meniscus. J Orthop Res. 2014 Jan;32(1):46-53.

Petit F, Sears KE, Ahituv N. Limb development: a paradigm of gene regulation. Nat Rev Genet. 2017 Apr;18(4):245-58.

Pękowska A, Klaus B, Xiang W, Severino J, Daigle N, Klein FA, Oleś M, Casellas R, Ellenberg J, Steinmetz LM, Bertone P, Huber W. Gain of CTCF-Anchored Chromatin Loops Marks the Exit from Naive Pluripotency. Cell Syst. 2018 Nov 28;7(5):482-95.e10.

Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A. 2011 Aug 16;108(33):13570-5.

Pikaart MJ, Recillas-Targa F, Felsenfeld G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. Genes Dev. 1998 Sep 15;12(18):2852-62.

Popay TM, Dixon JR. Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. J Biol Chem. 2022 Jun 9;298(8):102117. doi: 10.1016/j.jbc.2022.102117. Online ahead of print.

Preker PJ, Ohnacker M, Minvielle-Sebastia L, Keller W. A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. EMBO J. 1997 Aug 1;16(15):4727-37.

Prioleau MN, Nony P, Simpson M, Felsenfeld G. An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. EMBO J. 1999 Jul 15;18(14):4035-48.

Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol. 2009 Mar;10(3):192-206.

Prpar Mihevc S, Kokondoska Grgich V, Kopitar AN, Mohorič L, Majdič G. Neural differentiation of canine mesenchymal stem cells/multipotent mesenchymal stromal cells. BMC Vet Res. 2020 Aug 10;16(1):282.

Qi H, Liu M, Emery DW, Stamatoyannopoulos G. Functional validation of a constitutive autonomous silencer element. PLoS One. 2015 Apr 24;10(4):e0124588.

Quang DX, Erdos MR, Parker SCJ, Collins FS. Motif signatures in stretch enhancers are enriched for disease-associated genetic variants. Epigenetics Chromatin. 2015 Jul 16;8:23.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011 Feb 10;470(7333):279-83.

Ramanathan A, Robb GB, Chan SH. mRNA capping: biological functions and applications. Nucleic Acids Res. 2016 Sep 19;44(16):7511-26.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun. 2018 Jan 15;9(1):189.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665-80.

Ray A, Singh PN, Sohaskey ML, Harland RM, Bandyopadhyay A. Precise spatial restriction of BMP signaling is essential for articular cartilage differentiation. Development. 2015 Mar 15;142(6):1169-79.

Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci USA. 2002 May 14;99(10):6883-8.

Riddle RD, Johnson RL, Laufer E, Tabin C. Sonic hedgehog mediates the polarizing activity of the ZPA. Cell. 1993 Dec 31;75(7):1401-16.

Richter WF, Nayak S, Iwasa J, Taatjes DJ. The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat Rev Mol Cell Biol. 2022 Jun 20:1-18. doi: 10.1038/s41580-022-00498-3. Online ahead of print.

Roelofs AJ, Zupan J, Riemen AHK, Kania K, Ansboro S, White N, Clark SM, De Bari C. Joint morphogenetic cells in the adult mammalian synovium. Nat Commun. 2017 May 16;8:15040.

Sabitha KR, Shetty AK, Upadhya D. Patient-derived iPSC modeling of rare neurodevelopmental disorders: Molecular pathophysiology and prospective therapies. Neurosci Biobehav Rev. 2021 Feb;121:201-19.

Sandmann T, Jensen LJ, Jakobsen JS, Karzynski MM, Eichenlaub MP, Bork P, Furlong EE. A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev Cell. 2006 Jun;10(6):797-807.

Sauerwald N, Singhal A, Kingsford C. Analysis of the structural variability of topologically associated domains as revealed by Hi-C. NAR Genom Bioinform. 2020 Mar;2(1):lqz008.

Saunders JW Jr. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. 1948. J Exp Zool. 1998 Dec 15;282(6):628-68.

Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA. 2006 Jan 31;103(5):1412-7.

Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019 Aug;20(8):437-55.

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, Dimitrova E, Dimond A, Edelman LB, Elderkin S, Tabbada K, Darbo E, Andrews S, Herman B, Higgs A, LeProust E, Osborne CS, Mitchell JA, Luscombe NM, Fraser P. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 2015 Apr;25(4):582-97.

Schrode N, Saiz N, Di Talia S, Hadjantonakis AK. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Dev Cell. 2014 May 27;29(4):454-67.

Schuldiner M, Eiges R, Eden A, Yanuka O, Itskovitz-Eldor J, Goldstein RS, Benvenisty N. Induced neuronal differentiation of human embryonic stem cells. Brain Res. 2001 Sep 21;913(2):201-5.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, Spitz F. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017 Nov 2;551(7678):51-6.

Scott KC, Taubman AD, Geyer PK. Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. Genetics. 1999 Oct;153(2):787-98.

Segert JA, Gisselbrecht SS, Bulyk ML. Transcriptional Silencers: Driving Gene Expression with the Brakes On. Trends Genet. 2021 Jun;37(6):514-27.

Seita J, Weissman IL. Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip Rev Syst Biol Med. 2010 Nov-Dec;2(6):640-53.

Selvamurugan N, Kwok S, Vasilov A, Jefcoat SC, Partridge NC. Effects of BMP-2 and pulsed electromagnetic field (PEMF) on rat primary osteoblastic cell proliferation and gene expression. J Orthop Res. 2007 Sep;25(9):1213-20.

Sheaffer KL, Kim R, Aoki R, Elliott EN, Schug J, Burger L, Schübeler D, Kaestner KH. DNA methylation is required for the control of stem cell differentiation in the small intestine. Genes Dev. 2014 Mar 15;28(6):652-64.

Shen Y, Chow J, Wang Z, Fan G. Abnormal CpG island methylation occurs during *in vitro* differentiation of human embryonic stem cells. Hum Mol Genet. 2006 Sep 1;15(17):2623-35.

Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 2004 Jun 1;18(11):1263-71.

Shwartz Y, Viukov S, Krief S, Zelzer E. Joint Development Involves a Continuous Influx of Gdf5-Positive Cells. Cell Rep. 2016 Jun 21;15(12):2577-87.

Singec I, Crain AM, Hou J, Tobe BTD, Talantova M, Winquist AA, Doctor KS, Choy J, Huang X, La Monaca E, Horn DM, Wolf DA, Lipton SA, Gutierrez GJ, Brill LM, Snyder EY. Quantitative Analysis of Human Pluripotency and Neural Specification by In-Depth (Phospho)Proteomic Profiling. Stem Cell Reports. 2016 Sep 13;7(3):527-42.

Sloutskin A, Shir-Shapira H, Freiman RN, Juven-Gershon T. The Core Promoter Is a Regulatory Hub for Developmental Gene Expression. Front Cell Dev Biol. 2021 Sep 10;9:666508.

Soeda T, Deng JM, de Crombrugghe B, Behringer RR, Nakamura T, Akiyama H. Sox9-expressing precursors are the cellular origin of the cruciate ligament of the knee joint and the limb tendons. Genesis. 2010 Nov;48(11):635-44.

Solovei I, Thanisch K, Feodorova Y. How to rule the nucleus: divide et impera. Curr Opin Cell Biol. 2016 Jun;40:47-59.

Song Y, van den Berg PR, Markoulaki S, Soldner F, Dall'Agnese A, Henninger JE, Drotar J, Rosenau N, Cohen MA, Young RA, Semrau S, Stelzer Y, Jaenisch R. Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. Mol Cell. 2019 Sep 5;75(5):905-20.e6.

Später D, Hill TP, Gruber M, Hartmann C. Role of canonical Wnt-signaling in joint formation. Eur Cell Mater. 2006 Nov 17;12:71-80.

Spielmann M, Brancati F, Krawitz PM, Robinson PN, Ibrahim DM, Franke M, Hecht J, Lohan S, Dathe K, Nardone AM, Ferrari P, Landi A, Wittler L, Timmermann B, Chan D, Mennen U, Klopocki E, Mundlos S. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am J Hum Genet. 2012 Oct 5;91(4):629-35.

Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012 Sep;13(9):613-26.

Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, Gomez A, Collombet S, Berenguer C, Cuartero Y, Hecht J, Filion GJ, Beato M, Marti-Renom MA, Graf T. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet. 2018 Feb;50(2):238-49.

Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science. 1991 Nov 29;254(5036):1385-7.

Stik G, Vidal E, Barrero M, Cuartero S, Vila-Casadesús M, Mendieta-Esteban J, Tian TV, Choi J, Berenguer C, Abad A, Borsari B, le Dily F, Cramer P, Marti-Renom MA, Stadhouders R, Graf T. CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. Nat Genet. 2020 Jul;52(7):655-61.

Storm EE, Huynh TV, Copeland NG, Jenkins NA, Kingsley DM, Lee SJ. Limb alterations in brachypodism mice due to mutations in a new member of the TGF beta-superfamily. Nature. 1994 Apr 14;368(6472):639-43.

Storm EE, Kingsley DM. Joint patterning defects caused by single and double mutations in members of the bone morphogenetic protein (BMP) family. Development. 1996 Dec;122(12):3969-79.

Stryjewska A, Dries R, Pieters T, Verstappen G, Conidi A, Coddens K, Francis A, Umans L, van IJcken WF, Berx G, van Grunsven LA, Grosveld FG, Goossens S, Haigh JJ, Huylebroeck D. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells. 2017 Mar;35(3):611-25.

Stuart HT, van Oosten AL, Radzisheuskaya A, Martello G, Miller A, Dietmann S, Nichols J, Silva JC. NANOG amplifies STAT3 activation and they synergistically induce the naive pluripotent program. Curr Biol. 2014 Feb 3;24(3):340-6.

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

Sun J, Yoon J, Lee M, Lee HK, Hwang YS, Daar IO. Zic5 stabilizes Gli3 via a non-transcriptional mechanism during retinal development. Cell Rep. 2022 Feb 1;38(5):110312.

Sun Y, Hamilton K, Tong L. Recent molecular insights into canonical pre-mRNA 3'-end processing. Transcription. 2020 Apr;11(2):83-96.

Szczepińska T, Rusek AM, Plewczynski D. Intermingling of chromosome territories. Genes Chromosomes Cancer. 2019 Jul;58(7):500-6.

Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. Nat Rev Mol Cell Biol. 2016 Mar;17(3):183-93.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug 25;126(4):663-76.

Takeda S, Bonnamy JP, Owen MJ, Ducy P, Karsenty G. Continuous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa1-deficient mice. Genes Dev. 2001 Feb 15;15(4):467-81.

Tam PP, Rossant J. Mouse embryonic chimeras: tools for studying mammalian development. Development. 2003 Dec;130(25):6155-63.

Tatavosian R, Kent S, Brown K, Yao T, Duc HN, Huynh TN, Zhen CY, Ma B, Wang H, Ren X. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. J Biol Chem. 2019 Feb 1;294(5):1451-63.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. Science. 1998 Nov 6;282(5391):1145-7.

Tickle C. How the embryo makes a limb: determination, polarity and identity. J Anat. 2015 Oct;227(4):418-30.

Tickle C, Towers M. Sonic Hedgehog Signaling in Limb Development. Front Cell Dev Biol. 2017 Feb 28;5:14.

Tropel P, Jung L, André C, Ndandougou A, Viville S. CpG Island Methylation Correlates with the Use of Alternative Promoters for USP44 Gene Expression in Human Pluripotent Stem Cells and Testes. Stem Cells Dev. 2017 Aug 1;26(15):1100-10.

Tsumaki N, Tanaka K, Arikawa-Hirasawa E, Nakase T, Kimura T, Thomas JT, Ochi T, Luyten FP, Yamada Y. Role of CDMP-1 in skeletal morphogenesis: promotion of mesenchymal cell recruitment and chondrocyte differentiation. J Cell Biol. 1999 Jan 11;144(1):161-73.

Tylzanowski P, Mebis L, Luyten FP. The Noggin null mouse phenotype is strain dependent and haploinsufficiency leads to skeletal defects. Dev Dyn. 2006 Jun;235(6):1599-607.

van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell. 2017 May 18;169(5):780-91.

van den Berg DLC, Azzarelli R, Oishi K, Martynoga B, Urbán N, Dekkers DHW, Demmers JA, Guillemot F. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. Neuron. 2017 Jan 18;93(2):348-61.

VanderMeer JE, Smith RP, Jones SL, Ahituv N. Genome-wide identification of signaling center enhancers in the developing limb. Development. 2014 Nov;141(21):4194-8.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, Ballester B, Andrau JC, Spicuglia S. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. Nat Commun. 2015 Apr 15;6:6905.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 2015 Mar 3;10(8):1297-309.

von Meyenn F, Berrens RV, Andrews S, Santos F, Collier AJ, Krueger F, Osorno R, Dean W, Rugg-Gunn PJ, Reik W. Comparative Principles of DNA Methylation Reprogramming during Human and Mouse *In vitro* Primordial Germ Cell Specification. Dev Cell. 2016 Oct 10;39(1):104-15.

Voss TC, Schiltz RL, Sung MH, Yen PM, Stamatoyannopoulos JA, Biddie SC, Johnson TA, Miranda TB, John S, Hager GL. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell. 2011 Aug 19;146(4):544-54.

Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, Scherer G. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell. 1994 Dec 16;79(6):1111-20.

Wang B, Hasan MK, Alvarado E, Yuan H, Wu H, Chen WY. NAMPT overexpression in prostate cancer and its contribution to tumor cell survival and stress response. Oncogene. 2011 Feb 24;30(8):907-21.

Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. Spatial organization of chromatin domains and compartments in single chromosomes. Science. 2016 Aug 5;353(6299):598-602.

Wang Y, Ni T, Wang W, Liu F. Gene transcription in bursting: a unified mode for realizing accuracy and stochasticity. Biol Rev Camb Philos Soc. 2018 Jul 19;94(1):248–58.

West AG, Gaszner M, Felsenfeld G. Insulators: many functions, many mechanisms. Genes Dev. 2002 Feb 1;16(3):271-88.

Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature. 2012 Feb 1;482(7384):221-5.

Williamson I, Lettice LA, Hill RE, Bickmore WA. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. Development. 2016 Aug 15;143(16):2994-3001.

Wuelling M, Vortkamp A. Chondrocyte proliferation and differentiation. Endocr Dev. 2011;21:1-11.

Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A. Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. Proc Natl Acad Sci USA. 1998 Sep 1;95(18):10649-54.

Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. Cell. 2004 May 28;117(5):663-76.

Xie L, Torigoe SE, Xiao J, Mai DH, Li L, Davis FP, Dong P, Marie-Nelly H, Grimm J, Lavis L, Darzacq X, Cattoglio C, Liu Z, Tjian R. A dynamic interplay of enhancer elements regulates *Klf4* expression in naïve pluripotency. Genes Dev. 2017 Sep 1;31(17):1795-1808.

Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. Genes Dev. 2009 Dec 15;23(24):2824-38.

Yao S, Osborne CS, Bharadwaj RR, Pasceri P, Sukonnik T, Pannell D, Recillas-Targa F, West AG, Ellis J. Retrovirus silencer blocking by the cHS4 insulator is CTCF independent. Nucleic Acids Res. 2003 Sep 15;31(18):5317-23.

Yoon Y, Shi Y. Human pre-mRNA 3' end processing: reconstituting is believing. Genes Dev. 2022 Feb 1;36(3-4):106-7.

Zeitlinger J. Seven myths of how transcription factors read the cis-regulatory code. Curr Opin Syst Biol. 2020 Oct;23:22-31.

Zeller R, López-Ríos J, Zuniga A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nat Rev Genet. 2009 Dec;10(12):845-58.

Zhang SC, Wernig M, Duncan ID, Brüstle O, Thomson JA. *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. Nat Biotechnol. 2001 Dec;19(12):1129-33.

Zhang Y, See YX, Tergaonkar V, Fullwood MJ. Long-Distance Repression by Human Silencers: Chromatin Interactions and Phase Separation in Silencers. Cells. 2022 May 5;11(9):1560.

Zhang Z, Shen L, Law K, Zhang Z, Liu X, Hua H, Li S, Huang H, Yue S, Hui CC, Cheng SY. Suppressor of Fused Chaperones Gli Proteins To Generate Transcriptional Responses to Sonic Hedgehog Signaling. Mol Cell Biol. 2017 Jan 19;37(3):e00421-16.

Zhao H, Zhang Y, Zhang SB, Jiang C, He QY, Li MQ, Qian RL. The structure of the nucleosome core particle of chromatin in chicken erythrocytes visualized by using atomic force microscopy. Cell Res. 1999 Dec;9(4):255-60.

Zhao MT, Shao NY, Hu S, Ma N, Srinivasan R, Jahanbani F, Lee J, Zhang SL, Snyder MP, Wu JC. Cell Type-Specific Chromatin Signatures Underline Regulatory DNA Elements in Human Induced Pluripotent Stem Cells and Somatic Cells. Circ Res. 2017 Nov 10;121(11):1237-50.

Zhou Q, Li T, Price DH. RNA polymerase II elongation control. Annu Rev Biochem. 2012;81:119–43.

Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet. 2016 Aug 1;17(9):551-65.

Zhu M, Zernicka-Goetz M. Principles of Self-Organization of the Mammalian Embryo. Cell. 2020 Dec 10;183(6):1467-78.

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci USA. 2014 Jan 21;111(3):996-1001.

Chapter 1
Part b

Mutations in gene regulatory elements linked to human limb malformations

Karol Nowosad [1,2], Ewa Hordyjewska-Kowalczyk [1,2], Przemko Tylzanowski [1,3]

[1] Department of Biochemistry and Molecular Biology, Medical University of Lublin, Lublin, Poland
[2] The Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland
[3] Department of Development and Regeneration, Skeletal Biology and Engineering Research Centre, University of Leuven, Leuven, Belgium

This chapter 1b has been published in
J Med Genet. *2020 Jun;57(6):361-70*

1. Summary

Most of the human genome has a regulatory function in gene expression. The technological progress made in recent years permitted the revision of old and discovery of new mutations outside of the protein-coding regions that do affect human limb morphology. Steadily increasing discovery rate of such mutations suggests that the until now largely neglected part of the genome rises to its well-deserved prominence. In this review, we describe the recent technological advances permitting this unprecedented advance in identifying non-coding mutations. We especially focus on the mutations in *cis*-regulatory elements such as enhancers, and *trans*-regulatory elements such as miRNA and lncRNA, linked to hereditary or inborn limb defects. We also discuss the role of chromatin organization and enhancer-promoter interactions in the etiology of limb malformations.

2. Introduction

Congenital limb malformations are a group of developmental disorders with the estimated prevalence of 1 in 500 of life births (Giele *et al.*, 2001; Ekblom *et al.*, 2010). They may occur in isolation or be associated with additional defects (Jamsheer *et al.*, 2012; Franke *et al.*, 2016). Major risk factors leading to limb abnormalities include exposure to teratogens or genome alterations (Giele *et al.*, 2001; Lupiañez *et al.*, 2015). Studies of congenital limb malformation have allowed to better understand the causative role of genes in limb development. Intriguingly, different mutations within one gene may lead to distinct skeletal phenotypes, whereas mutations in various genes of the same developmental pathway may result in similar or identical limb abnormalities (reviewed in Zelzer and Olsen, 2003). Importantly, an increasing number of mutations identified outside of the protein-coding areas supports the notion that alterations in non-coding regulatory regions may lead to similar skeletal defects as mutations in their target genes. For instance, mutations in enhancers involved in regulation of the short stature homeobox-containing gene *SHOX* result in its downregulation, which lead to similar or identical skeletal phenotype as caused by loss-of-function mutations in (one allele of) *SHOX* (Benito-Sanz *et al.*, 2012). In both cases, the level of SHOX mRNA is significantly reduced, which affects limb development (Bunyan *et al.*, 2016). Several mutations in *SHOX* enhancers have been characterized as molecular factors involved in the etiology of SHOX-related disorders, which is described in details below.

A vertebrate limb emerges from the lateral plate mesoderm and develops under the coordinated action of three signaling centers. The zone of polarizing activity (ZPA), expressing Sonic hedgehog (*SHH*), regulates anterior-posterior axis of limb formation. The apical ectodermal ridge (AER) engages FGF signaling pathways, controlling the proximal-distal (P-D) axis. Finally, the wingless-type MMTV integration site family member 7A (*Wnt7A*) signal from the dorsal ectoderm is responsible for the coordination of the dorsal-ventral (D-V) limb axis (Petit *et al.*, 2017). The crosstalk and coordination among these signaling centers are thus crucial for correct patterning of the limb. During evolution, the differences in the timing and intensity of the expression of genes involved in these signaling pathways are partly responsible for the distinct limb morphology among species (Kvon *et al.*, 2016). Within one species however, even discrete changes to the spatio-temporal regulation of gene expression, caused by mutation

in regulatory elements frequently result in limb malformations. For instance, alterations of ZPA regulatory sequences (ZRSs) involved in regulation of *SHH* result in a range of skeletal abnormalities caused by ectopic expression of *SHH* (Van der Meer *et al.*, 2012; Cho *et al.*, 2013; Lohan *et al.*, 2014). Thus, it becomes apparent that not only the protein-coding regions but also the regulatory apparatus is critical for the limb shape diversity.

In this review, we provide an update on the recent technological developments and the results of studies on the role of mutations in *cis* and *trans*-regulatory elements affecting limb development.

3. Enhancers and enhancer-promoter interactions in regulation of gene expression

3.1. Enhancers and their status

Enhancers are regulatory elements that coordinate the spatio-temporal gene expression by interaction with gene promoters (Schoenfelder and Fraser, 2019). They are located genome-wide within introns, intergenic regions, gene desert areas or even gene exons (Birnbaum *et al.*, 2012). Enhancers can control the expression of target genes at very variable distance, in an orientation-independent manner, and sometimes across chromosomes (Monahan *et al.*, 2019). Therefore, they may act *in cis* (within the same chromosome) or *in trans* (between chromosomes). Enhancers provide a structural scaffold for transcription factors (TFs) and co-factors, bridging them to promoters (Sagai *et al.*, 2005; Fishilevich *et al.*, 2017). Promoters are regulatory elements located upstream from transcription start sites (TSSs) and are involved in the initiation of gene transcription by recruitment of TFs and RNAPol. The function of an enhancer is to augment the level of expression of its target gene (Schoenfelder and Fraser, 2019). Consequently, mutations in enhancers and reorganization of chromatin 3D architecture that affect the enhancer-promoter interaction have been correlated with developmental disorders, including limb malformations (Lohan *et al.*, 2014; Lupiañez *et al.*, 2015).

The development of next-generation sequencing (NGS) techniques has siginificantly contributed to resulted in the identification of many novel enhancers and helped in mapping of their interaction with target promoters. Several strategies have been developed to interrogate the enhancer regulatory regions and enhancer-promoter (E-P) connections. The first one examines DNA-accessibility using two techniques: DNase-I hypersensitive sites sequencing (DNase-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), to investigate the “open state” of chromatin structure, which is correlated with active enhancers and promoters (Buenrostro *et al.*, 2015). The second one is focused on the analysis of interactions between chromatin regions, among others allowing to map the physical E-P connections. It applies the Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) as well as Chromatin Conformation Capture (3C) or its several variants, which include 4C, 5C, T2C and Hi-C (Mishra and Hawkins, 2017). The third strategy interrogates the promoter- and enhancer-specific epigenetic marks using Chromatin Immunoprecipitation followed by sequencing (ChIP-seq). This technique examines histone modifications (e.g., H3K27ac and H3K4me1 specific for enhancers, and H3K4me3 characteristic for promoters), as well as co-factors and TFs interacting with either enhancers regions or gene promoters (e.g., P300 specific for enhancers; CTCF for

insulators etc., *see* below) (Gorkin *et al.*, 2012; Chokeshaiusaha *et al.*, 2018). The abovementioned techniques are frequently used in parallel to better map regulatory elements and E-P connections.

Large-scale epigenetic analysis of human and mouse genomes has identified three different states of enhancers: primed, active or poised. Each of the states is characterized by a specific profile of histone modifications, for instance, primed enhancers are enriched in H3K4me1 (Cruz-Molina *et al.*, 2017). They are able to bind pioneer TFs, proteins that play a role in blocking DNA-methylation, recruiting of chromatin modifiers and repositioning of nucleosomes. Pioneer TFs increase the accessibility of the regions for other TFs (Zaret and Carroll, 2011). Interestingly, primed enhancers are associated with inactive genes "waiting" for a specific trigger prior to transcription (Cruz-Molina *et al.*, 2017). Active enhancers are enriched in H3K27ac and H3K4me1, and also bind pioneer TFs and co-activators. One of the co-activators interacting with active enhancers is histone acetyltransferase P300, an enzyme involved in the regulation of chromatin structure by promoting chromatin decompaction (Gorkin *et al.*, 2012). Analysis of DNAse-I hypersensitive sites showed that they are located within the "open structure" regions of chromatin, suggesting that this state is involved in the regulation of gene expression (Thurman *et al.*, 2012).

The signature of poised enhancers has been characterized by enrichment of H3K4me1 and low level of H3K27ac. It is also marked by the interaction with Polycomb Repressive Complex 2 (PRC2) and the presence of H3K27me3. The poised enhancers have been associated with early development and pluripotent cells, such as human and mouse embryonic stem cells (ESCs) (Cruz-Molina *et al.*, 2017). Nevertheless, the role of enhancers at the poised state is poorly understood. They interact with gene promoters but they are unable to initiate their expression. The inactive state of poised enhancers is lost during the differentiation, and is correlated with loss of H3K27me3 and increase of H3K27ac, suggesting that PRC2 may be involved in specific regulation of this type of enhancers (Cruz-Molina *et al.*, 2017; Koenecke *et al.*, 2017).

Currently, enhancers are considered as transcriptional units producing unique enhancer RNAs (eRNAs) (De Santa *et al.*, 2010). The eRNAs differ from other functional non-coding RNAs, such as lncRNAs or miRNAs. They are typically shorter than 2 kb, highly unstable, not polyadenylated RNAs without additional post-transcriptional modifications (Mikhaylichenko *et al.*, 2018). In addition, eRNAs are typically bidirectionally produced and are characterized by a low-level expression, which can be measured by CAGE (cap analysis of gene expression), or PRO- cap (a precision nuclear run-on sequencing variant). eRNA transcription is a mark of active enhancers and is used to help in identifying functional enhancers (Mikhaylichenko *et al.*, 2018). Despite the unstable nature of eRNAs, they have been shown to be biologically relevant. For example, the recruitment of the RNA-Polymerase to the promoter of Myogenic Differentiation 1 (*MYOD1)* was shown to be facilitated by eRNAs (Mousavi *et al.*, 2013).

3.2. Mutations in enhancers linked to human limb malformations

Brachydactylies (BDs) are one of the examples linking limb malformation to mutations in enhancers. Several types of BDs have been classified from A to E. They are characterized by shortening of selected fingers and/or toes (Temtamy and Aglan, 2008). Most of type-E BDs were correlated with mutations in protein-coding regions. For instance, Brachydactyly Type E (DBE, MIM 113300) was genetically linked to mutations in the open reading frame of the gene encoding Homeobox D13 (*HOXD13*) (Jamsheer *et al.*, 2012). An interesting exception

is a 17-member family with DBE where no mutations in *HOXD13* protein coding sequence were identified, suggesting that mutations in the non-coding regions may be responsible for this phenotype. High-resolution Array Comparative Genome Hybridization (aCGH) analysis has identified a novel ~440 kb microdeletion within the so-called regulatory archipelago of the *HOXD* cluster (*see* **Table 1** below) (Flottmann *et al.*, 2018). This regulatory region is located centromeric to the *HOXD* cluster and contains several limb-specific enhancers (Montavon *et al.*, 2011). Generation of mice harboring a similar mutation resulted in shortening of metacarpals, recapitulating the human phenotype. Deletion of the enhancers located in regulatory archipelago in the mouse led to low expression of *Hoxd13* down to ~10%, which is most likely responsible for the phenotype (Montavon *et al.*, 2011; Flottmann *et al.*, 2018).

The next skeletal disorder associated with enhancer mutations is Split Hand/Foot Malformation 1 (SHFM1, MIM 183600). It is a rare, inherited disorder characterized by aplasia/hypoplasia of phalanges, metacarpals and metatarsals, lack of selected digits and cleft of hands and feet. The SHFM1 is also correlated with deafness, intellectual disability and craniofacial malformations (Lango Allen *et al.*, 2014; Rattanasopha *et al.*, 2018). Seven independent genetic loci have been linked to this malformation so far, including one in the 7q21.3 region (Rattanasopha *et al.*, 2018). Reporter studies in zebrafish and mice have revealed that this region encompasses two conserved limb-specific enhancers (Williamson *et al.*, 2019). They are located in exons 15 and 17 of the dynein cytoplasmic 1 intermediate chain 1 (*DYNC1I1*) gene. Analysis of a 134-sized unrelated family cohort revealed that alteration in these enhancers led to limb malformation (*see* **Table 1** below) (Tayebi *et al.*, 2014). The mechanism of action is most likely based on the misregulation of their target genes: distal-less homeobox 5 (*DLX5*) and distal-less homeobox 6 (*DLX6*). Importantly, both the *DLX5* and *DLX6* have been previously linked to SHFM type-I (Birnbaum *et al.*, 2012).

One of the most prominent examples of non-coding mutations linked to skeletal disorders are located in ZRS, a highly conserved, limb-specific intragenic enhancer, located around 1 Mb upstream of *SHH* (Van der Meer *et al.*, 2012). Mutations within this region lead to ectopic expression of *SHH*, which results in alteration of the A-P axis and patterning of the digits. Interestingly, distinct size mutations in the ZRS region may lead to different phenotypes (**Table 1**). For instance, two microduplications (~255 kb and ~179 kb) have been linked to Hass-type polysyndactyly [MIM 186200] featured by supernumerary fingers with cup-shaped syndactyly of hands (Lohan *et al.*, 2014). Another group of microduplications (~75 kb; ~47 kb; ~16 kb) has been associated with the Laurin-Sandrow syndrome [MIM 135750] characterized by polysyndactyly of hands, overlapping with Hass-type polysyndactyly, but also additional skeletal defect such as mirror image polysyndactyly of feet, duplication of fibula with lack of tibia, and nasal defects Lohan *et al.*, 2014). Finally, point mutations within ZRS have been associated with tibial hemimelia-polydactyly-triphalangeal thumb syndrome [MIM 188740], characterized by short lower limbs with lack of tibias, polydactyly and digitalization of the thumbs Cho *et al.*, 2013). Also patients with pre-axial polydactyly, post-axial polydactyly, syndactyly and triphalangeal thumb have been linked to point mutation in ZRS [MIM 174500] (Van der Meer *et al.*, 2012). Correlation between the size of mutations and diverse skeletal phenotypes may suggest that not only alteration in ZRS, but also disruption of local chromatin architecture may play a key role in the etiology of the limb malformations. However, in contrast to predictions, the genetic manipulations of the local chromatin structure of ZRS-*SHH* promoter region have shown limited or no change of gene regulation during limb development, and no limb malformations in animals with rearranged chromatin architecture (Williamson *et al.*, 2019).

Table 1. Non-coding variants linked to human limb malformations.

Type of element	element/ region	Target	size/ type of variant	Variant localization	phenotype	ref
Enhancer	ZRS	*SHH*	255 kb dup	Chr7:156,437,229-156,692,706 dup (hg19)	Hass-type polysyndactyly (MIM: 186200)	Lohan et al. 2014
			179 kb dup	Chr7:156,491,887-156,671,016 dup (hg19)		
			75 kb dup	Chr7:156,570,780-156,646,750 dup (hg19)	Laurin-Sandrow syndrome (MIM: 135750)	
			47 kb dup	Chr7:156,563,856-156,610,632 dup (hg19)		
			16 kb dup	Chr7:156,578,108-156,594,751 dup (hg19)		
			SNVs	LMBR1:c.423+5252A>G LMBR1:c.423+5134C>G LMBR1:c.423+4915C>T LMBR1:c.423+4842T>C LMBR1:c.423+4808T>C	Tibial hemimelia-polydactyly-triphalangeal thumb syndrome (MIM: 188740)	ClinVar variation IDs: 4902, 4903, 126371, 4900, 4906 Cho *et al.* 2013
			SNVs	LMBR1:c.423+5252A>G LMBR1:c.423+5176A>G LMBR1:c.423+5134C>G LMBR1:c.423+4917G>A LMBR1:c.423+4915C>T LMBR1:c.423+4909C>T LMBR1:c.423+4847T>G LMBR1:c.423+4842T>C LMBR1:c.423+4818A>T LMBR1:c.423+4810G>A LMBR1:c.423+4808T>C LMBR1:c.423+4618C>G LMBR1:c.423+4545T>C	Preaxial polydactyly (MIM: 174500)	ClinVar variation IDs: 4902, 515097, 4903, 4898, 126371, 4907, 30497, 4900, 4899, 30496, 4906, 4897, 426635 Lettice *et al.* 2003
	eExons of DYNC1I1	*DLX5*, *DLX6*	167 kb del	Chr7:95,615,187-95,783,313 (hg19)	Split hand/foot malformation (MIM: 183600)	Tayebi *et al.* 2014
			510 kb del	Chr7:95,624,825-96,135,521 (hg19)		
			205 kb del	Chr7: 95,667,046-95,872,044 (hg19)		
			169 kb del	Chr7:95,693,341-95,862,369 (hg19)		
	Regulatory archipelago of HOXD cluster	*HOXD13*	~440 kb del	Chr2:176,065,894-176,504,173 (hg19)*	Brachydactyly type E (MIM: 113300)	Flottmann *et al.* 2018
	CNE8, CNE9	*SHOX*	~423 dup	ChrX:8 09 125-1 232 802 (hg19)*	Leri-Weill dyschondrosteosis (MIM: 127300)	Bunyan *et al.* 2016
	CNEs		~290 kb dup	ChrX:6 75 000-9 64 000 (hg19)*		
	CNE3, CNE2		~103 kb dup	ChrX:4 40 364-5 43 359 (hg19)*		
	CNE7		~47.5 kb del	Xp22.33 chrX:7 00 549-7 48 093, (GRCh37)		ClinVar variation ID: 66087[7]
	CNE3, CNE5		~286 kb del	ChrY:1 24 349-4 09 949 (GRCh36)*	Idiopathic short stature (MIM: 300582)	Benito-Sanz *et al.* 2012
	CNE3, CNE5		~287 kb dup	ChrX:2 19 634-5 06 573 (hg19)*		Bunyan *et al.* 2016
	CNE9		~113 dup	ChrX:8 28 453-9 41 084*		
CTCF	Boundary site of PAX3 TAD and upstream TAD	*PAX3*	~1.75-1.9 Mb dup	2q35-36†	Brachydactyly	Lupianez *et al.* 2015
TAD	17q24.3 region	*KCNJ2*	~2 Mb dup	17q24.3†	Cooks syndrome (MIM: 106995)	Franke et al. 2016
	2q35 region	*IHH*	~900 kb dup	2q35†	Polysyndactyly	Lupianez et al. 2015
	2q36 region	*WNT6*	1.1 Mb inv	2q36†	Feingold syndrome (MIM: 164280)	
	2q36 region		1.5M dup	2q36†		
miRNA	miR-17~92		165 kb del	13q31.3†	Feingold syndrome (MIM: 164280)	de Pontual *et al.* 2011
	miR199 and miR214	*SP7*, *SMAD1*	~2.4 .Mb del	Chr1:168,007,914-170 448 198 (hg18)*	1q24q25 deletion syndrome	Ashraf *et al.* 2015
	RMRP-S1 and RMRP-S2	Multiple targets including *PTCH2* and *BMPR2*	70A>G Compound heterozygosity for a ~25-5 dup and a 146G>A	9p13†	Cartilage-hair hypoplasia (MIM: 250250)	Rogler *et al.* 2014
lncRNA	HOXC-AS2 and HOXC-AS3	*HOXD* genes	175 kb del	Chr12:54,165,001-54,335,668 (hg19)	Congenital vertical talus	Alvarado *et al.* 2016

*Minimal range of mutation based on aCGH.
†Location based on aCGH. The authors did not provide exact breaking points.
aCGH, array comparative genome hybridisation; CNE, non-coding DNA element; CTCF, CCCTC-binding factor; del, deletion; dup, duplication; inv, inversion; lncRNA, long non-coding RNA;SNV, single nucleotide variant; TAD, topologically associating domain; ZRS, zone of polarising activity regulatory sequence.

SHOX-related congenital limb disorders also have been linked to enhancer mutations. Alterations in enhancers located within the *SHOX* genomic region have been reported in Leri-Weill dyschondrosteosis (LWD; MIM 127300) and Idiopathic Short Stature (ISS; MIM 300582) (Benito-Sanz *et al.*, 2012). The phenotypic spectrum of these disorders includes short stature, mesomelia and Madelung deformity of the wrist (Benito-Sanz *et al.*, 2012; Bunyan *et al.*, 2016). Further comparative genomic and functional studies of the *SHOX cis*-regulatory landscape have identified eight downstream and three upstream conserved non-coding DNA elements (named CNEs) with enhancer activity. Four of these produce eRNA (CNE4, CNE5, CNE7/ECR1 and CNE9/ECS4) (Bunyan *et al.*, 2016). Importantly, mutations in these CNEs affect the formation and growth of limb bones. They most likely lead to defects in *SHOX* expression, resulting in skeletal phenotypes (**Table 1**). For instance, a ~47.5 kb downstream deletion identified in 30 patients with LWD or ISS encompasses the CNE7-regulatory element, leading to misregulation of *SHOX* (Benito-Sanz *et al.*, 2012). In another case, a ~286 kb deletion of two upstream enhancers has been associated with the ISS phenotype of two probands. Both patients have been reported with haploinsufficiency of *SHOX* (Benito-Sanz *et al.*, 2012). Moreover, duplication of CNEs may also lead to a reduction of SHOX mRNA levels and cause the skeletal phenotype. For example, duplication encompassing upstream *cis*-regulatory elements CNE3 and CNE5 have been reported in one family with ISS; another duplication including CNE9 has been associated with the phenotype of four other ISS patients (Bunyan *et al.*, 2016). Moreover, a ~290 kb duplication affecting all of the identified upstream CNEs was reported in one family with LWD; other cases of LWD were associated with duplications of downstream CNE8, CNE9, and X:970,000 putative element (Bunyan *et al.*, 2016). The mechanism of pathogenic effects is most likely based on the disruption of local chromatin architecture, which affects the long-range interaction between *cis*-regulatory elements and the *SHOX* promoter.

4. Chromatin organization and topologically associating domains (TADs)

4.1. TAD boundaries and intra-TAD DNA-looping

Studies of the three-dimensional (3D) chromatin structure have provided new insights into the role of chromatin architecture in the gene regulation and enhancer specificity. Chromatin is organized in the 3D nuclear space in a specific fashion. Each chromosome is located in a definite nucleus location, termed chromosome territory (Habermann *et al.*, 2001). Additionally, analysis of Hi-C contact maps has shown that mammalian genomes are organized into different size TADs, delineated by the boundaries enriched in CTCF binding-sites. The size of TADs may be hundreds of kilobases up to several millions bases, with average length of ~880 kb (Dixon *et al.*, 2016). The CCCTC-binding factor CTCF is an insulator protein that inhibits the interaction between enhancers and gene promoters (**Fig. 1b.1**). Thus, binding of CTCF proteins defines the TADs boundaries and is responsible for a specific pattern of enhancer-promoter interaction, which when mutated, may lead to limb malformations (Lupiañez *et al.*, 2015; Barrington *et al.*, 2017). The polymer simulations and chromatin architecture studies have discovered that intra-TAD interactions are dynamic (Fudenberg *et al.*, 2016), whereas the TADs organization within the genome is highly stable and conserved across cell types, tissues, and related species (Lupiañez *et al.*, 2015; Krefting *et al.*, 2018).

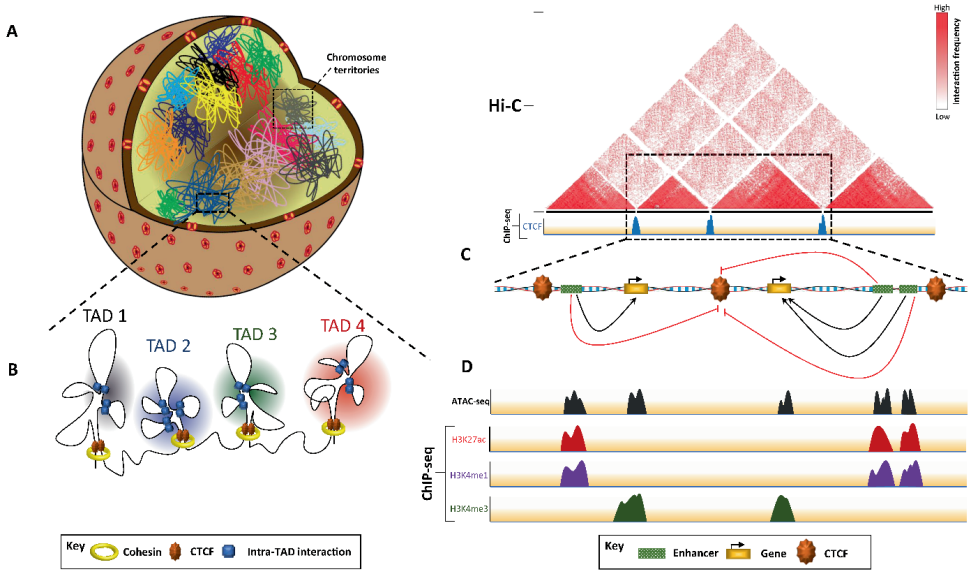


Figure 1b.1. Chromatin organization in gene regulation.

A. At the nuclear level, chromatin is organized in chromosomes, which occupy the defined territories. At the chromosomal level, chromatin is organized in topologically associating domains (TADs). The TADs organization is stable and conserved among tissues and species, while intra-TAD interactions are highly dynamic. The cohesin complex and CTCF are main factors which play a role in the formation of TADs and intra-TAD interactions. **B.** TADs have been originally discovered by Hi-C technique. They act as regulatory units within which *cis*-regulatory elements can interact with gene promotes. The interaction between regulatory elements from neighbor TADs is inhibited by boundary sides separating TADs, which are enriched in CTCF binding regions. The ChIP-seq together with ATAC-seq and functional assay data have helped to characterize the content of regulatory elements within the TADs. Several epigenetic marks may be used to identify active enhancers, such as Histone H3 Lys27 acetylation (H3K27ac), or Histone H3 Lys4 monomethylation (H3K4me1). Other marks may be used to interrogate promoter regions, such as Histone H3 Lys4 trimethylation (H3K4me3).

Based on the loop extrusion model, TADs are formed due to the cohesin-CTCF dependent chromatin loop formation (Habermann *et al.*, 2001; Gassler *et al.*, 2017). The process begins with attaching a cohesin complex to the chromatin. The ring structure of cohesin complex allows interacting with two chromatin regions at the same time leading to formation and extrusion of a chromatin loop until the cohesin reaches the CTCF boundary sites (Dixon *et al.*, 2016; Fudenberg *et al.*, 2016; Barrington *et al.*, 2017). Therefore, the interactions between *cis*-regulatory elements and gene promoters are coordinated by both the presence of cohesin complex and CTCF binding-sites. Importantly, the deletion of CTCF binding-sites causes loss of the "insulation effect", resulting in TADs rearrangements and gene misregulation, correlated with the etiology of some limb malformations (Lunyak *et al.*, 2007; Dixon *et al.*, 2012; Lupiañez *et al.*, 2015). However, it is important to mention that despite overwhelming evidence about the role of TADs boundaries in gene regulation, the genome-wide depletion of CTCF binding has a limited effect on global gene transcription (described in Despang *et al.*, 2019). Additionally, the deletion of CTCF binding-sites located at the TAD boundaries around *Sox9* and *Kcnj2* loci has caused an increase in inter-TAD interaction without TADs' fusion. The deletions of all major intra-TAD CTCF binding-sites resulted in TADs' fusion suggesting that CTCF sites located at the boundaries as well as intra-CTCF sites play a role in formation of TADs (Despang *et al.*, 2019). Surprisingly, it was shown that fusion of *Sox9-Kcnj2* TADs caused by deletion of CTCF binding-sites do not cause strong changes in gene

expression of the genes, suggesting that in this case TADs boundaries were non-essential for gene regulation. However, inversions and/or insertions of *Sox9-Kcnj2* TADs' boundaries resulted in gene misexpression (Despang *et al.*, 2019). Therefore, it suggests that CTCF located at TADs boundaries may not be always essential for gene regulation, however, when redirected they may result in gene misexpression.

The chromatin architecture also has an impact on gene expression by controlling the accessibility of proteins to transcriptionally active or inactive genomic regions. Interestingly, while the transcriptionally active chromatin tends to localize away from the nuclear envelope, transcriptionally silent chromatin locates within the nuclear space close to the membrane, and interact with the nuclear lamina (NL) (van Steensel and Belmont, 2017). The chromatin localized near to the NL is termed lamina-associated domains (LADs) and is marked by the features typical for heterochromatin, such as absence or low transcription level, and H3K27me3 histone modification. Collectively, the chromatin compaction together with its architecture and nuclear localization play a crucial role in gene regulation, which when disrupted may be responsible for molecular etiology of limb abnormalities.

4.2. Disruption of TADs leads to deregulation of genes and hereditary diseases

Studies of three unrelated patient families with a novel type of brachydactyly have revealed that deletion of CTCF binding-sites led to local gene deregulation. Specifically, the 4C-seq data have shown that absence of the CTCF boundary site separating Paired Box-3 (*PAX3*) from the next TAD, put *PAX3* under the control of the normally inaccessible enhancer, so rewiring the enhancers-*PAX3* promoter interaction (**Table 1**) (Lupiañez *et al.*, 2015). Such interaction is termed "enhancer adoption" and may be caused either by a fusion of two TADs or formation of a new TAD with the novel regulatory landscape. Importantly, the generation of mice harboring a similar deletion resulted in upregulation of *Pax3* and in brachydactyly-like phenotype, indicating that deregulation of *PAX3*, caused by rewiring E-P interactions, plays a crucial role in the skeletal phenotypes (Lupiañez *et al.*, 2015).

Cooks syndrome (MIM 106995) is another disorder linking reorganization of TADs with etiology of limb malformations (Franke *et al.*, 2016). It is an autosomal dominant ectodermal dysplasia characterized by brachydactyly, lack of phalanges, absence of the nails, and abnormal length of phalanges (Kurth *et al.*, 2009). Genetic analysis of seven affected individuals revealed that ~2 Mb duplications in the 17q24.3 region led to the disruption of local chromatin architecture and formation of a new TAD (*see* **Table 1**). This mutation changed the regulatory landscape of the potassium voltage-gated channel subfamily J member-2 *KCNJ2*, leading to its misexpression. Functional studies of inwardly rectifying K+ channel Irk2 (Irk2), a *KCNJ2 Drosophila* homolog, have shown that Irk2 is involved it dpp/BMP signaling (Dahal *et al.*, 2012). Since several mutations in the components of the BMP pathway play a role in etiology of skeletal disorders, it is tempting to suggest that the defect in the expression of *KCNJ2* may be an important factor in this limb malformation.

Feingold-syndrome (F-syndrome, MIM 164280) is a limb disorder, which also has been linked to alteration in chromatin architecture of TADs (Lupiañez *et al.*, 2015). Additionally, F-syndrome has been correlated with mutation in proto-oncogene *MYCN* and microdeletion encompassing the miR-17~92 region (*see* below). Thus, this is an example of a condition wherein the same phenotype is induced by different types of mutation. It is an autosomal dominant

disorder described by short stature, thumb hypoplasia, brachymesophalangy of second and fifth fingers, microcephaly and toe syndactyly. Additionally, F-syndrome has been correlated with learning disability, heart and kidney anomalies, hearing loss and asplenia (Mathonnet *et al.*, 2007; Marcelis *et al.*, 2008). The 4-C data analysis of two unrelated families has shown that a 1.1 Mb inversion or 1.5 Mb duplication led to alterations in chromatin architecture, affecting the organization of the TADs (**Table 1**). Thus, rearrangement of chromatin structure led to disruption of the Wnt family member-6 gene (*WNT6*) *cis*-regulatory landscape, causing its deregulation by rewiring E-P connections (Lupiañez *et al.*, 2015). *WNT6* is involved in chondrogenesis and bone formation, and its ectopic expression has been associated with limb malformation in chicken (Geetha-Loganathan *et al.*, 2010). Finally, reconstruction of a 1.1 Mb inversion in mice resulted in enhancer adoption and ectopic expression of *Wnt6* in distal limb autopod mesenchyme (Lupiañez *et al.*, 2015). Altogether, this strongly suggests the causative role of *WNT6* misexpression in etiology of human F-syndrome.

Studies of two probands with a polysyndactyly have correlated rearrangements within TADs with limb malformation. Polysyndactyly is a common skeletal defect characterized by an additional number of fingers and/or toes that are connected by interdigital webbing (Yuksel-Apak *et al.*, 2012). The 4C data analysis of two human samples carrying a ~900 kb duplication encompassing the Indian hedgehog (*IHH*) locus (2q35) has revealed that this mutation led to the shift of the *IHH* gene to an *EPHA4*-containing TAD (**Table 1**). In addition, a transposition of *IHH* to a different TAD led to enhancer adoption and *IHH* misexpression (Lupiañez *et al.*, 2015), and patients carrying such mutation have been also diagnosed with craniofacial abnormalities. The ~600 kb deletion region encompassing the *Ihh* locus (2q35) in mice also resulted in *Ihh* enhancer adoption, causing severe polydactyly and craniofacial dysmorphism (Babbs *et al.*, 2008), altogether supporting the role of chromatin rearrangements in etiology of polysyndactyly.

5. MicroRNAs – post-transcriptional gene regulators

5.1. Formation and actions of miRNAs

MicroRNAs (miRNA) are small, ~22 nucleotides-long non-coding RNAs (Landgraf *et al.*, 2007). Depending on the origin, miRNAs are synthesized by canonical or non-canonical pathways (reviewed in Ha and Kim, 2014). They act as *trans*-regulatory elements, coordinating gene expression through two main mechanisms: translational repression, or mRNA deadenylation followed by decay (**Fig. 1b.2**) (Mathonnet *et al.*, 2007; Huntzinger *et al.*, 2013).

The miRNAs cooperate with several proteins leading to the assembly of the RNA-induced silencing complex (RISC). Once miRNA attach to Argonaute (AGO) protein, the RISC is able to locate cognate mRNA and usually binds to its 3'-UTR region. The target site is recognized by the complementarity with 2-7 nt/2-8 nt miRNA "seed sequence" (Lewis *et al.*, 2005; Braun *et al.*, 2011). Another molecule interacting with AGO is TNRC6/GW182. It recruits the CCR4–NOT and PAN2-PAN3 deadenylase complexes responsible for the shortening of the 3′-poly(A) tail of mRNA (Lewis *et al.*, 2005; Braun *et al.*, 2011; Huntzinger *et al.*, 2013). Moreover, the TNRC6/GW182 and AGO interact with proteins involved in removing of the mRNA 5'-^{7me}G cap initiating targeted degradation of mRNA (Behm-Ansmant *et al.*, 2006; Rouya *et al.*, 2014). The RISC is also involved in the inhibition the formation of translation initiation complex, crucial for proper

interaction of mRNA with the ribosome (Mathonnet *et al.*, 2007; Zekri *et al.*, 2013; Chen *et al.*, 2017; Amaya Ramirez *et al.*, 2018).

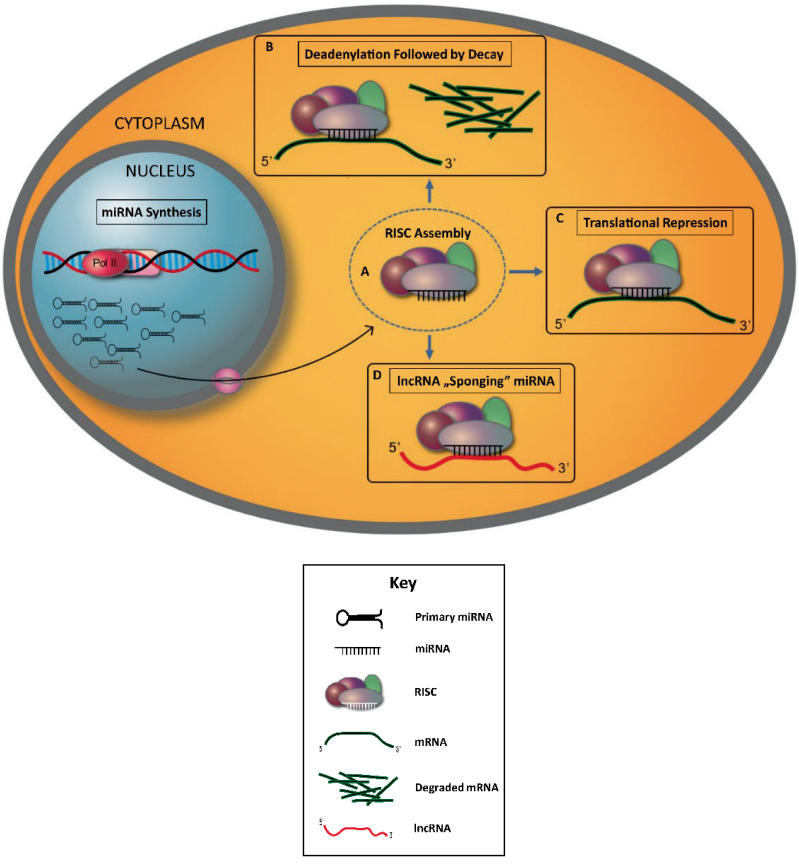


Figure 1b.2. Role of miRNA in post-transcriptional regulation of gene expression.

A. After transcription and maturation, the miRNAs interact with several proteins (e.g., Ago and TNRC6) leading to the assembly of RNA-induced silencing complex RISC. **B.** The miRNA is involved in the coordination of mRNAs level via initiation of targeted deadenylation followed by mRNA degradation. **C.** The miRNAs regulate gene expression of target genes due to a negative influence on the proper formation of initiation complex, leading to translational repression. **D.** A small fraction of "free" miRNA in the cytoplasm is caused by lncRNA "sponging", which affect the dynamics of miRNA-mRNA interaction events and regulate gene expression either negatively, or positively, depending on the target mRNA.

5.2. Disruption of miRNA genes leads to limb abnormalities

An added level of regulatory complexity is caused by the fact that a single miRNA may target several genes, while one gene can be regulated by different miRNAs (Friedman *et al.*, 2009). In those cases, miRNAs compete for target mRNA with other class of non-coding RNAs, such as long non-coding RNAs (lncRNAs) (reviewed in Yoon *et al.*, 2014). The crosstalk between miRNAs, lncRNAs and mRNAs creates a complex gene regulatory network, which when disrupted by mutation, or misexpression of miRNA may lead to disorders, including limb malformations.

One of the examples linking limb abnormalities to mutations in miRNAs is the aforementioned F-syndrome. As previously described, this disorder was correlated with mutations in the MYCN and disruption of *WNT6 cis*-regulatory landscape caused by chromatin rearrangement (Marcelis *et al.*, 2008; Grote *et al.*, 2015; Lupiañez *et al.*, 2015). Nevertheless, the analysis of high-resolution comparative genomic hybridization (CGH) data from ten F-syndrome patients has identified that microdeletion affecting the miR-17~92 region also plays a crucial role in the molecular etiology of this disorder (de Pontual *et al.*, 2011). Reconstruction of the mutation in mouse led to brachymesophalangy of the second and fifth digit, similar to the patient's phenotypes. The bi-allelic mutation of miR-17~92 was lethal. Mouse mutant embryos (*miR-17~92*$^{\Delta/\Delta}$) had hypoplasia of the mesophalangx of second digit and lack of the mesophalangx of fifth digit (de Ponutal *et al.*, 2011). Since the patients with F-syndrome skeletal defects have been characterized to carry mutations either in *MYCN* or *miR-17~92*, or have altered *WNT6* expression caused by TAD reorganization, it is tempting to speculate that these three components may be involved in the same developmental pathway. Indeed, MYCN may directly bind to the region encoding miR-17~92, leading to transcriptional activation. Furthermore, the studies of neuroblastoma cell lines with or without expression of *MYCN* suggest that MYCN is involved in regulation of *WNT6* (Liu *et al.*, 2008). Altogether, these data strongly imply that mutations in all abovementioned regions are involved in molecular etiology of F-syndrome through the modulation of a WNT6-dependent pathway.

Another example of limb malformation linked to a mutation in miRNAs is the 1q24q25 Deletion Syndrome. It is a rare inherited disorder characterized by short stature, small hands and feet, brachydactyly, clinodactyly, microcephaly, and severe cognitive disability (Burkardt *et al.*, 2011). Currently, only few patients with skeletal abnormalities and lack of cognitive disability have been reported, suggesting that microdeletion narrowed to the dynamin-3 (*DNM3*) gene region could play a vital role in the skeletal phenotype (**Table 1**) (Ashraf *et al.*, 2015). Yet, the expression of DNM3 mRNA was detected at high levels in the brain and low levels in the skeleton, excluding the *DNM3* gene as a potential candidate for the syndrome (Burkardt *et al.*, 2011). Interestingly, two miRNAs (miR199 and miR214) are encoded within intron-14 of *DNM3*. miR214 is involved in the inhibition of osteogenic differentiation through regulation of osteoblast-specific TF SP7, while miR199 affects BMP signaling by targeting SMAD1 (Desvignes *et al.*, 2014; Zhao *et al.*, 2015). Moreover, alteration of the miR199-miR214 cluster in a mutant mouse model led to short stature and cranial deformity, partially recapitulating the patients' phenotype, and strongly suggesting the role of the miR199-miR214 cluster in the etiology of limb malformation of 1q24q25 Deletion Syndrome (Watanabe *et al.*, 2008).

Cartilage-Hair Hypoplasia (CHH, MIM 250250) is another example of limb disorder associated with the misexpression of miRNA (**Table 1**). It is an autosomal recessive disorder characterized by severe short-limbed dwarfism, metaphyseal dysplasia, immunodeficiency, predisposition to cancers and severe anemia. The phenotypes are linked to mutations in *RMRP*, a gene encoding one of the units of RNase MRP ribonucleoprotein complex (Rogler *et al.*, 2014; Cherkaoui Jaouad *et al.*, 2015). Interestingly, expression analysis of fibroblast and B-cells from CHH patients has revealed that alteration in *RMRP* expression led to a significant reduction of two miRNAs, RMRP-S1 and RMRP-S2. Both miRNAs are involved in the regulation of more than 900 targets including downregulation of Patched 2 (encoded by *PTCH2*), which results in induction of SHH signaling and downregulation of BMP receptor type 2 *BMPR2*, decreasing BMP signaling in limb development (Rogler *et al.*, 2014). Collectively, it suggests that miRNAs may be important factors in the skeletal phenotype of CHH.

6. Long non-coding RNAs

6.1. lncRNAs as versatile regulators

Long non-coding RNAs (lncRNAs) are a class of functional RNAs longer than 200 bp and devoid of protein-coding capability. They are typically tissue-specific with characteristic spatio-temporal expression pattern (Yan *et al.*, 2017). A number of studies have shown that lncRNAs regulate various biological processes including chromatin remodeling, epigenetic modification of DNA, transcription, translation or mRNA splicing (**Fig. 1b.3**) (Tripathi *et al.*, 2010; Gong and Maquat, 2011; Su *et al.*, 2017). Importantly, they can affect gene expression not only by acting in *cis* but also *in trans* (reviewed in Yan *et al.*, 2017). Thus, lncRNAs most likely function as a structural scaffold, recruiting regulatory complexes or interacting with the gene promoter itself. For instance, ANRIL lncRNA (an antisense non-coding RNA in the *INK4* locus), by recruiting PRC2, leads to repression of *p15*^{INK4B} (Kotake *et al.*, 2011). EZR-AS1, by interaction with Histone-Lysine N-methyltransferase SMYD3, causes enrichment of H3K4me3 in the Ezrin gene (*EZR*) promoter region, leading to higher expression of EZR (Zhang *et al.*, 2018). Another lncRNA, derived from the region upstream of the dihydrofolate reductase gene (*DHFR*), acts by the formation of a stable complex with the *DHFR* major promoter, repressing its expression (Martianov *et al.*, 2007).

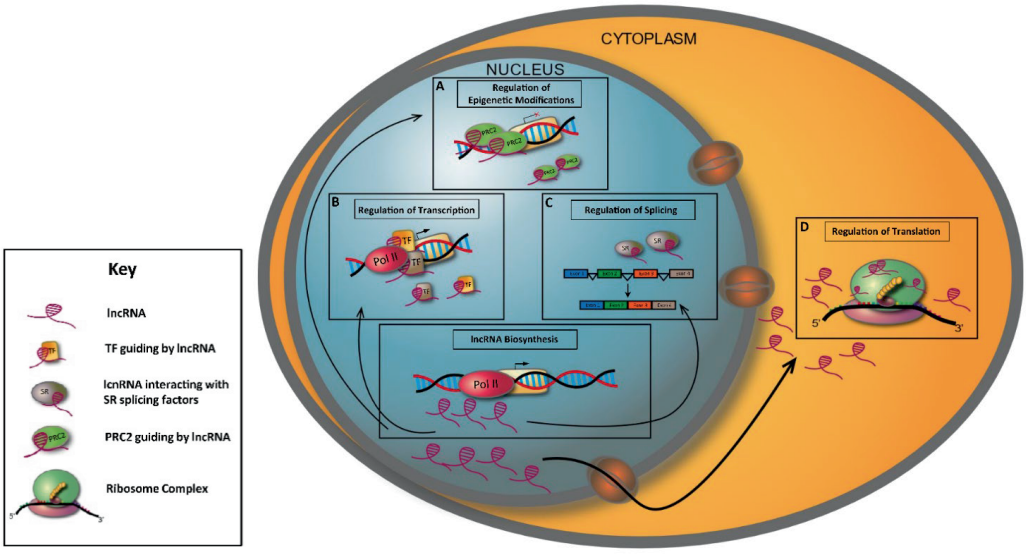


Figure 1b.3. Role of lncRNA in the regulation of gene expression.

A. The lncRNAs indirectly regulate epigenetic silencing of gene expression by bringing silencing complexes, i.e. PRC2, to target genes. **B.** The lncRNAs interact with transcription factors and transport them to gene promoters, modulating the transcription of specific genes. **C.** The lncRNA via interaction with SR splicing factors facilitates the alternative splicing. **D.** Several lncRNAs interact with ribosome complexes which suggest their regulatory roles in the translation process.

Additionally, lncRNAs through bridging TFs to gene regulatory elements, coordinate expression of multiple genes at the same time. For example, the E2F1 TF due to interaction with lncRNA-HIT modulates the expression of its target genes, such as *FOXM1*, S-phase kinase associated protein 2 *SKP2*, Neural EGFL like-2 *NELL2* and Docking protein-1 *DOK1* (Yu *et al.*, 2017). Besides transcription control, lncRNAs are also involved in the regulation of gene expression at the posttranscriptional level. For instance, MALAT1 lncRNA coordinates alternative splicing of mRNA by interaction with SR splicing factors (Tripahti *et al.*, 2010). Another lncRNA 5S-OT affects splicing due to interplay with U2AF65 splicing factor (Hu *et al.*, 2016). Additionally, numerous lncRNAs participate in Staufen STAU1-mediated mRNA decay. They are mainly involved in the formation of the binding site for STAU1 protein (Gong and Maquat, 2011; Lucas *et al.*, 2018). LncRNAs play also a role in the crosstalk between competing endogenous RNAs (ceRNA), within mRNA-miRNA-lncRNA regulation network (Fan *et al.*, 2018). For instance, they can function as a "sponge" for miRNAs or compete with miRNAs for mRNA binding-sites. For instance, MALAT1 and ANCR lncRNAs play a vital role in modulation of osteogenic differentiation, due to "sponging" of miRNA-143 and miRNA-758 (Gao *et al.*, 2018; Peng *et al.*, 2018).

Intriguingly, the lncRNA synthesis process itself may regulate the gene expression. For example, the transcription of Airn lncRNA affects the expression of the insulin-like growth factor 2 receptor *IGF2R* gene and does not require Airn transcript. The mechanism of action is based on the transcriptional overlap of the promoter region of the target gene, which reduces the RNAPol2 recruitment (Latos *et al.*, 2012).

6.2. Potential role of lncRNA genes in the development of limb abnormalities

Currently, more than 51,000 lncRNA genes have been annotated in the human genome (Volders *et al.*, 2015). Taking under the consideration the functional role of lncRNAs in gene regulation, it opens a possibility that alteration in lncRNAs may play a role in limb malformations.

Genetic characterization of patients with clubfoot and congenital vertical talus supports this hypothesis, both disorders are closely related and may occur in the same family, or individual. The former is characterized by rigid flatfoot deformity, the latter is manifested by the inward position of the foot/feet (Alvarado *et al.*, 2016). Analysis of Affymetrix genome-wide human SNP array 6.0 and multiplexed direct genomic selection (MDiGS) form 226 individuals with clubfoot and 27 with congenital vertical talus pointed out four deletions in the 5' of *HOXC* gene cluster, ranging from 13 to 175 kb (**Table 1**) (Alvarado *et al.*, 2016). Several lncRNAs are located within this region, such as HOTAIR, HOXC-AS2, HOXC-AS3, HOXC13-AS. HOTAIR play a role in the repression of *HOXD10* by guiding PRC2 silencing complexes to its region, and mutation in *HOXD10* has been linked with congenital vertical talus (Rinn and Chang, 2012). The gene expression analysis in the patient with the 175-kb deletion revealed a significant reduction of *HOXD10* and two lncRNAs, HOXC-AS2 and HOXC-AS3. Despite the fact that expression of HOTAIR was unaffected, it could be possible that other lncRNAs located in 5' of the *HOXC* gene cluster might play a role in the regulation of *HOXD* genes (Alvarado *et al.*, 2016). Altogether, it is suggestive that alteration of trans-regulation of the *HOXD* genes due to misexpression, or mutation in lncRNAs may be involved in the etiology of clubfoot and congenital vertical talus.

7. Conclusions and future perspectives

Recent technological advances in NGS-based methods, resulted in discovery of new layers of gene regulation. It also allowed for the identification of numerous mutations in gene regulatory elements linked to congenital limb disorders. Therefore, the examination of regulatory elements, or regulome, provides the answers not only to the basic science questions but, importantly, gives clinically relevant information applicable to the diagnosis of patients. Better understanding of *cis*/*trans* regulatory elements may also lead to the development of potentially promising pharmacological solutions in limb malformations treatment. For instance, a small molecule targeting enhancers regions could be applied for correction of gene misexpression, whereas targeting of miRNA and lncRNA could be used to specific regulation of genes.

Competing Interests

None declared.

Funding

This work was supported by the Polish National Science Centre [PRO-2013/11/B/NZ4/03660 to PT, UMO-2015/19/B/NZ4/03184 to PT, and PRO-2017/27/N/NZ5/02940 to EHK].

Contributorship statement

KN was responsible for drafting, writing of the manuscript, preparing the figures, and corrections.
EHK was responsible for drafting, writing, preparing Table 1, and corrections.
PT was responsible for drafting, writing, and corrections.

WEB resources

https://www.omim.org/entry/186200?search=186200&highlight=186200
https://www.omim.org/entry/135750?search=135750%20&highlight=135750
https://www.omim.org/entry/188740?search=188740&highlight=188740
https://www.omim.org/entry/174500?search=174500&highlight=174500
https://www.omim.org/entry/188740
https://www.omim.org/entry/183600?search=183600&highlight=183600
https://www.omim.org/entry/113300?search=113300&highlight=113300
https://www.omim.org/entry/127300?search=127300&highlight=127300
https://www.omim.org/entry/300582?search=300582&highlight=300582
https://www.omim.org/entry/106995?search=106995&highlight=106995
https://www.omim.org/entry/164280?search=164280&highlight=164280
https://www.omim.org/entry/250250?search=250250&highlight=250250

Literature references

Alvarado DM, McCall K, Hecht JT, Dobbs MB, Gurnett CA. Deletions of 5' HOXC genes are associated with lower extremity malformations, including clubfoot and vertical talus. J Med Genet. 2016;53(4):250-5.

Amaya Ramirez CC, Hubbe P, Mandel N, Bethune J. 4EHP-independent repression of endogenous mRNAs by the RNA-binding protein GIGYF2. Nucleic Acids Res. 2018;46(11):5792-808.

Ashraf T, Collinson MN, Fairhurst J, Wang R, Wilson LC, Foulds N. Two further patients with the 1q24 deletion syndrome expand the phenotype: A possible role for the miR199-214 cluster in the skeletal features of the condition. Am J Med Genet A. 2015;167A(12):3153-60.

Babbs C, Furniss D, Morriss-Kay GM, Wilkie AO. Polydactyly in the mouse mutant Doublefoot involves altered Gli3 processing and is caused by a large deletion in cis to Indian hedgehog. Mech Dev. 2008;125(5-6):517-26.

Barrington C, Finn R, Hadjur S. Cohesin biology meets the loop extrusion model. Chromosome Res. 2017;25(1):51-60.

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev. 2006;20(14):1885-98.

Benito-Sanz S, Aza-Carmona M, Rodriguez-Estevez A, Rica-Etxebarria I, Gracia R, Campos-Barros A, Heath KE. Identification of the first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature. Eur J Hum Genet. 2012;20(1):125-7.

Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernandez B, Barreda-Bonis AC, Liu P, Gracia R, Lupski JR, Campos-Barros A, Gomez-Skarmeta JL, Heath KE. Identification of the first recurrent PAR1 deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. J Med Genet. 2012;49(7):442-50.

Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, Gurrieri F, Everman DB, Schwartz CE, Birk OS, Bejerano G, Lomvardas S, Ahituv N. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012;22(6):1059-68.

Birnbaum RY, Everman DB, Murphy KK, Gurrieri F, Schwartz CE, Ahituv N. Functional characterization of tissue-specific enhancers in the DLX5/6 locus. Hum Mol Genet. 2012;21(22):4930-8.

Braun JE, Huntzinger E, Fauser M, Izaurralde E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell. 2011;44(1):120-33.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21 29 1-9.

Bunyan DJ, Baffico M, Capone L, Vannelli S, Iughetti L, Schmitt S, Taylor EJ, Herridge AA, Shears D, Forabosco A, Coviello DA. Duplications upstream and downstream of SHOX identified as novel causes of Leri-Weill dyschondrosteosis or idiopathic short stature. Am J Med Genet A. 2016;170A(4):949-57.

Burkardt DD, Rosenfeld JA, Helgeson ML, Angle B, Banks V, Smith WE, Gripp KW, Moline J, Moran RT, Niyazov DM, Stevens CA, Zackai E, Lebel RR, Ashley DG, Kramer N, Lachman RS, Graham JM, Jr. Distinctive phenotype in 9 patients with deletion of chromosome 1q24-q25. Am J Med Genet A. 2011;155A(6):1336-51.

Chen S, Gao G. MicroRNAs recruit eIF4E2 to repress translation of target mRNAs. Protein Cell. 2017;8(10):750-61.

Cherkaoui Jaouad I, Laarabi FZ, Chafai Elalaoui S, Lyonnet S, Henrion-Caude A, Sefiani A. Novel Mutation and Structural RNA Analysis of the Non-coding RNase MRP Gene in Cartilage-Hair Hypoplasia. Mol Syndromol. 2015;6(2):77-82.

Cho TJ, Baek GH, Lee HR, Moon HJ, Yoo WJ, Choi IH. Tibial hemimelia-polydactyly-five-fingered hand syndrome associated with a 404 G>A mutation in a distant sonic hedgehog cis-regulator (ZRS): a case report. J Pediatr Orthop B. 2013;22(3):219-21.

Chokeshaiusaha K, Puthier D, Nguyen C, Sananmuang T. A demonstration of the H3 trimethylation ChIP-seq analysis of galline follicular mesenchymal cells and male germ cells. Asian-Australas J Anim Sci. 2018;31(6):791-97.

Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, van Ijcken WFJ, Grosveld F, Frommolt P, Bazzi H, Rada-Iglesias A. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell. 2017;20(5):689-705 e9.

Dahal GR, Rawson J, Gassaway B, Kwok B, Tong Y, Ptacek LJ, Bates E. An inwardly rectifying K+ channel is required for patterning. Development. 2012;139(19):3653-64.

de Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, Van Haeringen A, Genevieve D, Goldenberg A, Oufadem M, Manouvrier S, Munnich A, Vidigal JA, Vekemans M, Lyonnet S, Henrion-Caude A, Ventura A, Amiel J. Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. Nat Genet. 2011;43(10):1026-30.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010;8(5):e1000384.

Despang A, Schopflin R, Franke M, Ali S, Jerkovic I, Paliou C, Chan WL, Timmermann B, Wittler L, Vingron M, Mundlos S, Ibrahim DM. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 2019;51(8):1263-71.

Desvignes T, Contreras A, Postlethwait JH. Evolution of the miR199-214 cluster and vertebrate skeletal development. RNA Biol. 2014;11(4):281-94.

Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol Cell. 2016;62(5):668-80.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376-80.

Ekblom AG, Laurell T, Arner M. Epidemiology of congenital upper limb anomalies in 562 children born in 1997 to 2007: a total population study from stockholm, sweden. J Hand Surg Am. 2010;35(11):1742-54.

Fan CN, Ma L, Liu N. Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. J Transl Med. 2018;16(1):264.

Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017;2017.

Flottmann R, Kragesteen BK, Geuer S, Socha M, Allou L, Sowinska-Seidler A, Bosquillon de Jarcy L, Wagner J, Jamsheer A, Oehl-Jaschkowitz B, Wittler L, de Silva D, Kurth I, Maya I, Santos-Simarro F, Hulsemann W, Klopocki E, Mountford R, Fryer A, Borck G, Horn D, Lapunzina P, Wilson M, Mascrez B, Duboule D, Mundlos S, Spielmann M. Non-coding copy-number variations are associated with congenital limb malformation. Genet Med. 2018;20(6):599-607.

Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schopflin R, Kraft K, Kempfer R, Jerkovic I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538(7624):265-69.

Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15(9):2038-49.

Gao Y, Xiao F, Wang C, Wang C, Cui P, Zhang X, Chen X. Long non-coding RNA MALAT1 promotes osterix expression to regulate osteogenic differentiation by targeting miRNA-143 in human bone marrow-derived mesenchymal stem cells. J Cell Biochem. 2018;119(8):6986-96.

Gassler J, Brandao HB, Imakaev M, Flyamer IM, Ladstatter S, Bickmore WA, Peters JM, Mirny LA, Tachibana K. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. EMBO J. 2017;36(24):3600-18.

Geetha-Loganathan P, Nimmagadda S, Christ B, Huang RJ, Scaal M. Ectodermal Wnt6 is an early negative regulator of limb chondrogenesis in the chicken embryo. BMC Dev Biol. 2010;10.

Giele H, Giele C, Bower C, Allison M. The incidence and epidemiology of congenital upper limb anomalies: a total population study. J Hand Surg Am. 2001;26(4):628-34.

Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011;470(7333):284-8 doi: 10.1038/nature09701.

Gorkin DU, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, Beer MA, Pavan WJ, McCallion AS. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. Genome Res. 2012;22(11):2290-301.

Grote LE, Repnikova EA, Amudhavalli SM. Expanding the phenotype of feingold syndrome-2. Am J Med Genet A. 2015;167A(12):3219-25.

Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.

Habermann FA, Cremer M, Walter J, Kreth G, von Hase J, Bauer K, Wienberg J, Cremer C, Cremer T, Solovei I. Arrangements of macro- and microchromosomes in chicken cells. Chromosome Res. 2001;9(7):569-84.

Hu S, Wang X, Shan G. Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing. Nat Struct Mol Biol. 2016;23(11):1011-19.

Huntzinger E, Kuzuoglu-Ozturk D, Braun JE, Eulalio A, Wohlbold L, Izaurralde E. The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res. 2013;41(2):978-94.

Jamsheer A, Sowinska A, Kaczmarek L, Latos-Bielenska A. Isolated brachydactyly type E caused by a HOXD13 nonsense mutation: a case report. BMC Med Genet. 2012;13:4.

Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J. Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res. 2017;27(1):64-74.

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011;30(16):1956-62.

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

Kurth I, Klopocki E, Stricker S, van Oosterwijk J, Vanek S, Altmann J, Santos HG, van Harssel JJ, de Ravel T, Wilkie AO, Gal A, Mundlos S. Duplications of non-coding elements 5' of SOX9 are associated with brachydactyly-anonychia. Nat Genet. 2009;41(8):862-3.

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissieres V, Pickle CS, Plajzer-Frick I, Lee EA, Kato M, Garvin TH, Akiyama JA, Afzal V, Lopez-Rios J, Rubin EM, Dickel DE, Pennacchio LA, Visel A. Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell. 2016;167(3):633-42 e11.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007;129(7):1401-14.

Lango Allen H, Caswell R, Xie W, Xu X, Wragg C, Turnpenny PD, Turner CL, Weedon MN, Ellard S. Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC1I1 exonic enhancers of DLX5/6 expression in humans. J Med Genet. 2014;51(4):264-7.

Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. 2012;338(6113):1469-72.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003;12(14):1725-35.

Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.

Liu X, Mazanek P, Dam V, Wang Q, Zhao H, Guo R, Jagannathan J, Cnaan A, Maris JM, Hogarty MD. Deregulated Wnt/beta-catenin program in high-risk neuroblastomas without MYCN amplification. Oncogene. 2008;27(10):1478-88.

Lohan S, Spielmann M, Doelken SC, Flottmann R, Muhammad F, Baig SM, Wajid M, Hulsemann W, Habenicht R, Kjaer KW, Patil SJ, Girisha KM, Abarca-Barriga HH, Mundlos S, Klopocki E. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. Clin Genet. 2014;86(4):318-25.

Lucas BA, Lavi E, Shiue L, Cho H, Katzman S, Miyoshi K, Siomi MC, Carmel L, Ares M, Jr., Maquat LE. Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay. Proc Natl Acad Sci USA. 2018;115(5):968-73.

Lunyak VV, Prefontaine GG, Nunez E, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, Garcia-Diaz A, Zhu XY, Yung Y, Montoliu L, Glass CK, Rosenfeld MG. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science. 2007;317(5835):248-51.

Lupiañez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161(5):1012-25.

Marcelis CL, Hol FA, Graham GE, Rieu PN, Kellermayer R, Meijer RP, Lugtenberg D, Scheffer H, van Bokhoven H, Brunner HG, de Brouwer AP. Genotype-phenotype correlations in MYCN-related Feingold syndrome. Hum Mutat. 2008;29(9):1125-32.

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007;445(7128):666-70.

Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, Murata T, Biffo S, Merrick WC, Darzynkiewicz E, Pillai RS, Filipowicz W, Duchaine TF, Sonenberg N. MicroRNA inhibition of translation initiation *in vitro* by targeting the cap-binding complex eIF4F. Science. 2007;317(5845):1764-7.

Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. 2018;32(1):42-57.

Mishra A, Hawkins RD. Three-dimensional genome architecture and emerging technologies: looping in disease. Genome Med. 2017;9(1):87.

Monahan K, Horta A, Lomvardas S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature. 2019;565(7740):448-53.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. A regulatory archipelago controls Hox genes transcription in digits. Cell. 2011;147(5):1132-45.

Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013;51(5):606-17.

Peng W, Deng W, Zhang J, Pei G, Rong Q, Zhu S. Long non-coding RNA ANCR suppresses bone formation of periodontal ligament stem cells via sponging miRNA-758. Biochem Biophys Res Commun. 2018;503(2):815-21.

Petit F, Sears KE, Ahituv N. Limb development: a paradigm of gene regulation. Nat Rev Genet 2017;18(4):245-58.

Rattanasopha S, Tongkobpetc.h S, Srichomthong C, Kitidumrongsook P, Suphapeetiporn K, Shotelersuk V. Absent expression of the osteoblast-specific maternally imprinted genes, DLX5 and DLX6, causes split hand/split foot malformation type I. J Med Genet. 2014;51(12):817-23.

Rinn JL, Chang HY. Genome regulation by long non-coding RNAs. Annu Rev Biochem. 2012;81:145-66.

Rogler LE, Kosmyna B, Moskowitz D, Bebawee R, Rahimzadeh J, Kutchko K, Laederach A, Notarangelo LD, Giliani S, Bouhassira E, Frenette P, Roy-Chowdhury J, Rogler CE. Small RNAs derived from lncRNA RNase MRP have gene-silencing activity relevant to human cartilage-hair hypoplasia. Hum Mol Genet. 2014;23(2):368-82.

Rouya C, Siddiqui N, Morita M, Duchaine TF, Fabian MR, Sonenberg N. Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA. 2014;20(9):1398-409.

Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development. 2005;132(4):797-803.

Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019;20(8):437-55.

Su W, Xu M, Chen X, Chen N, Gong J, Nie L, Li L, Li X, Zhang M, Zhou Q. Long non-coding RNA ZEB1-AS1 epigenetically regulates the expressions of ZEB1 and downstream molecules in prostate cancer. Mol Cancer. 2017;16(1):142.

Tayebi N, Jamsheer A, Flottmann R, Sowinska-Seidler A, Doelken SC, Oehl-Jaschkowitz B, Hulsemann W, Habenicht R, Klopocki E, Mundlos S, Spielmann M. Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. Orphanet J Rare Dis. 2014;9.

Tayebi N, Jamsheer A, Flottmann R, Sowinska-Seidler A, Doelken SC, Oehl-Jaschkowitz B, Hulsemann W, Habenicht R, Klopocki E, Mundlos S, Spielmann M. Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. Orphanet J Rare Dis. 2014;9:108.

Temtamy SA, Aglan MS. Brachydactyly. Orphanet J Rare Dis. 2008;3:15.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82.

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. The nuclear-retained non-coding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010;39(6):925-38.

Van der Meer JE, Afzal M, Alyas S, Haque S, Ahituv N, Malik S. A novel ZRS mutation in a Balochi tribal family with triphalangeal thumb, pre-axial polydactyly, post-axial polydactyly, and syndactyly. Am J Med Genet A. 2012;158A(8):2031-5.

van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell. 2017;169(5):780-91.

Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. 2015;43(Database issue):D174-80.

Watanabe T, Sato T, Amano T, Kawamura Y, Kawamura N, Kawaguchi H, Yamashita N, Kurihara H, Nakaoka T. Dnm3os, a non-coding RNA, is required for normal growth and skeletal development in mice. Dev Dyn. 2008;237(12):3738-48.

Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA. Developmentally regulated Shh expression is robust to TAD perturbations. Development. 2019;146(19):dev179523.

Yan P, Luo S, Lu JY, Shen X. Cis- and trans-acting lncRNAs in pluripotency and reprogramming. Curr Opin Genet Dev. 2017;46:170-78.

Yoon JH, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long non-coding RNAs. Semin Cell Dev Biol. 2014;34:9-14.

Yu L, Fang F, Lu S, Li X, Yang Y, Wang Z. lncRNA-HIT promotes cell proliferation of non-small cell lung cancer by association with E2F1. Cancer Gene Ther. 2017;24(5):221-26.

Yuksel-Apak M, Bogershausen N, Pawlik B, Li Y, Apak S, Uyguner O, Milz E, Nurnberg G, Karaman B, Gulgoren A, Grzeschik KH, Nurnberg P, Kayserili H, Wollnik B. A large duplication involving the IHH locus mimics acrocallosal syndrome. Eur J Hum Genet. 2012;20(6):639-44.

Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011;25(21):2227-41.

Zekri L, Kuzuoglu-Ozturk D, Izaurralde E. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. EMBO J. 2013;32(7):1052-65.

Zelzer E, Olsen BR. The genetic basis for skeletal diseases. Nature. 2003;423(6937):343-8.

Zhang XD, Huang GW, Xie YH, He JZ, Guo JC, Xu XE, Liao LD, Xie YM, Song YM, Li EM, Xu LY. The interaction of lncRNA EZR-AS1 with SMYD3 maintains overexpression of EZR in ESCC cells. Nucleic Acids Res. 2018;46(4):1793-809.

Zhao C, Sun W, Zhang P, Ling S, Li Y, Zhao D, Peng J, Wang A, Li Q, Song J, Wang C, Xu X, Xu Z, Zhong G, Han B, Chang YZ, Li Y. miR-214 promotes osteoclastogenesis by targeting Pten/PI3k/Akt pathway. RNA Biol. 2015;12(3):343-53.

Chapter 2

Scope of this thesis

The research in this PhD thesis involves different models: (1) the developing limb as *in vivo* model, and (2) neural differentiation of pluripotent stem cells as *in vitro* model, both for investigating the role of 3D chromatin structure in development, cell differentiation and cell identity, and of gene regulation, using Targeted Chromatin Capture (T2C). In addition, focus is on the characterization and functional analysis of *cis*-REs (enhancers) and *trans*-REs (TF Zeb2) in the regulation of expression of genes crucial for growth/differentiation signaling, in particular by the Wnt and the TGFβ/BMP families. The selection of the two aforementioned models was based on the expertise of my supervisors and my own experience, which originates from the research in the team of P. Tylzanowski at MUL and the team of D. Huylebroeck at Erasmus MC.

The focus of the first part of my PhD thesis, and perhaps the main part, is on the investigation of enhancers and 3D chromatin architecture during limb development, in particular during formation of synovial joint and phalange. Further, via expertise building in the fields of 3D chromatin architecture and enhancers, the aim was to learn about and contribute to mechanistic studies of Zeb2, the Mowat-Wilson Syndrome (MOWS) TF. The latter studies would be carried out in cultured ESCs, and (mainly) their neural and (also) mesendodermal differentiation, respectively.

The specific objectives of the respective projects were then:

(i) To produce a genome-wide candidate-enhancer atlas of the joint interzone and adjacent phalange, respectively. This work includes integrative analysis of transcriptomic data from bulk RNA-seq together with histone-3 H3K27ac and H3K4me1 signatures obtained by ChIP-seq (**Chapter 3**).

(ii) To establish a low-T2C protocol for studying cell populations or *in vivo* samples available as low cell numbers, ideally 100,000 cells or even less (**Chapter 4**).

(iii) To use low-T2C to investigate the *Dact2-Smoc2* genomic region, and identify and characterize genomic enhancers in the interzone during synovial joint development; these enhancers were also validated functionally at te MUL lab using a zebrafish enhancer assay (**Chapter 5**).

(iv) Similar to (iii), and using neural differentiating ESCs, contribute to the demonstration of dynamic DNA-loops in and around the human *ZEB2* locus (including its 3.5 Mb-long gene desert), and show co-operation between the newly identified enhancers, including in human neuroprogenitor cells (NPCs) (**Chapter 6**).

(v) To contribute to the mapping of genome-wide binding sites (by ChIP-seq) of Zeb2 for the first time in this field, in particular in NPCs, but also start with robust ESC differentiation (by adding BMP+Activin) towards mesendodermal cells for repeating this ChIP-seq in these. This work as a whole aims at identifying Zeb2-dependent and directly controlled target genes, as well as candidate Zeb2 co-operating TFs and/or other Zeb2 partner proteins, including those that act via these identified regions (**Chapter 7**).

The focus of the introductory chapter to this thesis manuscript (**Chapters 1a/b**) is therefore on selected mechanisms involved in the regulation of gene transcription, in particular on enhancers and their status, and chromatin architecture and its dynamics. **Chapter 1a** also briefly introduces the biological systems we worked on and their link with disease, the latter also being worked out in detail for limb development in **Chapter 1b**.

Rather than recapitulating in **Chapter 8** the discussion sections of the individual experimental chapters on each of these two large lines of my PhD research (synovial joint formation and Zeb2, respectively), the focus of this chapter is on the importance and perspectives of endured application of integrative omics to better understand global/general as well as selected/specific regulations, players and action modes in the first place, in the fields of synovial joint formation, and forebrain development and modeling of MOWS, respectively. Specific challenges for the enhancer and 3D chromatin architecture fields are also discussed.

Chapter 3

Identification of candidate enhancers controlling the transcriptome during the formation of interphalangeal joints

Karol Nowosad [1,2,3], Rutger W. W. Brouwer [1,4], Adrian Odrzywolski [5,6], Anne L. Korporaal [1], Bartłomiej Gielniewski [7], Bartosz Wojtaś [8], Wilfred F. J. van IJcken [1,4], Frank Grosveld [1], Danny Huylebroeck [1], Przemko Tylzanowski [2,9,*]

[1] Department of Cell Biology, Erasmus University Medical Center, 3015 CN Rotterdam, the Netherlands;
[2] Department of Biomedical Sciences, Laboratory of Molecular Genetics, Medical University of Lublin, Lublin, Poland;
[3] The Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland;
[4] Center for Biomics-Genomics, Erasmus University Medical Center, 3015 CN Rotterdam, the Netherlands;
[5] Department of Biochemistry and Molecular Biology, Medical University of Lublin, Lublin, Poland;
[6] Department of Human Genetics, KU Leuven, B-3000 Leuven, Belgium;
[7] Laboratory of Molecular Neurobiology, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, Warsaw, Poland;
[8] Laboratory of Sequencing, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, Warsaw, Poland;
[9] Department of Development and Regeneration, KU Leuven, B-3000 Leuven, Belgium

Published in

Sci Rep. 2022;12:12835

3.1. Summary

The formation of the synovial joint begins with the visible emergence of a stripe of densely packed mesenchymal cells located between distal ends of the developing skeletal anlagen called the interzone. Recently the transcriptome of the early synovial joint was reported. Knowledge about enhancers would complement these data and lead to a better understanding of the control of gene transcription at the onset of joint development. Using ChIP-sequencing we have mapped the H3-signatures H3K27ac and H3K4me1 to locate regulatory elements specific for the interzone and adjacent phalange, respectively. This one-stage atlas of candidate enhancers (CEs) was used to map the association between these respective joint tissue specific CEs and biological processes. Subsequently, integrative analysis of transcriptomic data and CEs identified new putative regulatory elements of genes expressed in interzone (e.g., *GDF5*, *BMP2* and *DACT2*) and phalange (e.g., *MATN1*, *HAPLN1* and *SNAI1*). We also linked such CEs to genes known as crucial in synovial joint hypermobility and osteoarthritis, as well as phalange malformations. These analyses show that the CE atlas can serve as resource for identifying, and as starting point for experimentally validating, putative disease-causing genomic regulatory regions in patients with synovial joint dysfunctions and/or phalange disorders, and enhancer-controlled synovial joint and phalange formation.

3.2. Introduction

Synovial joints, organs present at the articular ends of long bones, are essential for vertebrate mobility. They comprise of articular cartilage, synovium, ligaments and the synovium capsule (Khan *et al.*, 2007). Due to their function, the joints are frequently exposed to mechanical stress and thus prone to injuries. Congenital malformations and a number of diseases affect joint structure, thereby causing a decrease of joint functionality. For instance, misexpression of *PITX1* caused by enhancer adoption results in dysplastic elbow joints in Liebenberg syndrome (OMIM #186550; Spielmann *et al.*, 2012), a homozygous mutation in *IMPAD1* leads to chondrodysplasia with joint dislocations (OMIM #614078; Vissers *et al.*, 2011), and loss of *EXOC6B* causes joint dislocations and defects in joint mobility, characteristic for patients with spondylo-epimetaphyseal dysplasia with joint laxity, type 3 (OMIM #618395; Girisha *et al.*, 2016). Osteoarthritis (OA) is the most prevalent synovial joint disease affecting adults (Boer *et al.*, 2021). Typical in OA is the progressive degeneration of articular cartilage and accompanying subchondral bone sclerosis, joint space narrowing, osteophyte formation and the variable degree of synovium inflammation (Chen *et al.*, 2017), eventually causing joint destruction. The latter frequently needs intervention by joint replacement (Dieppe *et al.*, 2011).

A comprehensive understanding of gene regulatory networks (GRNs) orchestrating synovial joint formation will contribute to the understanding of both healthy and pathological processes taking place in this organ. This knowledge will also help in the development of novel cell and/or gene-based strategies for treatment of injured articular cartilage within the developmental engineering paradigm (Roelofs *et al.*, 2013). The first morphologically distinguishable event in joint development is the formation of interzones, with distinct progenitor cells giving rise to the majority of articular tissues (Decker *et al.*, 2014). The condensing cartilage anlage, at the locations of future joints, undergoes several rounds of cell proliferation

(Sagnoli *et al.*, 2007). The discovery that the influx of cells from the outside of the interzone contributes to overall increased interzone cell density, points to an important mechanism in interzone formation (Ray *et al.*, 2015; Schwartz *et al.*, 2016). At the same time, *SOX9* (a member of Sry family of transcription factors, TFs) expression becomes repressed, arresting the chondrogenic program and allowing the interzone to form (Luyten *et al.*, 2009). The formed joint interzone comprises of two layers of cells, named the outer and intermediate interzone layers, with differentially expressed genes (DEGs), including *COL2A1* (encoding α(II)-collagen) and *MATN1* (Matrilin-1), which have higher expression in the outer interzone, and *GDF5* (Growth and Differentiation Factor-5, a ligand of the BMP subgroup of the TGFβ family) with higher mRNA level in the intermediate interzone (Jenner *et al.*, 2014). These layers will contact the ends of future bones, while an inner cell layer of yet to be defined function is also present.

Cells within the cartilage anlage change their phenotype progressively from round to columnar, pre- and eventually hypertrophic chondrocytes, contributing to longitudinal cartilage and bone growth. Acknowledged molecular markers for the round chondrocytes include the aforementioned *MATN1* and *COL2A1* (Hyde *et al.*, 2007). Additionally, chondrogenic induction and differentiation is accompanied by the expression of *RUNX2* (encoding a RUNT family TF) followed by the expression of *COL10A1* (α(X)-collagen). The latter is a specific marker gene for hypertrophic chondrocytes, and some of its enhancers have been mapped (Gu *et al.*, 2014). It has been suggested that the round chondrocytes may contribute to the articular cartilage, but their contribution to the interzone structures remains unclear and may depend on restricted exposure to BMP and/or WNT signals (Ray *et al.*, 2015). Thus, while new knowledge is emerging regarding early stages of joint formation, including interaction with adjacent cartilage, the insight into gene expression control within cells of the joint interzone remains incomplete. Indeed, while numerous reports have described the transcriptome in the synovial joint formation (Jenner *et al.*, 2014; Pazin *et al.*, 2014), including at single-cell level (Bian *et al.*, 2020), relatively little is known about the activity of enhancers during that process.

Enhancers regulate gene transcription mainly in *cis*, within the topologically-associating domains (TADs), where they promote intra-TAD control of transcription of loci by making TFs and co-factors bridge between their bound distal enhancer sites and the promoter-proximal region of the appropriate target gene(s), hence achieving physical proximity (Lupiañez *et al.*, 2015; Fudenberg *et al.*, 2016; Garcia-Gonzalez *et al.*, 2016). Enhancers are associated with histone modification signatures, such as H3K27ac and H3K4me1, and chromatin accessibility, however such biochemical marks may be absent in the so-called hidden enhancers at one or more stages of cell differentiation. In the enhancer-promoter complexes high-affinity binding of co-factors (e.g. histone-acetylation containing P300), TFs and RNAPol2 also can be shown (Calo and Wysocka, 2013). Defects in enhancer function have been linked to limb malformations as well, for example in Hass-type polysyndactyly (OMIM #186200; Lohan *et al.*, 2014), split hand/foot malformation (OMIM #183600; Tayebi *et al.*, 2014), Leri-Weill dyschondrosteosis (OMIM #127300) and Laurin-Sandrow syndrome (OMIM #135750) (Bunyan *et al.*, 2016).

Here, our focus was on identifying CEs based on biochemical H3-profiles active in the joint interzone and adjacent phalange. Next, we developed an atlas of candidate *cis*-regulatory elements at one developmental stage of chick embryos. In combination with the available transcriptomes, this atlas will help in elucidating the molecular mechanisms that control joint interzone formation and/or cause joint disease. We opted for microsurgical dissection of interzone as opposed to using *GDF5*-positive (+) cell selection procedures, because not all cells during early stages of joint formation are convincingly *GDF5*+ (Schwartz *et al.*, 2016; Bian *et al.*, 2020). We identified unique

interzone/phalange CEs that are conserved between chicken, mouse and human, and functionally annotated these CEs, followed by integrative analysis of cell-type specific CEs and DEGs. We also associated the CEs with synovial joint and phalange abnormalities, and a higher risk of OA.

3.3. Results

3.3.1. Microdissection of joint interzones and phalanges

The interzones and the adjacent proximal part of phalange were dissected from the third digit of the hindlimb of chick embryos (at stage HH32, when the interzone was distinguishable under the microscope; **Fig. 3.1a**), and RNA-sequencing (RNA-seq) was performed on the separated tissues (*see* Experimental procedures). Subsequently, we analyzed the interzone and phalange RNA-seq datasets using DESeq2. First, we checked whether biological replicates separated according to origin of the tissue. For this, we performed principal component analysis (PCA) revealing that component 1 (PC1) indeed separates interzone from phalange (**Fig. S3.1a**). Next, to ensure that our datasets fit the DESeq2 model we analyzed the dispersion estimates, showing that our data generates typical pattern of dispersion plot (**Fig. S3.1b**). Importantly, the curve presented at this plot has low dispersion values for high mean values of normalized counts, and high dispersion values for low mean values of normalized counts, which presents a general relationship between dispersion and gene expression for datasets fitting the DESeq2 model.

Next, we performed differential gene expression analysis (DEA) to identify key gene players specific either for interzone or adjacent phalange. The DEA with log2FC > 0.5 and p.adj < 0.05 identified 116 upregulated genes in interzone, and 61 genes upregulated in phalange (**Table S3.1**; **Fig. S3.1c**). Importantly, this analysis confirmed that the interzone samples had increased mRNA levels of *GDF5*, *ENPP2, COL3A1* and *ERG*, each already known to be expressed in joint interzones. Also, the DEA showed that phalange samples had significantly higher expression of well-described chondrocyte markers, such as *COL2A1*, *MATN1*, *SNAI1* and *RUNX2* (**Fig. 3.1b**; **Table S3.1**). Notably, *COL10A1* mRNA expression was not detected in phalange samples, supporting the notion that the collected phalange regions contain chondrocytes prior to hypertrophy, indicating an early stage of the limb development. Also, interzone samples were not expressing *COL10A1* (except one replicate with ultra-low count numbers, i.e. equal to 1.9, suggesting that this is an artefact, and not the product of gene expression) excluding potential contamination by hypertrophic chondrocytes (**Table S3.2**).

To further validate differential expression of selected interzone marker genes (i.e. *GDF5*, *ENPP2*, *ERG*) we performed RT-qPCR (**Fig. 3.1c**), which confirmed our RNA-seq data. Next, we compared gene expression of our identified DEGs with available single-cell (sc)RNA-seq data from the atlas of synovial joint development (Bian *et al.*, 2020). For this purpose, we focused on the Super Cluster 2 (SC2), shown to be composed of two populations: SC2_A, expressing chondrocyte related genes, and SC2_B, expressing interzone markers. Subsequently, we analyzed the subset of 60 genes due to lack of information about the other DEGs in scRNA-seq data. Many of the DEGs identified here presented high expression either in SC2_A, or SC2_B (**Figs. S3.2-S3.4**). Collectively, these results show that the we successfully dissected tissues of interest.

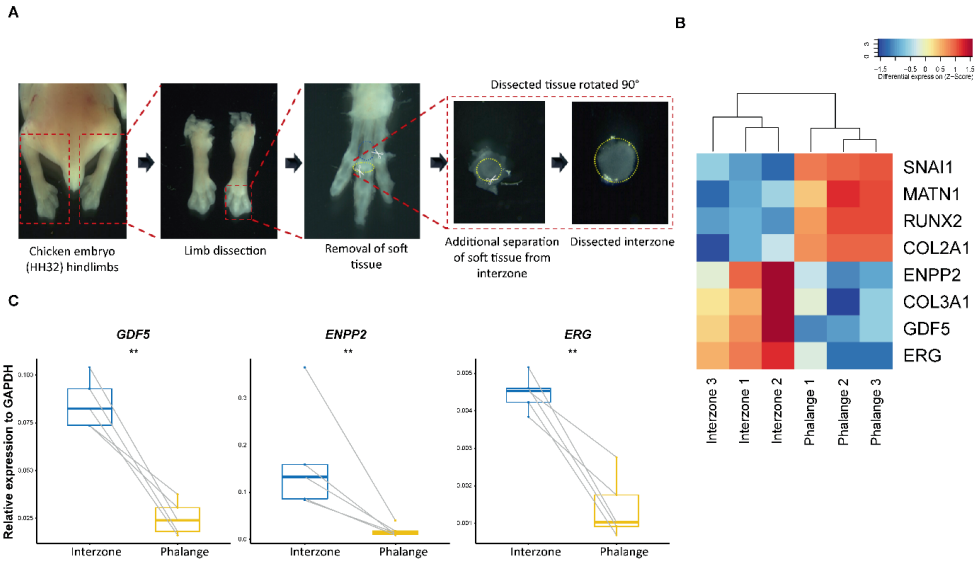


Figure 3.1. Dissection and transcriptome profiling of joint interzone and phalanges.

(a) Location and dissection of the interzone and adjacent proximal part of phalange from hindlimb digit 3 of chicken embryo (HH32). The dissection procedure includes separation of hindlimbs, removal of soft tissue from the digits, and subsequent separation of interzone from adjacent phalange. **(b)** Differences in expression of interzone (*ENPP2, COL3A1, GDF5*, *ERG*) and phalange (*SNAI1,* MATN1, *RUNX2, COL2A1*) marker genes based on the RNA-seq data. **(c)** mRNA steady-state level of selected interzone markers (*GDF5*, *ENPP2*, *ERG*) as determined by RT-qPCR (all data were normalized to the expression of *GAPDH*; lines combined the samples isolated from the same embryo; *$p < 0.05$; **$p < 0.01$ based on Mann-Whitney-Wilcoxon test).

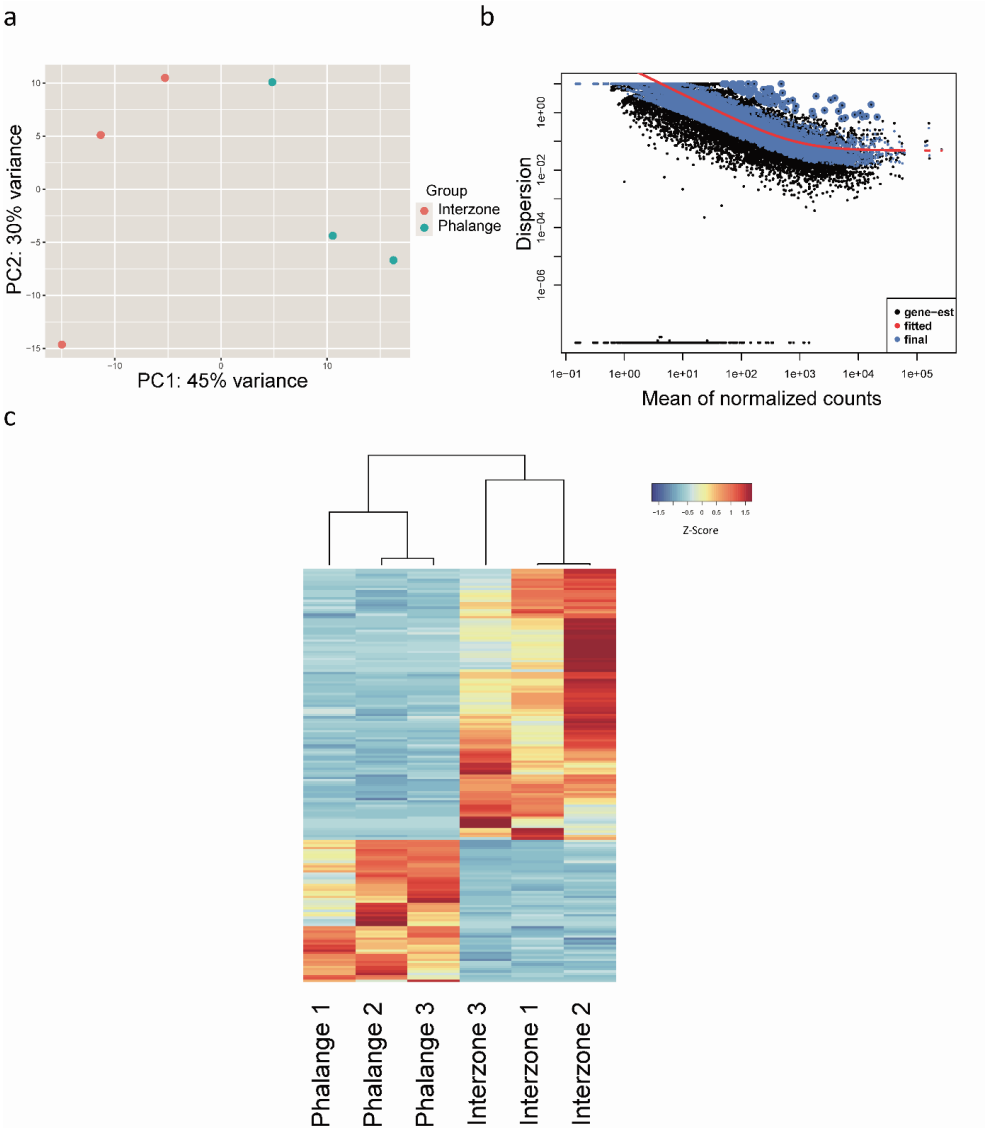


Figure S3.1. Analysis of RNA-seq data from isolated joint interzone and adjacent phalange.

(a) Principal component analysis (PCA) showing separation of analyzed interzone and phalange biological replicates based on the origin of the tissue. **(b)** Dispersion plot presenting the relationship between dispersion and gene expression. **(c)** Heatmap of all differentially expressed genes.

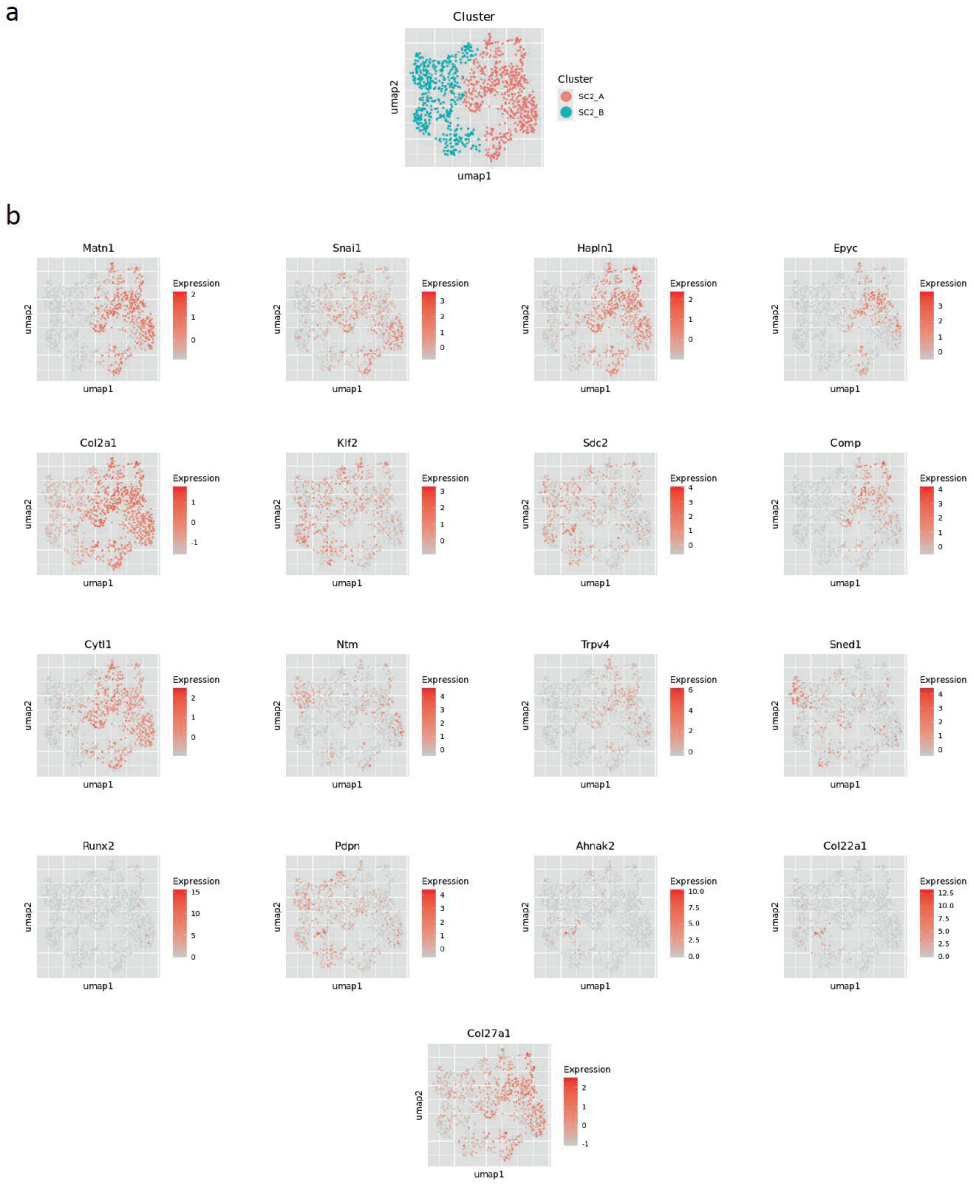


Figure S3.2. Analysis of gene expression of DEGs upregulated in phalange using data from a single cell transcriptional atlas of early synovial joint development.

(a) A UMAP with annotated clusters SC2_A and SC2_B. The SC2_A was characterized by Bian *et al.* (2020) as a cluster containing cells with high expression of chondrocyte markers. The SC2_B was described as a cluster containing cells with high expression of interzone markers. **(b)** Analysis of DEGs with significantly higher expression in phalange as compared to interzone.

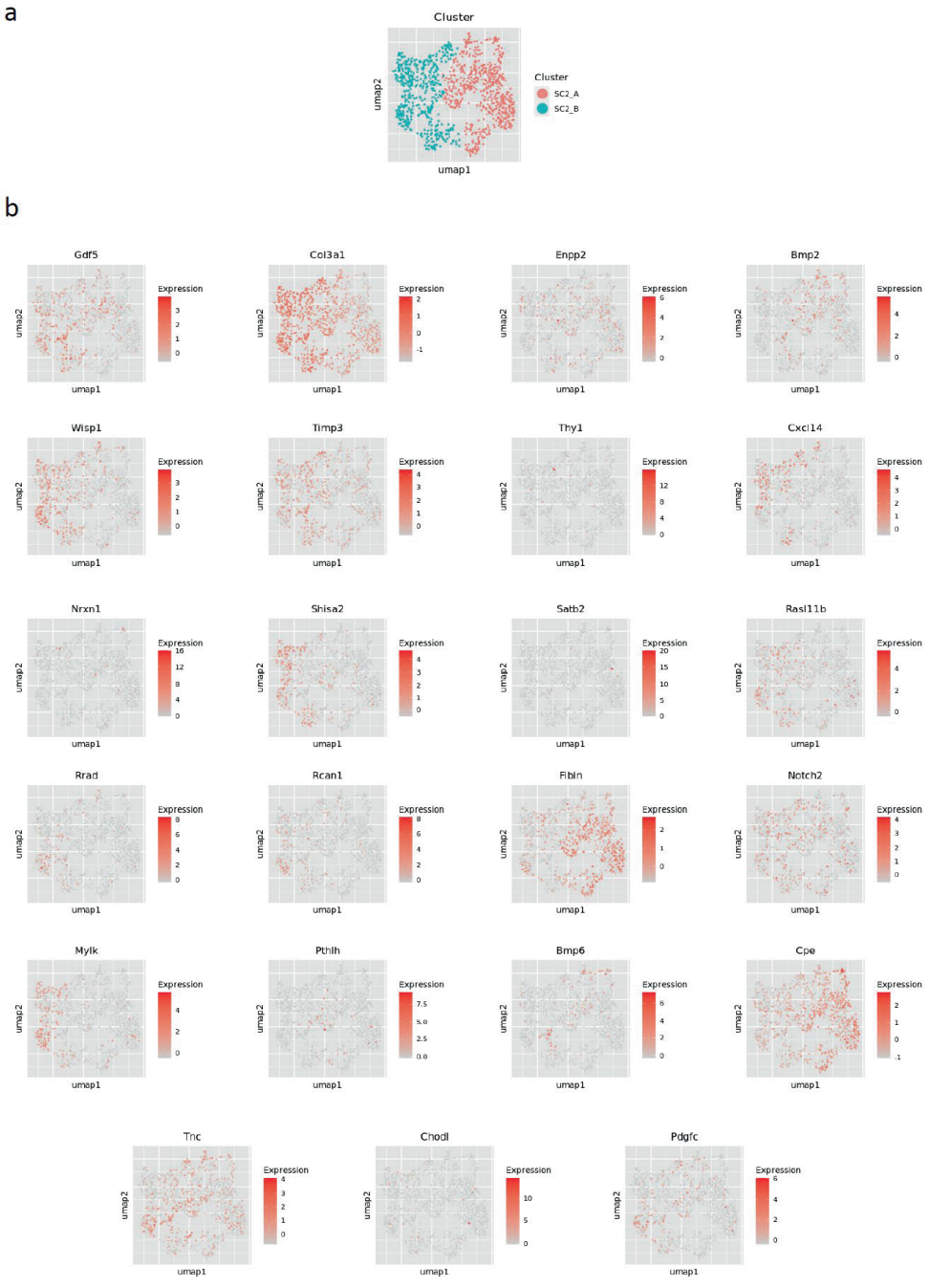


Figure S3.3. Analysis of gene expression of DEGs upregulated in interzone using data from a single cell transcriptional atlas of early synovial joint development.

(a) A UMAP with marked clusters SC2_A and SC2_B. The SC2_A and SC2_B contain cells as described in **Fig. S3.2**. **(b)** Analysis of DEGs with significantly higher expression in interzone as compared to phalange.

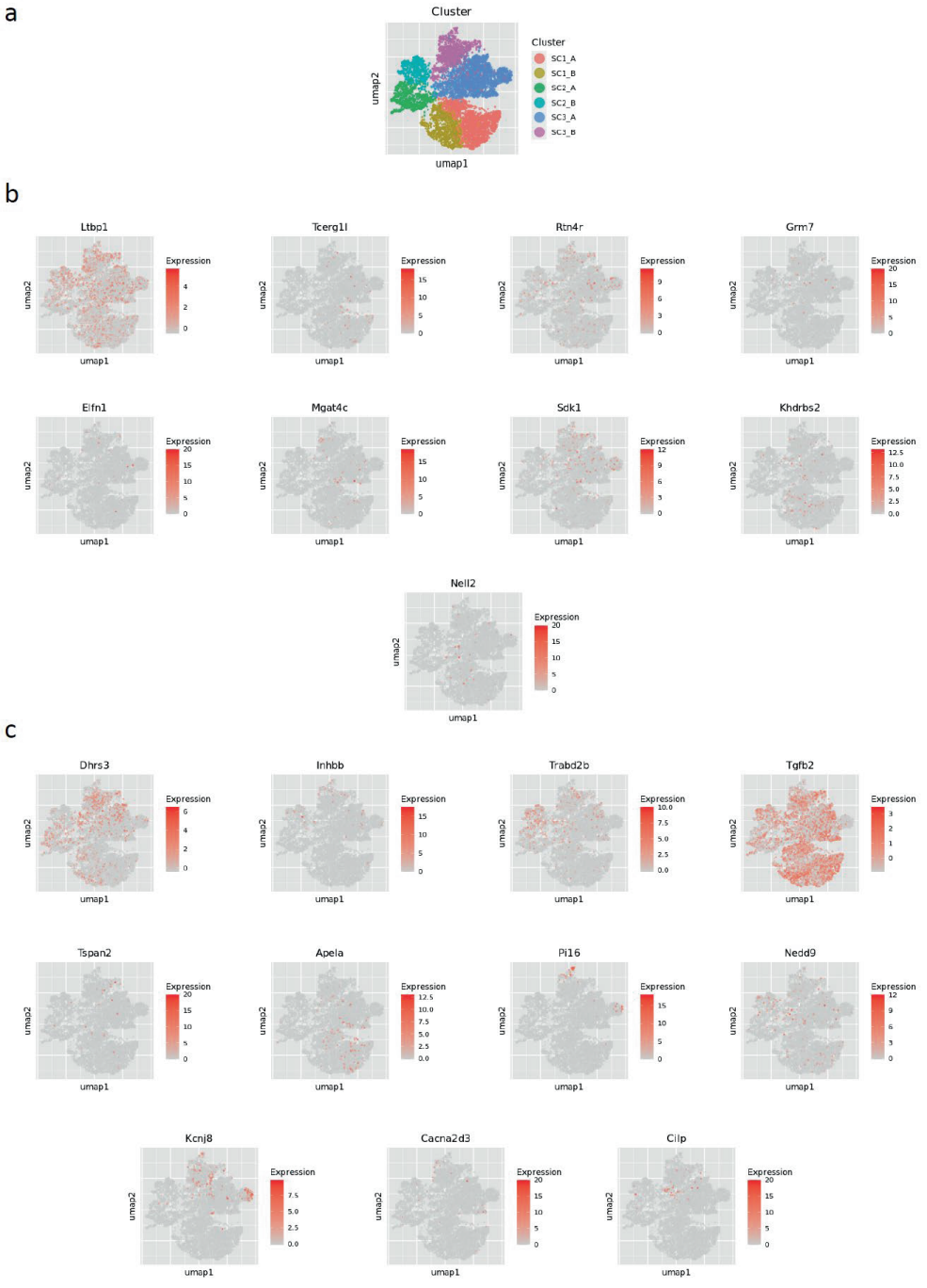


Figure S3.4. Analysis of gene expression of DEGs upregulated in either phalange or interzone using data from a single cell transcriptional atlas of early synovial joint development.

(a) A UMAP with annotated clusters SC1-SC3. The SC2_A and SC2_B contain cells marked by the phalange and interzone markers, respectively. **(b)** Analysis of DEGs upregulated in interzone. **(c)** Analysis of DEGs upregulated in phalange.

3.3.2. Atlas of putative enhancers of joint interzone and phalange identifies candidate enhancers involved in the regulation of cell identity

Next, we mapped the global H3K27ac and H3K4me1 signatures of joint and phalange by chromatin immunoprecipitation followed by Next-Generation Sequencing (ChIP-seq). The unsupervised clustering analysis revealed that interzone and phalange have distinct profiles for both H3K27ac and H3K4me1 (**Fig. 3.2 a,b**).

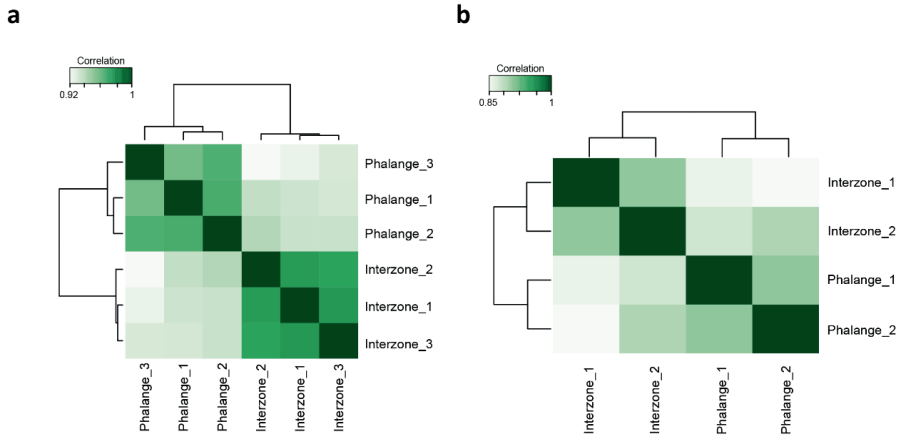


Figure 3.2. Clustering of genome-wide profiles of histone modifications in prepared interzone and phalange, in particular based on the mapping of H3K27ac and H4K4me1 genomic regions using ChIP-seq.

(a) Heatmap presenting correlation of interzone and phalange, based on the detection of H3K27ac. **(b)** Heatmap with correlation of interzone and phalange H3Kme1 mapped signatures.

The mapping of the regions enriched for H3 modifications enabled us to generate a joint/phalange CE atlas and regions enriched for both H3K27ac and H3K4me1 were denoted strongly-active enhancers, for H3K27ac active enhancers, and for H3K4me1 poised enhancers. Since enhancers are often evolutionarily conserved (Fish *et al.*, 2017), we decided to select only conserved regions among chick, mouse and human followed by merging nearby genomic intervals (for details, *see* Experimental procedures). Merging of these conserved regions reduces the probability of CEs separation into multiple short sequences and empowers the analysis, however, may lead to the generation of CEs containing multiple conserved regions separated by non-conserved genomic blocks. Using this approach, we identified 14,217 strongly-active, 5,479 active and 11,913 poised enhancers in the interzone, and 14,224; 6,041 and 12,997, respectively, in the phalange (**Table S3.3**; for their frequency and similar ratios in both samples, *see* **Fig. 3.3a**). For functional annotation of these CEs we used GREAT, which extracts gene ontology (GO) terms linked to biological processes (McLean *et al.*, 2010). The denoted strongly-active enhancers associated with cartilage and skeletal development (**Fig. 3.3b**), whereas active and poised enhancers mostly linked to general cell functions or processes not specific for skeletal development (**Fig. S3.5**). These results suggest that, among all CEs, only the strongly-active enhancers are associated with genes involved in regulation of processes crucial for limb development. Therefore, we hypothesized that strongly-active enhancers play important role in cell-type specific biological processes.

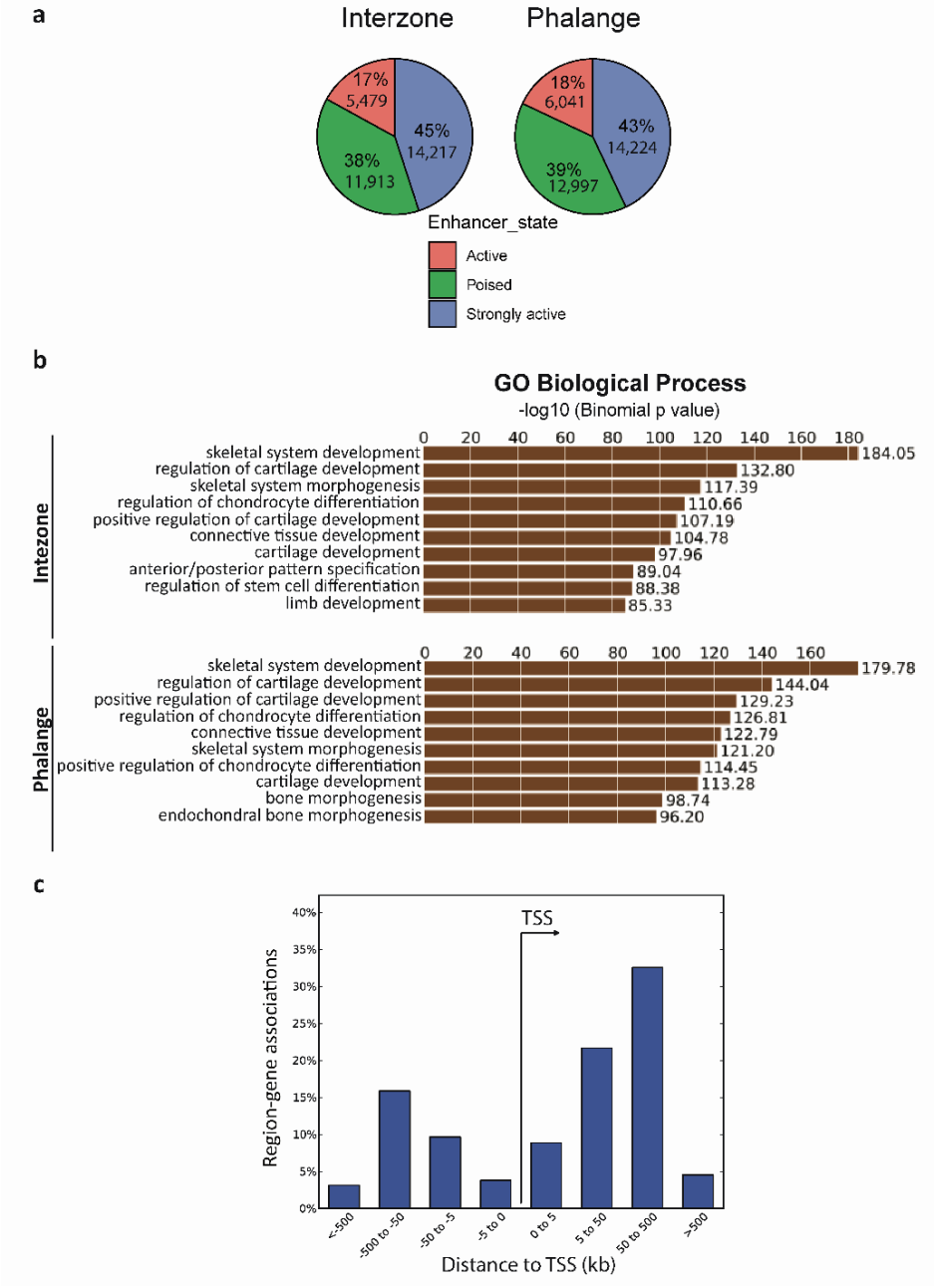


Figure 3.3. Characterization of the strongly-active CEs.

(a) Distribution of strongly-active, active and poised candidate enhancers in the interzone and phalange enhancer atlas. **(b)** Biological process GO terms associated with the strongly-active CEs. The functional annotation of CEs was again carried out using GREAT. **(c)** Genomic localization of strongly-active CEs in relation to nearby transcription start site (TSS). The CEs – target gene(s) association and calculation of the distances from the TSS was performed using GREAT.

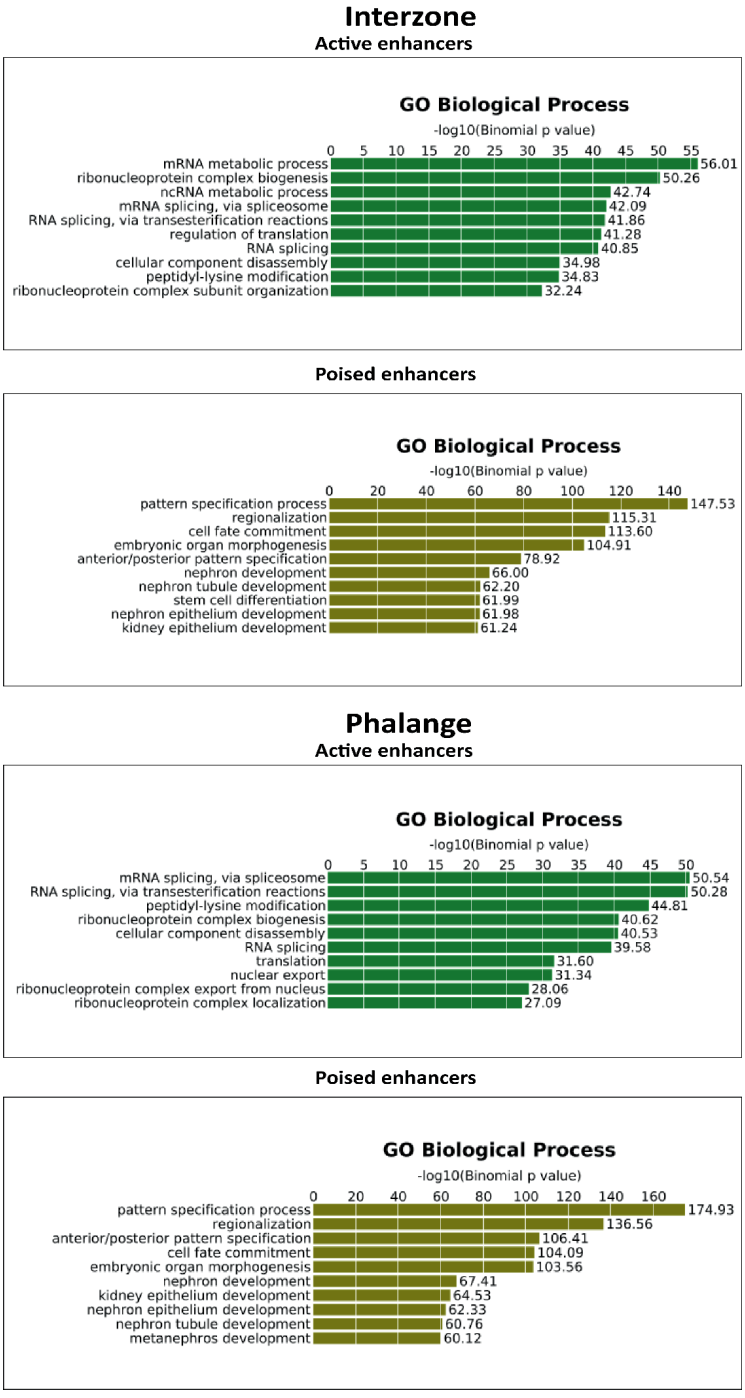


Figure S3.5. Functional annotation of the candidate enhancers (CEs) in interzone and phalange.

The CEs were associated with target genes using GREAT, followed by pathway enrichment analysis. Biological process GO terms associated with active and poised enhancers in interzone and phalange, respectively, are represented.

Association of strongly-active candidate enhancers with cell-lineage specific processes prompted us to focus on strongly-active CEs. These CEs are typically located >5 kb away from the respective transcription start site (TSS) (**Fig. 3.3c**). We could confirm characterized enhancers of well-studied loci (**Fig. 3.4**), specifically those expressed in interzone (e.g., *GDF5*; Chen *et al.*, 2016) or chondrocytes (e.g., phalangeal *IHH*, *SOX9, ACAN*; Hu *et al.*, 2012; Yao *et al.*, 2015; Will *et al.*, 2017). In parallel, and further validating our *in silico* selection approach, we extended the analysis by using Vista Enhancer Browser dataset (Visel *et al.*, 2007), leading to identification of 257 enhancers (**Table S3.4**), which have been functionally validated during embryogenesis. The enhancers from the Vista Enhancer Browser dataset were tested at E11.5 in mouse. Using our atlas, we showed that these 257 enhancers also present marks of active enhancers at later developmental stage (HH32, an equivalent of E14.5 in mouse) and also present conserved activity in chicken. A majority of them (203/257) have been defined as strongly-active CEs in both interzone and phalange (for illustration of 6 of these, *see* **Fig. S3.6**).

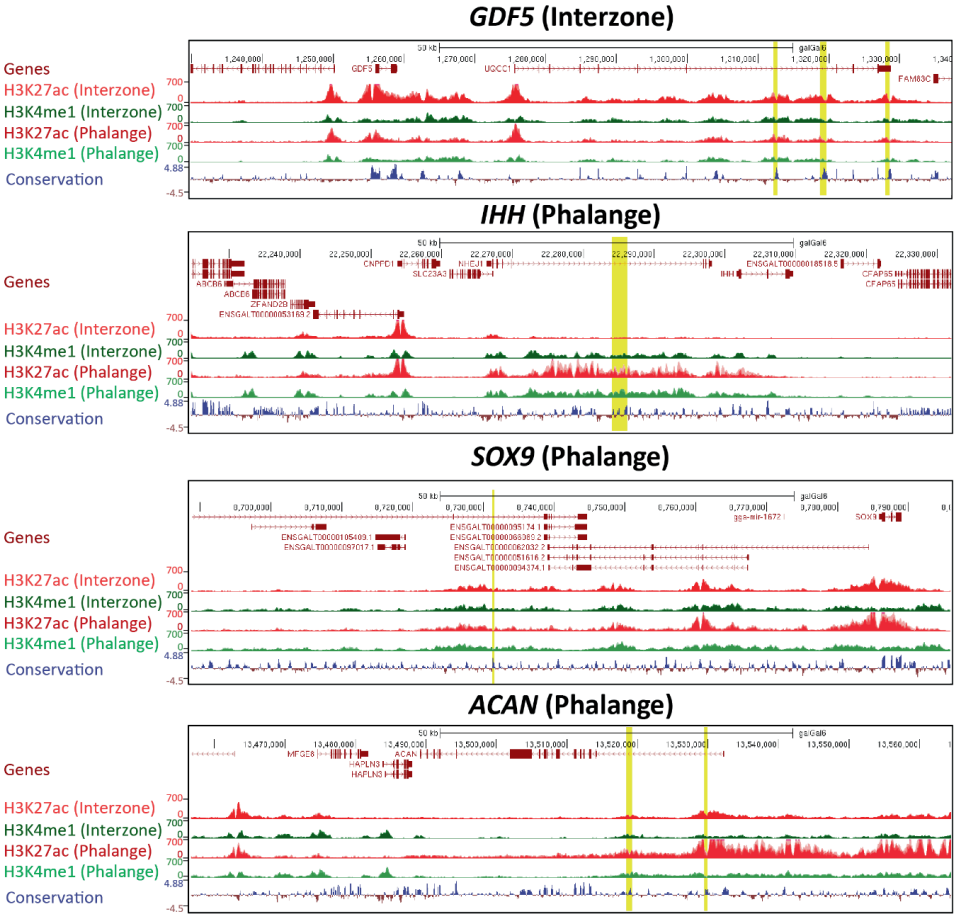
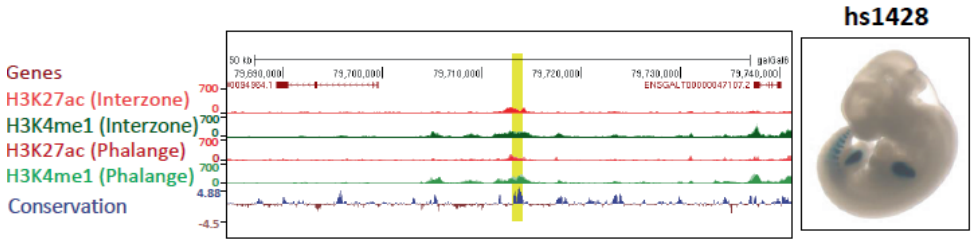


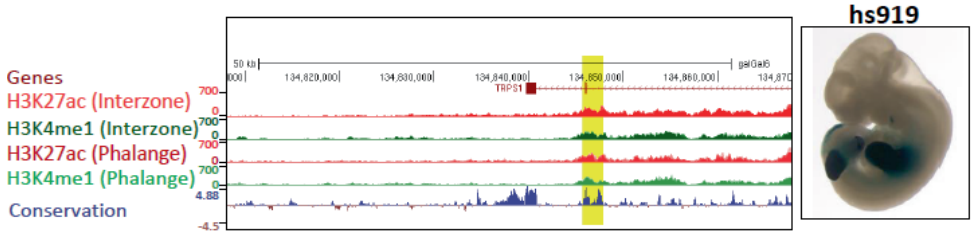
Figure 3.4. Examples of functionally validated enhancers characterized in the literature.

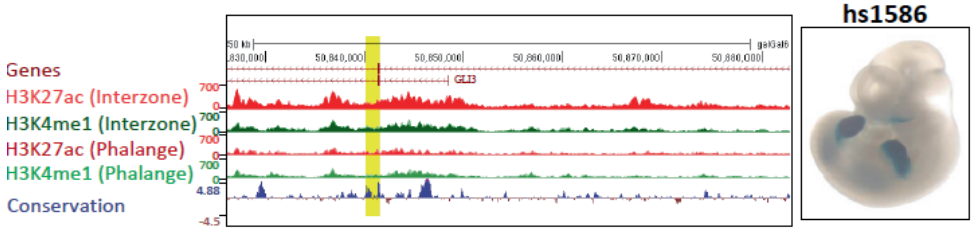
For the typical four loci shown, the H3K27ac enrichment track is marked in red, the H3K4me1 enrichment in green, and the conservation track in blue (together with gene structure information in brown). Regions marked by yellow present functionally validated enhancers described in Hu *et al.* (2012), Yao *et al.* (2015), Chen *et al.* (2016), and Will *et al.* (2017). The same enhancers have been identified as strongly-active enhancers in our Enhancer Atlas.

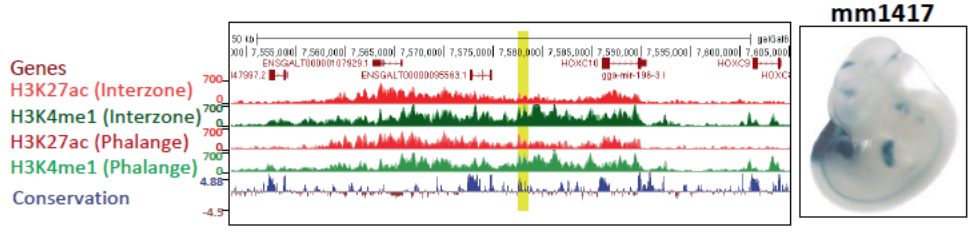
Figure S3.6. (*next page*). Validated active enhancers identified in our CE atlas.

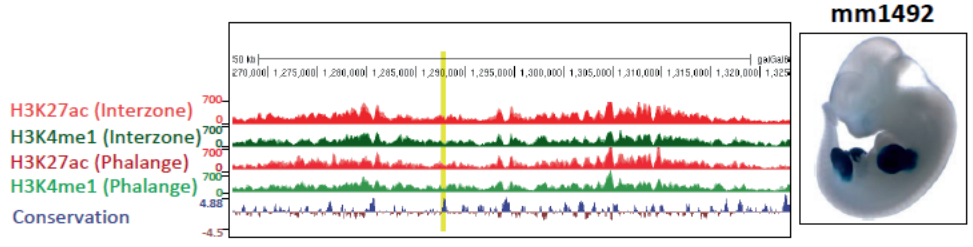
Examples of experimentally validated enhancers with limb activity (collected from Vista Enhancer Browser). H3K27ac enrichment track is colored in red, the H3K4me1 enrichment in green, and the conservation track in blue. The gene structure information is marked in brown. Regions marked by yellow present functionally validated enhancers. Activity of these enhancers is shown by the β-galactosidase (blue) staining present in embryonic regions.

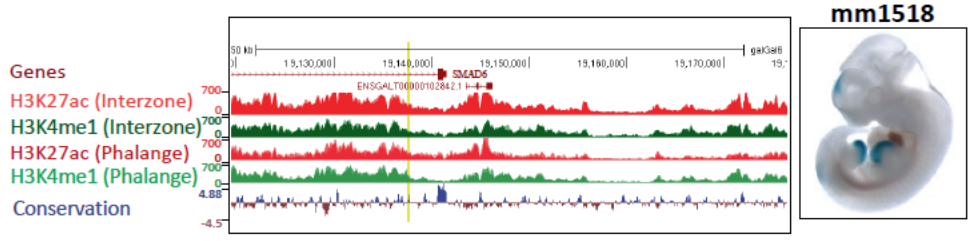

hs1428
Genes
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation


hs919
Genes
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation


hs1586
Genes
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation


mm1417
Genes
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation


mm1492
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation


mm1518
Genes
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation

To further characterize tissue-specific strongly-active CEs in the developing interzone and phalange, we selected the mutually exclusive strongly-active enhancers, yielding 3,406 CEs (out of the aforementioned 14,217 in total) unique for interzone and 3,407 (out of 14,224) for phalange (**Table S3.5; Fig. 3.5**). GREAT linked many of such interzone-specific CEs to mesenchymal cell differentiation, and regulation of transmembrane receptor protein serine/threonine kinase signaling (**Fig. S3.7a**, top panel). In contrast, CEs exclusive for phalange retrieved GO terms including chondrocyte differentiation and endochondral bone morphogenesis (**Fig. S3.7b**, top panel).

Next, we investigated whether the change of enhancer state from strongly-active to poised would be relevant for the regulation of tissue-specific genes. Indeed, 2,111 changes occurred with strongly-active CEs (out of the aforementioned identified 14,217 in total) in interzone and were found poised in phalange; in comparison, 1,502 changes occurred with strongly-active CEs (out of 14,224 in total) in phalange and were found poised in interzone (**Fig. 3.5**; **Table S3.6**). Strikingly, the interzone strongly-active enhancers that are poised in phalange were found to associate with regulation of transmembrane receptor protein serine/threonine kinase signaling (**Fig. S3.7a**; bottom panel), which is consistent with our functional annotation of interzone-specific CEs. In contrast, phalange strongly-active enhancers that are poised in interzone linked to positive regulation of cartilage differentiation (**Fig. S3.7b**; bottom panel).

Figure 3.5. Analysis of unique CEs and CEs with different enhancer states.

Visualization of all CEs with a subset of unique CEs and a subset of CEs which change the state between tissues.

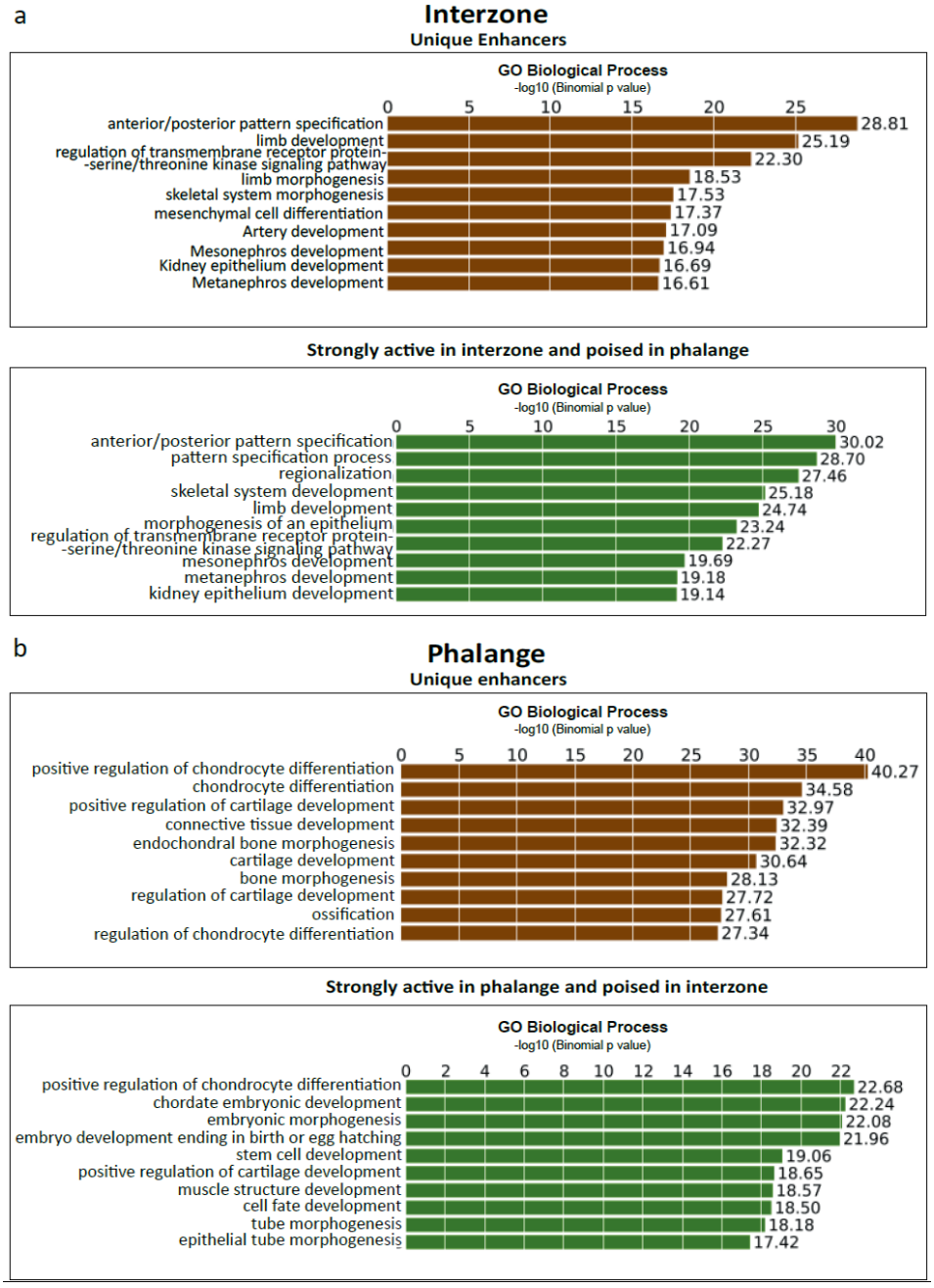


Figure S3.7. Mutually exclusive strongly-active enhancers in biological processes in interzone and phalange.

(a) GO terms associated with interzone-exclusive strongly-active enhancers, and strongly active enhancers in interzone that are poised in phalange, and **(b)** vice versa.

3.3.3. Transcription factor binding at strongly active CEs of developing interzone and phalange

Enhancers contain multiple TF binding-sites (TFBSs), involved in enhancer activation. The formation of TFBSs into clusters within enhancer regions allows enhancers to be bound by a set of tissue-specific TFs, and consequently be regulated in a spatio-temporal manner. Therefore, to predict the binding of TFs with function/s in limb development (including synovial joint development), we performed motif enrichment analysis of interzone and phalange strongly-active CEs identified in this study.

The analysis of TFBSs using HOMER (Heinz *et al.*, 2010) revealed that strongly-active CEs from both tissues are indeed enriched in TFs motifs pivotal for limb development (**Tables S3.7 and S3.8**). A majority of identified motifs was common for both interzone and phalange CEs. Importantly, we identified that CEs were enriched in the motif belonging to P300, a co-factor associated with active enhancers. There was also a match to PITX1, a TF crucial for hindlimb identity (Infante *et al.*, 2013), as well as a match to *ERG*, a TF involved in synovial joint formation (Iwamoto *et al.*, 2007) and OA susceptibility (Ohta *et al.*, 2015), and to a motif of *HOXD13*, a TF important for phalange formation and disease (Brison *et al.*, 2012).

Next, we characterized tissue-specific TFs, which may regulate interzone and phalange CEs and be responsible for establishment of cell identity. For this purpose, we performed motif enrichment analysis using mutually exclusive strongly-active enhancers CEs specific for interzone and phalange (**Tables S3.9 and S3.10**). For instance, we identified that phalange-specific CEs contain exclusively enrichment of *RUNX2* motifs in comparison to interzone-unique CEs. Importantly, *RUNX2* was identified as differentially expressed gene (log2FC = 2.5; p.adj = 4.5e-23) in our interzone and phalange RNA-seq datasets (**Fig. 3.1b**). Collectively, the motif analysis showed that identified CEs are enriched in the motifs of TFs, which are important for development of synovial joint and phalange, as well as matching with our RNA-seq data also.

3.3.4. Integrative analysis of DEGs and CEs

To investigate if there was a correlation between gene transcription and CEs, we superimposed our RNA-seq and ChIP-seq data. Pathway enrichment analysis of DEGs showed that genes upregulated in interzone again linked to transmembrane receptor protein serine/threonine kinase signaling (**Fig. 3.6a**), in line with our preceding annotation of interzone-exclusive CEs (*see above*). In particular *RASL11B, LTBP1, TGFB2, GDF5, FSTL1, BMP2, DACT2, CCN3, BMP6, CILP, INHBB* and *BMPR2* were found upregulated in interzone as compared to phalange (**Fig. 3.6b**). Similarly, analysis of genes upregulated in phalange linked these to chondrocyte differentiation and also endochondral bone morphogenesis (**Fig. 3.6c**), which is consistent with functional annotation of phalange-specific CEs. The genes involved it these two latter processes are *RUNX2, COL2A1, TRPV4, COL27A1, MATN1, COMP* and *CYTL1* (**Fig. 3.6d**).

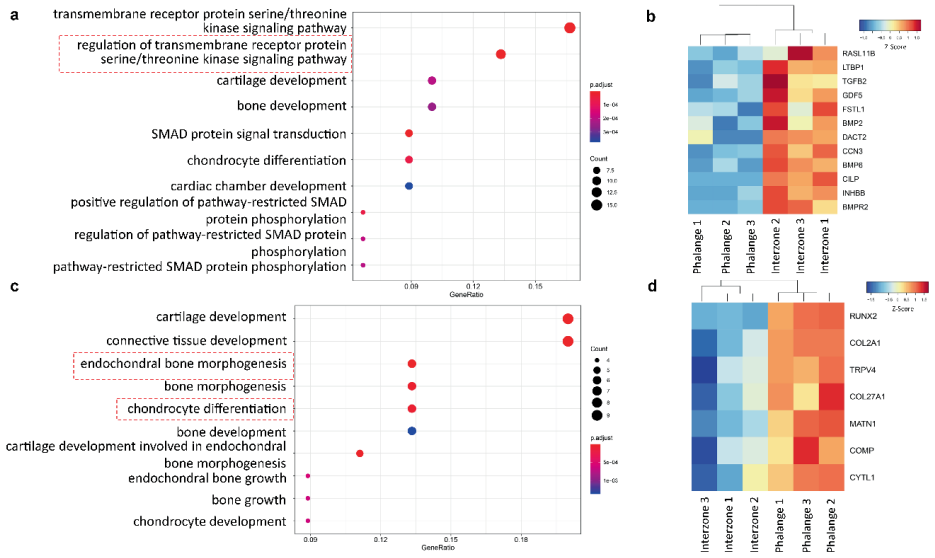


Figure 3.6. Analysis of the DEGs.

(a) Pathway enrichment analysis of genes significantly upregulated in interzone as compared to phalange. Color of the dot presents p.adj; size of the dot marks the number of genes involved in the pathway. The red rectangle marks pathway shown as enriched in the analysis of interzone-specific CEs as well as DEGs. **(b)** Heatmap presenting the differences in expression level of genes involved in the pathway: regulation of transmembrane receptor protein serine/threonine kinase signaling. **(c)** Pathway enrichment analysis of phalange upregulated DEGs. The red rectangle highlights pathways identified as enriched in for phalange specific CEs as well as DEGs. Color and size of the dot are as described above. **(d)** Heatmap presenting the expression level between interzone and phalange for genes involved in the pathways: endochondral bone morphogenesis and chondrocyte differentiation, respectively.

Next, in order to identify the CEs that control DEGs upregulated in interzone, the *cis*-regulatory landscapes were characterized by annotation of the TADs encompassing such genes. For this purpose, we used available chicken fibroblast Hi-C data (Fishman *et al.*, 2019). Next, we extracted CEs located within these TADs, and associated them with DEGs. The CEs were mapped to all DEGs located within the same TAD. Therefore, a CE can be associated with more than one gene, which is in line with studies showing that enhancers can indeed regulate more than one target gene. If the DEGs were located within the region not annotated by any TAD (TADs do not cover the entire chicken genome, as shown by Fishman *et al.*, 2019), we associated the CEs located -/+ 1 Mb from the TSS. Using this approach, we identified 486 interzone-specific CEs (**Table S3.11**; examples of enhancer analysis are given in **Fig. 3.7a**).

We then performed the same analysis for phalange-upregulated DEGs. This resulted in identification of 333 phalange-specific CEs (**Table S3.12**; with examples given in **Fig. 3.7b**). Collectively, the integrative analysis of transcriptome data with CEs assignment, and considering interzone vs. phalange signatures, showed that the DEGs involved in cell type-specific processes are regulated by cell-specific CEs.

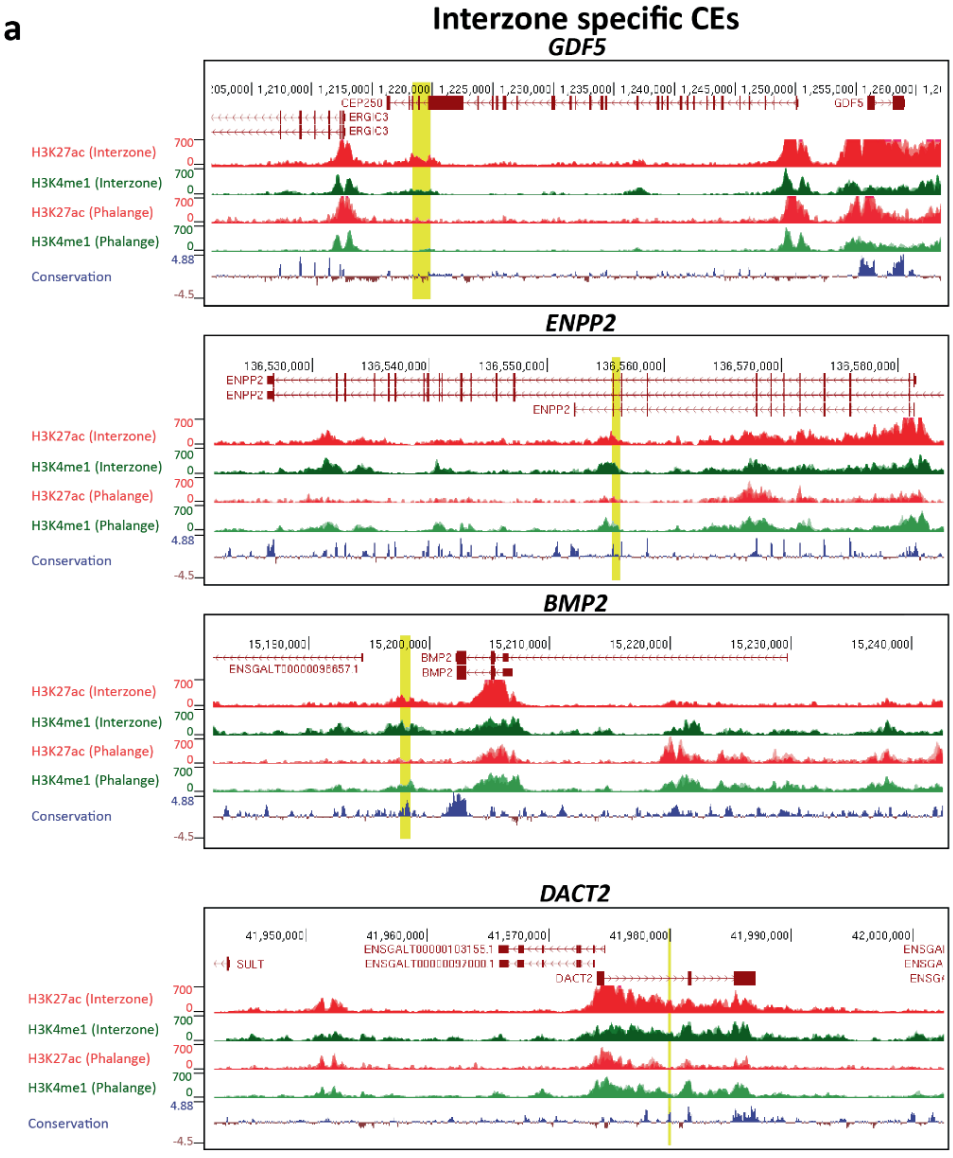
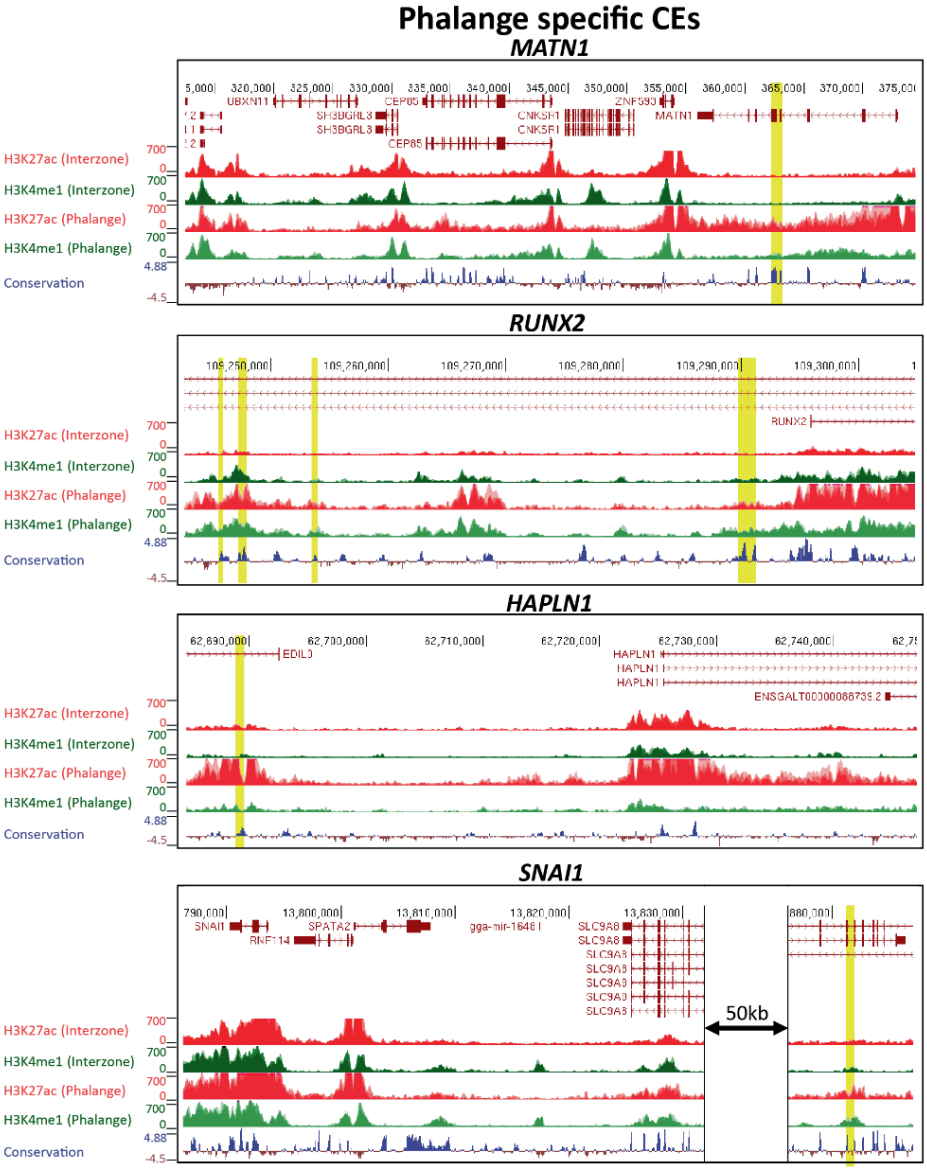


Figure 3.7. (*continued on next page*). Examples of interzone/phalange-specific CEs that are associated with DEGs.

(a) CEs associated with selected DEGs upregulated in interzone, for 4 typical loci. As in **Fig. 3.4**, the H3K27ac and H3K4me1 enrichment tracks are given in red and green, respectively. The conservation track is marked by blue/brown, and regions marked by yellow present CEs. **(b)** CEs associated with selected DEGs upregulated in phalange, again for 4 typical loci.

b

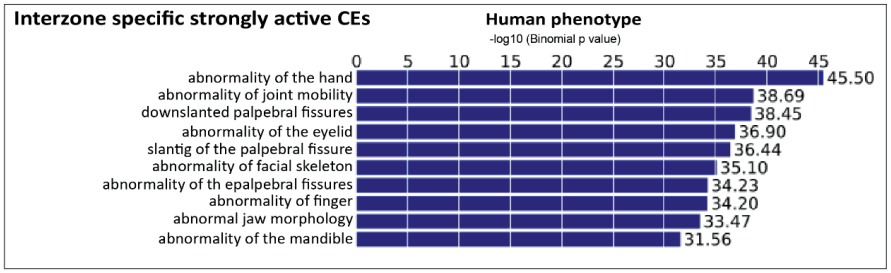

Phalange specific CEs
MATN1
H3K27ac (Interzone)
H3K4me1 (Interzone)
H3K27ac (Phalange)
H3K4me1 (Phalange)
Conservation
RUNX2
HAPLN1
SNAI1
50kb

3.3.5. CEs regulate skeletal malformation and disease-relevant genes, and are associated with a higher risk of OA

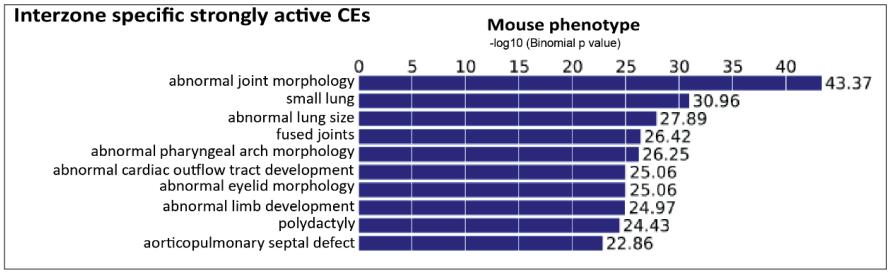
Mutations in genes and CEs have been linked to various limb malformations and skeletal defects (Lohan *et al.*, 2014; Tayebi *et al.*, 2014; Lupiañez *et al.*, 2015; Bunyan *et al.*, 2016). We applied two types of analysis to screen for CEs that link to molecular etiology of limb disorders in general. First, we assigned our strongly-active CEs to the proximal genes (including relevant respective marker genes and DEGs), and tested whether these genes have previously been associated with limb phenotypes, either in patients (including in syndromes) or mouse models. Analysis of interzone/phalange specific strongly-active CEs showed that these are indeed involved in the regulation of genes linked to joint and phalange abnormalities (**Table S3.13a-o**).

The interzone-specific, strongly-active CEs particularly associate with defective joint mobility in humans (**Fig. 3.8a**). For instance, we identified such putative enhancers of *OTX2* and *TGFB2*, which are genes that have been linked to joint laxity (OMIM #610125 and #614816, respectively); candidate CEs of *FLNB*, a gene associated with joint dislocation and carpal fusion (OMIM #150250 and #272460, respectively); we also predicted enhancers of *COL5A1*, a gene linked to joint hypermobility (OMIM #130000) (**Table S3.13a**). In mice the interzone-specific CEs associate with abnormal joint morphology and fused joints (**Fig. 3.8b**; *see* also **Table S3.13b-c**). Phalange-specific CEs have in humans been linked to aplasia/hypoplasia of the phalanges, short phalanges, and abnormality of the phalanges of the toe (**Fig. 3.8c**). For example, we identified putative enhancers of *BMPR1B* and *IHH* (**Table S3.13d**; both genes are associated with brachydactyly type-A (OMIM #112500), and candidate enhancers of *RUNX2,* a gene linked to cleidocranial dysplasia, with brachydactyly (OMIM #119600) (**Table S3.13f**). The phalange-specific CEs have also been linked to abnormal chondrocyte and cartilage morphology, chondrodystrophy, abnormal bone ossification and short limbs in mice (**Fig. 3.8d**; *see* also **Table S3.13j-o**).

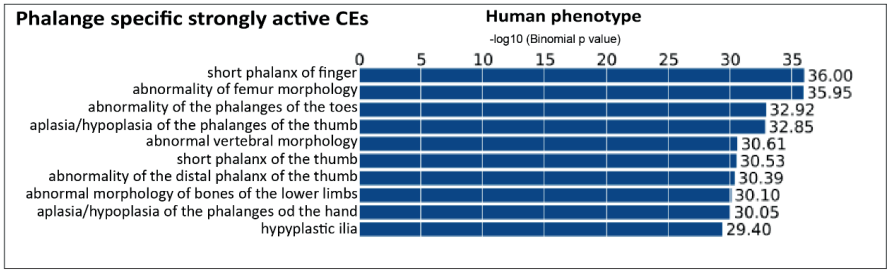
a



b



c



d

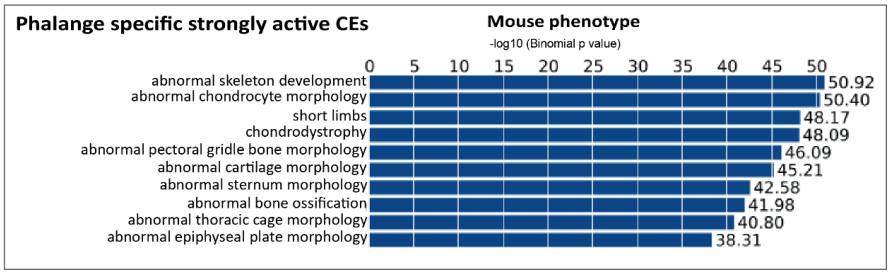


Figure 3.8. Identified CEs link to synovial joint/phalange disorders.

The candidate enhancer regions were assigned to target gene(s) using GREAT, followed by further association with human/mouse phenotypes (for details, *see* main text). Human **(a)** and mouse **(b)** phenotypes linked to interzone-specific CEs; human **(c)** and mouse **(d)** phenotypes associated with phalange-specific CEs.

Next, we screened the GWAS catalog (NHGRI-EBI; Buniello *et al.*, 2019), which resulted in identification of 3,263 single-nucleotide polymorphisms (SNPs) within CEs, with 232 of these linking to skeleton-related traits (**Table S3.14**). For instance, we identified single-nucleotide variations (SNVs) within CEs that have been associated with OA-relevant genes, such as for *ALDH1A2* (rs4775006, P-value $8x10^{-10}$; Styrkarsdottir *et al.*, 2014; Shepherd *et al.*, 2018) and *WWP2* (rs34195470, p-value $3x10^{-13}$; Mokuda *et al.*, 2019; Boer *et al.*, 2021). Additionally, we identified SNVs associated with increased risk of OA, which are located within CEs that map to *LRIG3* (rs79056043, P-value $1x10^{-9}$), *CRADD* (rs7953280, P-value $5x10^{-12}$) and *ROCR* (rs8067763, P-value $2x10^{-9}$) (Tachmazidou *et al.*, 2019; Boer *et al.*, 2021).

We then checked whether the SNPs located in CEs may affect the TF motifs. For this, we used the motifbreakR package, which provides scores for each SNP, both for reference and alternative allele, indicating the importance of the studied TF motifs (Coetzee *et al.*, 2015). The bigger the difference between these two alleles, the stronger motifbreakR predicts a variant effect on the analyzed TF motif. Scores below 0.4 are marked as neutral, <0.7 as weak, and >0.7 as strong. Using motifbreakR, we annotated 75% of SNPs located in CEs, out of which three-quarters are predicted to have at least one strong effect on any TF binding motif (**Tables S3.15 and S3.16**). Similar proportions were observed among variants related to skeletal traits. For example, rs34195470, associated with *WWP2*, may affect *TAL1* and *SRY* binding motifs (**Fig. S3.8**). rs4775006 linked to *ALDH1A2* may damage the binding site in *VDR* (**Fig. S3.9**). Also, rs8067763 associated with *ROCR*, may disrupt the *MECOM* motif (**Fig. S3.10**). Altogether, the analysis of the tissue-specific CEs showed the association with either synovial joint or phalange congenital abnormalities, which affect their function, as well as the identified CEs of genes relevant to joint degenerative disorders, in particular OA.

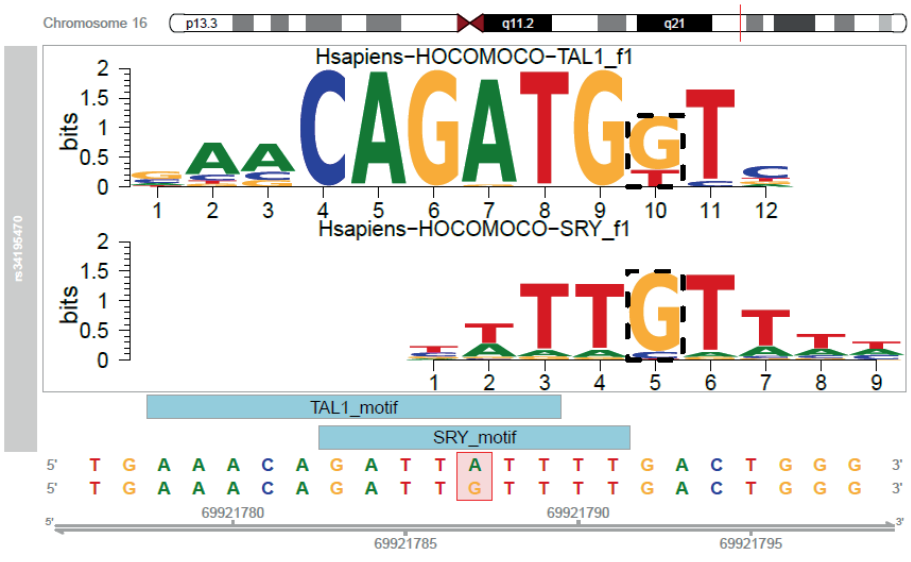


Figure S3.8. The rs34195470 is predicted to cause strong motif disturbance of TAL1 and SR binding sites.

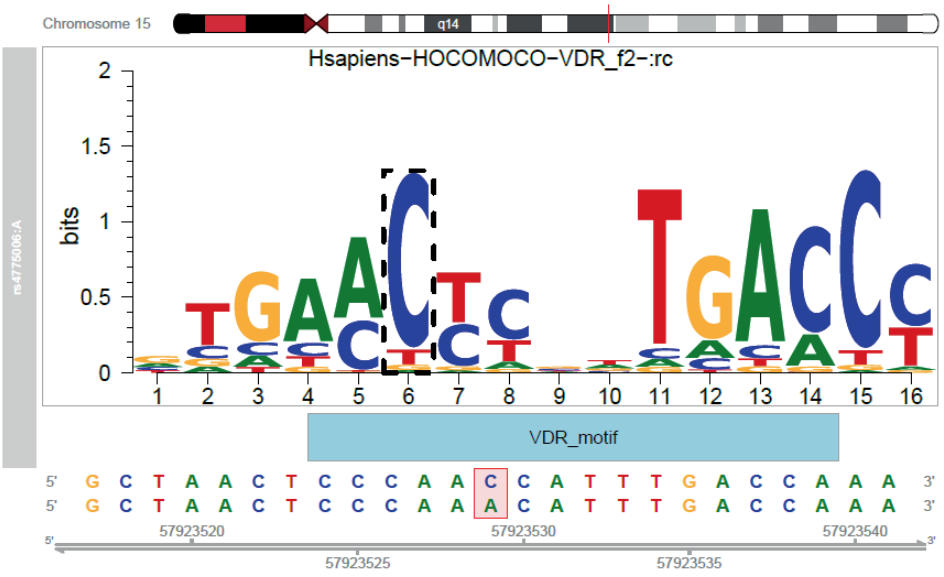


Figure S3.9. **The rs4775006 is predicted to affect the VDR binding site.**

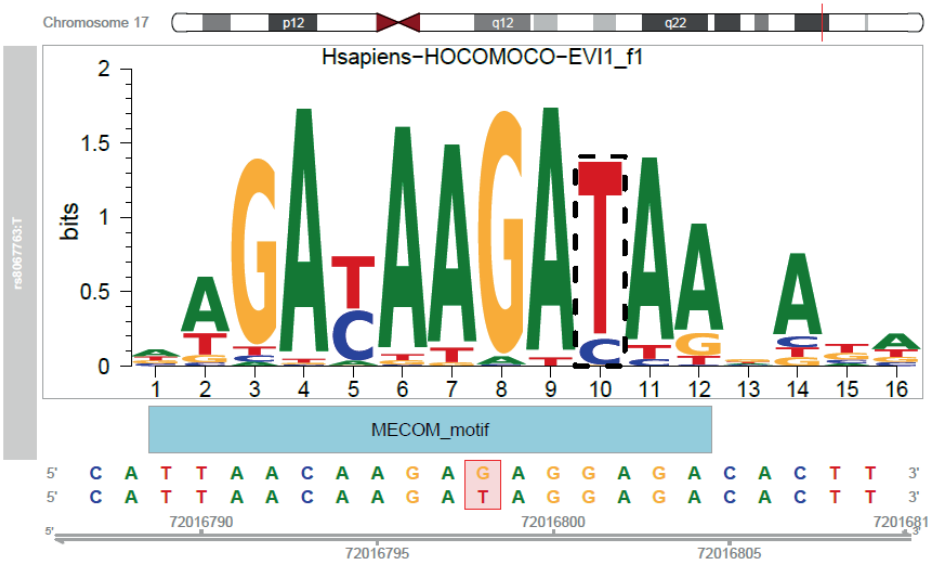


Figure S3.10. **The rs8067763 is predicted to damage the MECOM binding site.**

3.4. Discussion

Transcriptome analysis of joint interzone gained attention in recent years (Jenner *et al.*, 2014; Feregrino *et al.*, 2019; Bian *et al.*, 2020), but unlike in other fields, relatively little is known about the *cis*-regulatory elements involved in the establishment of interzone during limb development. Here, we carried out an integrative analysis of transcriptomic and epigenetic data, and subsequently generated a ChIP-seq based CE atlas, for separated interzone and phalange, respectively. For this, we optimized a fast-dissection protocol for careful collection of joint interzone samples. Using both RNA-seq and validation RT-qPCR with selected markers, we showed that such collected interzones have significantly higher expression of *GDF5*, *ENPP* and *ERG* as compared to adjacent phalange. At the same time, we validated the dissection protocol for collecting interzone cells, for using these in genome-wide experiments that require high numbers of input cells. This optimization of sample collection and separation permitted for the identification of CEs, and correlation of the latter with digit/joint formation. Functional annotation of the CEs illustrates that strongly-active CE regions enriched in both H3K27ac and H3K4me1 are associated with genes important for cartilage/skeletal development, unlike the CEs enriched in one of the two H3 marks only. Also, Cheung *et al.* (2020), who used differentiated chondrocytes, showed that strongly-active CEs identified by ChromHMM are linked to genes pivotal for regulation of the cell-type specific processes, like during chondrogenesis and in cartilage function. In contrast and similarly to our findings, active and poised CEs from differentiated chondrocytes have been associated with more general GO terms. The H3 signature determinations, just as often used extra ATAC-seq or DNase-seq results, likely still miss CEs in general or CEs that are active in the respective cells during earlier of later stages of joint interzone formation. However, to our knowledge this is the first time H3-signatures for these difficult-to-get cell populations in the joint interzone are obtained and provide an important starting point for subsequent studies in the field. Analysis of TF motifs located in strongly-active CEs shows that these regions are enriched in motifs crucial for both synovial joint and phalange development. Moreover, these CEs are enriched for the PITX1 motif, a TF broadly associated with limb enhancers and involved in establishment of hindlimb identity (Infante *et al.*, 2013). Therefore, these CEs might play role in establishment of the hindlimb identity. We also identify CEs unique for interzone/phalange linked to specific biological processes, which correlate with the pathway enrichment analysis of DEGs.

One of the pathways enriched in upregulated genes and CEs in interzones is transmembrane receptor protein serine/threonine kinase signaling, i.e. TGFβ/BMP family signaling. This signaling system has been well-characterized in the process of chondrogenic differentiation (Tsumaki *et al.*, 2002; Kobayashi *et al.*, 2005; Keller *et al.*, 2011; Wang *et al.*, 2014), but is still not well characterized in joint interzones. Suppression of BMP activity in the interzone region is essential for normal joint development (Brunet *et al.*, 1998; Ray *et al.*, 2015). In contrast, *GDF5* is expressed at high level in the interzone region, and *GDF5-null* mutations result in joint defects (Storm and Kingsley, 1996). In our study we confirm in an alternative way previously described enhancers of *GDF5*, and identify novel CEs of *GDF5*. Another example of a BMP-upregulated gene in the interzone region, consistent with reported *in situ* RNA-hybridization, is *BMP2* (Macias *et al.*, 1997). *BMP2* is involved in joint maturation; its genetic inactivation in synovial joint forming cells results in changes in extracellular matrix and also shape of the meniscus (Gamer *et al.*, 2018). Interestingly, within the genomic regulatory landscape of *BMP2*, we identified several CEs likely to be active in the interzone, but this will require further

investigation. We also documented a change of enhancer state between strongly-active and poised enhancers, which enabled us to identify CEs associated with unique biological processes. Some of these processes correlate with analysis of interzone/phalange-specific CEs and pathway enrichment analysis of certain DEGs.

When considering changes in gene expression causal for limb malformation, it becomes necessary to also include studies of enhancers, which may drive misexpression of disease-causing genes. We show that many CEs associate with genes important for normal development as well as etiology of both synovial joints and phalanges. For instance, in our study we identified CEs of *GDF5* (as shown by Chen *et al.*, 2016) to be active in the interzone region. Mutations in *GDF5* lead to joint malformations, and in a genomic region encompassing *GDF5* these experimentally tested enhancers have been linked to higher risk of OA (Miyamoto *et al.*, 2007; Egli *et al.*, 2009; Chen *et al.*, 2016; Capellini *et al.*, 2017). We also characterized interzone-specific CEs of the *BMP2* locus. Importantly, *BMP2* conditional knock-out mice develop progressive OA in the knees (Gamer *et al.*, 2018). Another set of gene-linked enhancers in joint abnormalities are CEs for *FSTL1*, associated with rheumatoid arthritis (li *et al.*, 2011). We also identified CEs linked to genes (e.g., *OTX2*, *TGFB2, COL5A1*) associated with defects in joint mobility.

We used the GWAS catalog NHGRI-EBI to identify SNVs located within CEs and that are associated with higher risk of OA. This yields CEs linked to the genes previously described in OA, such as *ALDH1A2* and *WWP2*. Next, we showed that SNPs located in these CEs disrupts binding motifs for TFs. In addition to interzone CEs, we also characterized several phalange CEs linked to the genes important in chondrocyte-related disorders. For instance, we identified CEs associated with *RUNX2* and *IHH*, both (also) pivotal genes in the molecular etiology of limb malformation, including brachydactyly (Kirkpatrick *et al.*, 2003; Hordyjewska *et al.*, 2017). Altogether, this illustrates that our CE atlas provides information on association of CEs with already existing gene-to-disease correlations. This will be helpful in studies of possible variations in genomic regions in patients without mutations in the protein-coding genes. We also discovered that *BMPR2* was co-expressed with *GDF5* in interzones. Therefore, it is possible that BMP signaling in the interzone region is prevented at the intracellular level, or BMP ligands play a dual function, which is mutually exclusive in the interzone cells and chondrocytes. Both *GDF5* and *BMPR2* are upregulated in chondrocyte de-differentiation *in vitro* (Schlegel *et al.*, 2009), which may support the hypothesis that interzone originates from de-differentiated chondrocytes (Decker *et al.*, 2014).

WNT signaling is crucial in the formation of joint interzone. We confirmed that *DACT2* (encoding an intracellular WNT beta-catenin dependent pathway inhibitor) was significantly upregulated in interzone as compared to phalange, consistent with reported *in situ* RNA-hybridization (Sensiate *et al.*, 2014). This shows again the complexity of WNT signaling regulation during joint formation. Within the genomic regulatory landscape of *DACT2* we identify several CEs that are likely active in the interzone. However, the more detailed characterization of these *DACT2* enhancers requires additional studies. Our work identified several genes from the TGFβ family ligands or other system components upregulated in interzone, for instance *TGFB2*, *LTBP1* and *INHBB.* Several studies showed antagonistic action of TGFβ/Activin-Nodal pathways on BMP signaling in several cell types (Chen *et al.*; 1991; Kawamura *et al.*, 2012). Therefore, it is tempting to hypothesize that the upregulation of TGFβ family system components in interzone may have effects on attenuation and/or inhibition of BMP signaling within prospective sites of synovial joint formation.

3.5. Experimental procedures

Tissue collection

All vertebrate animal experiments (with chick early embryos) were carried out in accordance with the relevant guidelines as applied and approved by the Ethical Committee at the Medical University in Lublin, where this work was performed, and also comply with the European regulations (directive 2010/63/EU). The tissue collection was performed on the chicken embryos until 7.5 days post fertilization which is exempt from the Ethical Committee Approval. Chick White Leghorn embryonated eggs were incubated at 38.5ºC in fixed humidity for 7.5 days, followed by evaluation of developmental stage based on the Hamilton Hamburger classification (HH), using a Zeiss Stereo Discovery V8 microscope equipped with 0.63x Plan Apo S Objective Lens. Selected embryos at HH32 were sacrificed for tissue microdissection. The joint interzones and adjacent phalange samples were microdissected from hindlimb digit-3 using Dumont No.5 forceps (tip dimensions: 0.005 x 0.025 mm)

RNA extraction

Total RNA was extracted and prepared using Syngen Tissue RNA Kit followed by DNA digestion with QIAGEN RNase-free DNase Set. The RNA quality was validated on 1% agarose gel (for RT-qPCR) or on Agilent 2100 Bioanalyzer system with RNA 6000 Nano Assay (for RNA-seq). All latter samples had a RIN value > 9.0.

RT-qPCR

Three independently extracted RNAs from both interzone and phalange were reverse transcribed to cDNA using Invitrogen™ SuperScript™ IV Reverse Transcriptase and Oligo(dT) primer. The qPCR was performed using PowerUp™ SYBR™ Green Master Mix II on LightCycler® 480 Instrument II. Gene expression was normalized to expression of *GAPDH*. Statistics were computed using Mann-Whitney-Wilcoxon Test. The list of used primers is given in **Table S3.17**.

RNA-seq library preparation

The 6 samples (3 of interzone, 3 of phalange-independent biological replicates) were prepared with the Smart-seq2 method (Picelli *et al.*, 2013). In brief, poly(A)-RNA was reverse transcribed using oligo(dT) primers. Template switching by reverse transcriptase was achieved by using a LNA-containing TSO oligonucleotide. The reverse-transcribed cDNA was pre-amplified with primers for 18 cycles, followed by clean-up. Tagmentation was performed on 500 pg of the pre-amplified cDNA with Tn5 followed by gap repair. The tagmented library was extended with Illumina adaptor sequences by PCR for 14 cycles and purified. The resulting sequencing library was measured on Bioanalyzer and equimolar amounts loaded onto a flowcell and sequenced according to the Illumina TruSeq v3 protocol on the HiSeq2500, with a single-read 50 bp and dual 9 bp indices.

RNA-seq data analysis

The fastq files were checked for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by removal of adapters using Trimmomatic (Bolger *et al.*, 2015). Further, reads were mapped to the *Gallus gallus* 6.0 reference genome using STAR with default parameters (Dobin *et al.*, 2013). Gene expression values were called using featureCounts with Ensemble release 104 annotation (Liao *et al.*, 2014). The differential data analysis has been performed with DEseq2 (langmead and Salzberg, 2012), and heatmaps have been created in R environment for statistical computing. The expression of DEGs were compared to aviable scRNA-seq data from Bian *et al.* (2020), using the web application: cahanlab.org/resources/joint_ontogeny.

Chromatin-immunoprecipitation (ChIP)

For each sample, either 100 interzones or phalanges were dissected from the 3rd hindlimb digit and pooled together. Further, tissues were dissociated for 3 hours at 37ºC using 2.4% Collagenase-II (Gibco™) resuspended in DMEM/high-glucose medium containing 10% fetal bovine serum (FBS). Cells were passed through a 40-µm cell strainer (BD Falcon) and then counted using a hemocytometer. 10^6 cells were cross-linked using 1% formaldehyde at room temperature (RT, 24ºC) for 9 min. Fixation was quenched with ice-cold 0.125 M glycine for 5 min at 4ºC. To remove excess formaldehyde, two rounds of centrifugation of the cells followed by resuspension in ice-cold PBS were carried out. Next, the cell nuclei were isolated using ice-cold nuclei extraction buffer (NEB) containing 10 mM Tris-HCl pH 7.5, 0.15 mM NaCl, 1 mM EDTA, 1%

IGEPAL® CA-630 and complete protease inhibitors (Roche). Subsequently, the nuclei were resuspended in SDS-containing lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 1% SDS), and chromatin was sheard obtaining the average size of 150 bp in AFA Fiber Pre-Slit Snap-Cap (130 µl) microtube using a S220 Focused-ultrasonicator.

For ChIP, 500 ng of sonicated chromatin was immunoprecipitated with 7.5 µg of anti-H3K27ac antibody (Active Motif, Cat. No. 39133) or anti-H3K4me1 (61781). Input sample was collected prior to immunoprecipitation reaction. Chromatin pre-cleaning incubation with protein-A and protein-G agarose beads (Millipore) was carried out in immunoprecipitation buffer (50 mM Tris-HCl pH 8.0, 0.15 M NaCl, 1 mM EDTA pH8.0, 1% Triton X-100 and 0.1% sodium deoxycholate) for 3 hours at 4ºC while rotating. In parallel, the antibodies were incubated with previously blocked A- and G-agarose beads (Millipore) also for 3 hours at 4ºC, again while rotating.

After pre-cleaning, the chromatin was mixed with the pre-bound antibodies with A- and G-agarose beads and incubated overnight (O/N) at 4ºC, rotating. The next day, multiple rounds of washes of the beads were conducted. Each wash was carried for 10 min at 4°C while rotating. The beads were washed once with RIPA-150 buffer (50 mM Tris-HCl pH8.0, 0.15M NaCl, 1mM EDTA pH8.0, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), twice with RIPA-500 (50 mM Tris-HCl pH8.0, 0.5 M NaCl, 1 mM EDTA pH8.0, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), once with RIPA-LiCl (50 mM Tris-HCl pH8.0, 1mM EDTA pH8.0, 1% Nonidet-P40, 0.7% sodium deoxycholate, 0.5M LiCl) and twice in TE buffer (10 mM Tris-HCl pH8.0, 1mM EDTA pH8.0). Subsequently, the chromatin was eluted with 200 µl of fresh elution buffer (1% SDS and 0.1 M $NaHCO_3$) followed by addition of 100 µl of TE buffer and 25 µl of 5 M NaCl prior to reverse-crosslinking at 65°C for 16 hours. The next day, chromatin was incubated with 2 µl of Proteinase-K (10 mg/ml) for 1 hour at 56°C, and 2 µl of RNaseA (10 mg/ml) for 45 min at 37ºC, and DNA was further purified using QIAquick PCR Purification Kit (QIAGEN). The size distribution of immunoprecipitated fragments was evaluated using Agilent 2100 Bioanalyzer system with High Sensitivity DNA. Additionally, the DNA-concentration of input and immunoprecipitated samples was measured on Qubit 2.0 Fluorometer (Invitrogen).

ChIP-sequencing

ChIP-seq libraries were prepared using QIAseq Ultra Low Input Library Kit (QIAGEN, Hilden, Germany). Briefly, DNA was end-repaired, adenosines were added to the 3′ ends of dsDNA and adapters were ligated (adapters from NEB, Ipswich, MA, USA). Following the adapter ligation, uracil was digested by USER enzyme from NEB (Ipswich, MA, USA) in a loop structure of the adapter. Adapters containing DNA fragments were amplified by PCR using NEB starters (Ipswich MA, USA). Library quality evaluation was done with Agilent 2100 Bioanalyzer using the Agilent DNA High Sensitivity chip (Agilent Technologies, Ltd.) Quantification and quality evaluation of obtained samples were done using Nanodrop spectrophotometer (Thermo Scientific, NanoDrop products, Wilmington, USA), Quantus fluorometer (Promega Corporation, Madison, USA) and 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Mean library size was 300 bp. Libraries were run in the rapid run flow cell and were single-end sequenced (65 bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

ChIP-seq data analysis, CEs identification and CE annotation

The quality of raw fastq files were validated using FastQC and adapters were removed using Trimmomatic. Next, reads were mapped to the *Gallus gallus* 6.0 reference genome using Bowtie 2 with default parameters (Langmead and Salzberg, 2012) and PCR-duplicates were marked and removed using Picard (http://broadinstitute.github.io/picard/). The peaks were called using MACS2 with significance level threshold FDR <0.05, and normalization to input sample (Zhang *et al.*, 2008). Further, biological replicates were merged using BEDTools (Quinlan and Hall, 2010).

The unsupervised clustering of H3K27ac and H3K4me1 peaks was performed using DiffBind (https://bioconductor.org/packages/DiffBind/) with normalized IP samples to input. The experiment-specific lists containing anomalously enriched regions were generated using the GreyListChIP and further removed from datasets.

CEs were identified using in-house script. Briefly, the promoter regions (1 kb -/+ from TSS) were filtered out from H3K27ac and K3K4me1 dataset and nearby peaks (< 1 kb) were merged with GenomicRanges::reduce(min.gapwidth=1000) (Lawrence *et al.*, 2013). Further, the conserved CEs were selected using BEDTools intersect followed by merging nearby genomic intervals with GenomicRanges::reduce (min.gapwidth=1000). The consensus and cell-specific CEs were identified using BEDTools::intersect -f 0.9 -r and BEDTools::intersect -v, respectively. The functional interpretation of CEs was performed using GREAT, with the genomic regions previously lifted to hg38 genome using liftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

Tissue-specific CEs have been mapped to the DEGs using our in-house script. Specifically, we characterized the *cis*-regulatory landscapes of such genes defined by the borders of TADs. To do this we utilized the annotated TADs from chicken fibroblasts (Fishman *et al.*, 2019). We identified the CEs located within the TADs encompassing DEGs and further associated them with DEGs. If the gene was located within genomic region not annotated by any TAD we defined cis-regulatory landscape as region -/+ 1 Mb from the TSS.

The enrichment tracks for H3K4me1 and H3K27ac ChIP-seq data were generated using deepTools (Ramirez *et al.*, 2016). Specifically, bamCoverage with reads per kilobase per million mapped reads (RPKM) per bin normalization was used. The enrichment tracks were visualized by loading to UCSC Genome Browser. The tracks visualized in UCSC Genome Browser were merged for biological replicates using transparent method (https://genome.ucsc.edu/cgi-bin/hgCollection).

Motif enrichment analysis of CEs

The motif enrichment analysis was preformed using HOMER (Heinz *et al.*, 2010) tool with function "findMotifsGenome.pl". The CEs coordinates were adjusted to 1kb from the center of CE region. The randomly selected regions (1 kb size) from chicken genome as background. A p value ≤ 0.01 was considered to select significantly enriched motifs.

Annotation of CEs to disease-relevant genes, and identification of CEs associated with a higher risk of OA

To annotate CEs to disease-relevant genes and locate SNPs within the CEs these genomic regions were lifted to hg38 genome using liftOver. The GREAT was used to associate CEs with genes and retriever human and mouse phenotypes. The GenomicRanges::findOverlaps was used identify SNPs located within the CEs. To assess whether detected SNPs might damage motifs recognized by Transcription Factors, the motifbreakR tool was used[45]. Output was generated based on: human reference genome hg38, the SNPs were liftover to human reference genome hg38, and dbSNP versioned 155. No upstream filtering was performed except for removing variants with the same rsID. Default settings were used with the maximum p-value for a match to be called or a minimum score threshold set to 10^{-4}.

Data availability

The RNA-seq and ChIP-seq data are available under the gene expression omnibus (GEO) accession number GSE198819

Acknowledgments

We thank the members of the Center for Biomics-Genomics at the Erasmus University Medical Center for technical support. We also thank all the members of the Cell Biology Department for fruitful discussion. This work was supported by Polish National Science Centre (UMO-2015/19/B/NZ4/03184) and primary funding to the Department of Cell Biology at Erasmus University Medical Center.

Author contribution

Conceptualization: DH, KN, PT, FG; Methodology: AK, BG, BW, KN, WI; Investigation in the lab: KN; Computational investigation: AO, KN, RB, Software: AO, K.; Resources: DH and PT; Data Curation: KN, PT; Writing Original draft: KN; Writing, Review & editing: DH, PT; Supervision: DH and PT; Project Administration: DH and PT; Funding Acquisition: DH and PT.

Supplementary tables

These supplementary tables are directly available as downloadable excel sheets at the journal's webpage of this Open Access publication:

https://www.nature.com/articles/s41598-022-16951-4#Sec20,

- with clickable Supplementary Information 1, containing **Tables S3.1 – S3.12**
- with clickable Supplementary Information 2, containing **Tables S3.13a-o**
- with clickable Supplementary Information 3, containing **Tables S.14 – S.17**

Literature references

Bian Q, Cheng YH, Wilson JP, Su EY, Kim DW, Wang H, Yoo S, Blackshaw S, Cahan P. A single cell transcriptional atlas of early synovial joint development. Development. 2020 Jul 20;147(14):dev185777.

Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, Wu TT, Zheng J, Hartley A, Teder-Laving M, Skogholt AH, Terao C, Zengini E, Alexiadis G, Barysenka A, Bjornsdottir G, Gabrielsen ME, Gilly A, Ingvarsson T, Johnsen MB, Jonsson H, Kloppenburg M, Luetge A, Lund SH, Mägi R, Mangino M, Nelissen RRGHH, Shivakumar M, Steinberg J, Takuwa H, Thomas LF, Tuerlings M; arcOGEN Consortium; HUNT All-In Pain; ARGO Consortium; Regeneron Genetics Center, Babis GC, Cheung JPY, Kang JH, Kraft P, Lietman SA, Samartzis D, Slagboom PE, Stefansson K, Thorsteinsdottir U, Tobias JH, Uitterlinden AG, Winsvold B, Zwart JA, Davey Smith G, Sham PC, Thorleifsson G, Gaunt TR, Morris AP, Valdes AM, Tsezou A, Cheah KSE, Ikegawa S, Hveem K, Esko T, Wilkinson JM, Meulenbelt I, Lee MTM, van Meurs JBJ, Styrkársdóttir U, Zeggini E. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021 Sep 2;184(18):4784-818.e17.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20.

Brison N, Tylzanowski P, Debeer P. Limb skeletal malformations - what the HOX is going on? Eur J Med Genet. 2012 Jan;55(1):1-7.

Brunet LJ, McMahon JA, McMahon AP, Harland RM. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. Science. 1998 May 29;280(5368):1455-7.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019 Jan 8;47(D1):D1005-D12.

Bunyan DJ, Baffico M, Capone L, Vannelli S, Iughetti L, Schmitt S, Taylor EJ, Herridge AA, Shears D, Forabosco A, Coviello DA. Duplications upstream and downstream of SHOX identified as novel causes of Leri-Weill dyschondrosteosis or idiopathic short stature. Am J Med Genet A. 2016 Apr;170A(4):949-57.

Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013 Mar 7;49(5):825-37.

Capellini TD, Chen H, Cao J, Doxey AC, Kiapour AM, Schoor M, Kingsley DM. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat Genet. 2017 Aug;49(8):1202-10.

Chen D, Shen J, Zhao W, Wang T, Han L, Hamilton JL, Im HJ. Osteoarthritis: toward a comprehensive understanding of pathological mechanism. Bone Res. 2017 Jan 17;5:16044.

Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, Shoulders, Elbows, Knees, and Toes: Modular Gdf5 Enhancers Control Different Joints in the Vertebrate Skeleton. PLoS Genet. 2016 Nov 30;12(11):e1006454.

Chen P, Carrington JL, Hammonds RG, Reddi AH. Stimulation of chondrogenesis in limb bud mesoderm cells by recombinant human bone morphogenetic protein 2B (BMP-2B) and modulation by transforming growth factor beta 1 and beta 2. Exp Cell Res. 1991 Aug;195(2):509-15.

Cheung K, Barter MJ, Falk J, Proctor CJ, Reynard LN, Young DA. Histone ChIP-Seq identifies differential enhancer usage during chondrogenesis as critical for defining cell-type specificity. FASEB J. 2020 Apr;34(4):5317-31.

Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics. 2015 Dec 1;31(23):3847-9.

Decker RS, Koyama E, Pacifici M. Genesis and morphogenesis of limb synovial joints and articular cartilage. Matrix Biol. 2014 Oct;39:5-10.

Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? Int J Rheum Dis. 2011 May;14(2):175-80.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21.

Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A, Carr A, Chapman K, Loughlin J. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. Arthritis Rheum. 2009 Jul;60(7):2055-64.

Feregrino C, Sacher F, Parnas O, Tschopp P. A single-cell transcriptomic atlas of the developing chicken limb. BMC Genomics. 2019 May 22;20(1):401.

Fish A, Chen L, Capra JA. Gene Regulatory Enhancers with Evolutionarily Conserved Activity Are More Pleiotropic than Those with Species-Specific Activity. Genome Biol Evol. 2017 Oct 1;9(10):2615-25.

Fishman V, Battulin N, Nuriddinov M, Maslova A, Zlotina A, Strunov A, Chervyakova D, Korablev A, Serov O, Krasikova A. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res. 2019 Jan 25;47(2):648-65.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016 May 31;15(9):2038-49.

Gamer LW, Pregizer S, Gamer J, Feigenson M, Ionescu A, Li Q, Han L, Rosen V. The Role of Bmp2 in the Maturation and Maintenance of the Murine Knee Joint. J Bone Miner Res. 2018 Sep;33(9):1708-17.

García-González E, Escamilla-Del-Arenal M, Arzate-Mejía R, Recillas-Targa F. Chromatin remodeling effects on enhancer activity. Cell Mol Life Sci. 2016 Aug;73(15):2897-910.

Girisha KM, Kortüm F, Shah H, Alawi M, Dalal A, Bhavani GS, Kutsche K. A novel multiple joint dislocation syndrome associated with a homozygous nonsense variant in the EXOC6B gene. Eur J Hum Genet. 2016 Aug;24(8):1206-10.

Gu J, Lu Y, Li F, Qiao L, Wang Q, Li N, Borgia JA, Deng Y, Lei G, Zheng Q. Identification and characterization of the novel Col10a1 regulatory mechanism during chondrocyte hypertrophic differentiation. Cell Death Dis. 2014 Oct 16;5(10):e1469.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576-89.

Hordyjewska E, Jaruga A, Kandzierski G, Tylzanowski P. Novel Mutation of the *RUNX2* Gene in Patients with Cleidocranial Dysplasia. Mol Syndromol. 2017 Aug;8(5):253-60.

Hu G, Codina M, Fisher S. Multiple enhancers associated with ACAN suggest highly redundant transcriptional regulation in cartilage. Matrix Biol. 2012 Jul;31(6):328-37.

Hyde G, Dover S, Aszodi A, Wallis GA, Boot-Handford RP. Lineage tracing using matrilin-1 gene expression reveals that articular chondrocytes exist as the joint interzone forms. Dev Biol. 2007 Apr 15;304(2):825-33.

Infante CR, Park S, Mihala AG, Kingsley DM, Menke DB. Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. Dev Biol. 2013 Feb 1;374(1):234-44.

Iwamoto M, Tamamura Y, Koyama E, Komori T, Takeshita N, Williams JA, Nakamura T, Enomoto-Iwamoto M, Pacifici M. Transcription factor ERG and joint and articular cartilage formation during mouse limb and spine skeletogenesis. Dev Biol. 2007 May 1;305(1):40-51.

Jenner F, IJpma A, Cleary M, Heijsman D, Narcisi R, van der Spek PJ, Kremer A, van Weeren R, Brama P, van Osch GJ. Differential gene expression of the intermediate and outer interzone layers of developing articular cartilage in murine embryos. Stem Cells Dev. 2014 Aug 15;23(16):1883-98.

Kawamura I, Maeda S, Imamura K, Setoguchi T, Yokouchi M, Ishidou Y, Komiya S. SnoN suppresses maturation of chondrocytes by mediating signal cross-talk between transforming growth factor-β and bone morphogenetic protein pathways. J Biol Chem. 2012 Aug 17;287(34):29101-13.

Keller B, Yang T, Chen Y, Munivez E, Bertin T, Zabel B, Lee B. Interaction of TGFβ and BMP signaling pathways during chondrogenesis. PLoS One. 2011 Jan 28;6(1):e16421.

Khan IM, Redman SN, Williams R, Dowthwaite GP, Oldfield SF, Archer CW. The development of synovial joints. Curr Top Dev Biol. 2007;79:1-36.

Kirkpatrick TJ, Au KS, Mastrobattista JM, McCready ME, Bulman DE, Northrup H. Identification of a mutation in the Indian Hedgehog (IHH) gene causing brachydactyly type A1 and evidence for a third locus. J Med Genet. 2003 Jan;40(1):42-4.

Kobayashi T, Lyons KM, McMahon AP, Kronenberg HM. BMP signaling stimulates cellular differentiation at multiple steps during cartilage development. Proc Natl Acad Sci USA. 2005 Dec 13;102(50):18023-7.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.

Li D, Wang Y, Xu N, Wei Q, Wu M, Li X, Zheng P, Sun S, Jin Y, Zhang G, Liao R, Zhang P. Follistatin-like protein 1 is elevated in systemic autoimmune diseases and correlated with disease activity in patients with rheumatoid arthritis. Arthritis Res Ther. 2011 Feb 8;13(1):R17.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923-30.

Lohan S, Spielmann M, Doelken SC, Flöttmann R, Muhammad F, Baig SM, Wajid M, Hülsemann W, Habenicht R, Kjaer KW, Patil SJ, Girisha KM, Abarca-Barriga HH, Mundlos S, Klopocki E. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. Clin Genet. 2014 Oct;86(4):318-25.

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

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 21;161(5):1012-25.

Luyten FP, Tylzanowski P, Lories RJ. Wnt signaling and osteoarthritis. Bone. 2009 Apr;44(4):522-7.

Macias D, Gañan Y, Sampath TK, Piedra ME, Ros MA, Hurle JM. Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. Development. 1997 Mar;124(6):1109-17.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495-501.

Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, Fujioka M, Sudo A, Uchida A, Yamamoto S, Ozaki K, Takigawa M, Tanaka T, Nakamura Y, Jiang Q, Ikegawa S. A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet. 2007 Apr;39(4):529-33.

Mokuda S, Nakamichi R, Matsuzaki T, Ito Y, Sato T, Miyata K, Inui M, Olmer M, Sugiyama E, Lotz M, Asahara H. Wwp2 maintains cartilage homeostasis through regulation of Adamts5. Nat Commun. 2019 Jun 3;10(1):2429.

Ohta Y, Okabe T, Larmour C, Di Rocco A, Maijenburg MW, Phillips A, Speck NA, Wakitani S, Nakamura T, Yamada Y, Enomoto-Iwamoto M, Pacifici M, Iwamoto M. Articular cartilage endurance and resistance to osteoarthritic changes require transcription factor Erg. Arthritis Rheumatol. 2015 Oct;67(10):2679-90.

Pazin DE, Gamer LW, Capelo LP, Cox KA, Rosen V. Gene signature of the embryonic meniscus. J Orthop Res. 2014 Jan;32(1):46-53.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W160-5.

Ray A, Singh PN, Sohaskey ML, Harland RM, Bandyopadhyay A. Precise spatial restriction of BMP signaling is essential for articular cartilage differentiation. Development. 2015 Mar 15;142(6):1169-79.

Roelofs AJ, Rocke JP, De Bari C. Cell-based approaches to joint surface repair: a research perspective. Osteoarthritis Cartilage. 2013 Jul;21(7):892-900.

Schlegel W, Albrecht C, Eckl P, Freudenthaler H, Berger A, Vécsei V, Marlovits S. Dedifferentiation of human articular chondrocytes is associated with alterations in expression patterns of GDF-5 and its receptors. J Cell Mol Med. 2009 Sep;13(9B):3398-404.

Sensiate LA, Sobreira DR, Da Veiga FC, Peterlini DJ, Pedrosa AV, Rirsch T, Joazeiro PP, Schubert FR, Collares-Buzato CB, Xavier-Neto J, Dietrich S, Alvares LE. Dact gene expression profiles suggest a role for this gene family in integrating Wnt and TGF-β signaling pathways during chicken limb development. Dev Dyn. 2014 Mar;243(3):428-39.

Shepherd C, Zhu D, Skelton AJ, Combe J, Threadgold H, Zhu L, Vincent TL, Stuart P, Reynard LN, Loughlin J. Functional Characterization of the Osteoarthritis Genetic Risk Residing at ALDH1A2 Identifies rs12915901 as a Key Target Variant. Arthritis Rheumatol. 2018 Oct;70(10):1577-87.

Shwartz Y, Viukov S, Krief S, Zelzer E. Joint Development Involves a Continuous Influx of Gdf5-Positive Cells. Cell Rep. 2016 Jun 21;15(12):2577-87.

Spagnoli A, O'Rear L, Chandler RL, Granero-Molto F, Mortlock DP, Gorska AE, Weis JA, Longobardi L, Chytil A, Shimer K, Moses HL. TGF-beta signaling is essential for joint morphogenesis. J Cell Biol. 2007 Jun 18;177(6):1105-17.

Spielmann M, Brancati F, Krawitz PM, Robinson PN, Ibrahim DM, Franke M, Hecht J, Lohan S, Dathe K, Nardone AM, Ferrari P, Landi A, Wittler L, Timmermann B, Chan D, Mennen U, Klopocki E, Mundlos S. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am J Hum Genet. 2012 Oct 5;91(4):629-35.

Storm EE, Kingsley DM. Joint patterning defects caused by single and double mutations in members of the bone morphogenetic protein (BMP) family. Development. 1996 Dec;122(12):3969-79.

Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, Strijbosch AM, Evangelou E, Hart D, Beekman M, Jonasdottir A, Sigurdsson A, Eiriksson FF, Thorsteinsdottir M, Frigge ML, Kong A, Gudjonsson SA, Magnusson OT, Masson G; TREAT-OA Consortium; arcOGEN Consortium, Hofman A, Arden NK, Ingvarsson T, Lohmander S, Kloppenburg M, Rivadeneira F, Nelissen RG, Spector T, Uitterlinden A, Slagboom PE, Thorsteinsdottir U, Jonsdottir I, Valdes AM, Meulenbelt I, van Meurs J, Jonsson H, Stefansson K. Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. Nat Genet. 2014 May;46(5):498-502.

Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, Johnson T, Koprulu M, Zengini E, Steinberg J, Wilkinson JM, Bhatnagar S, Hoffman JD, Buchan N, Süveges D; arcOGEN Consortium, Yerges-Armstrong L, Smith GD, Gaunt TR, Scott RA, McCarthy LC, Zeggini E. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019 Feb;51(2):230-6.

Tayebi N, Jamsheer A, Flöttmann R, Sowinska-Seidler A, Doelken SC, Oehl-Jaschkowitz B, Hülsemann W, Habenicht R, Klopocki E, Mundlos S, Spielmann M. Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. Orphanet J Rare Dis. 2014 Jul 29;9:108.

Tsumaki N, Nakase T, Miyaji T, Kakiuchi M, Kimura T, Ochi T, Yoshikawa H. Bone morphogenetic protein signals are required for cartilage formation and differently regulate joint development during skeletogenesis. J Bone Miner Res. 2002 May;17(5):898-906.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007 Jan;35(Database issue):D88-92.

Vissers LE, Lausch E, Unger S, Campos-Xavier AB, Gilissen C, Rossi A, Del Rosario M, Venselaar H, Knoll U, Nampoothiri S, Nair M, Spranger J, Brunner HG, Bonafé L, Veltman JA, Zabel B, Superti-Furga A. Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, encoding the Golgi-resident nucleotide phosphatase, gPAPP. Am J Hum Genet. 2011 May 13;88(5):608-15.

Wang W, Rigueur D, Lyons KM. TGFβ signaling in cartilage development and maintenance. Birth Defects Res C Embryo Today. 2014 Mar;102(1):37-51.

Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, Heinrich V, de Villartay JP, Vingron M, Klopocki E, Visel A, Lupiáñez DG, Mundlos S. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet. 2017 Oct;49(10):1539-45.

Yao B, Wang Q, Liu CF, Bhattaram P, Li W, Mead TJ, Crish JF, Lefebvre V. The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. Nucleic Acids Res. 2015 Jun 23;43(11):5394-408.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

Chapter 4

Low-input targeted chromatin capture (low-T2C): genome compartmentalization and interactome at high resolution

Ilias Boltsis [1*], **Karol Nowosad** [2,3*], Rutger W. W. Brouwer [1,4], Przemko Tylzanowski [2,5], Wilfred F.J. van IJcken [1,4], Danny Huylebroeck [1,6], Frank Grosveld [1], Petros Kolovos [7]

[1] Department of Cell Biology, Erasmus University Medical Center, Rotterdam 3015 CN, the Netherlands

[2] Department of Biochemistry and Molecular Biology, Medical University of Lublin, 20-093 Lublin, Poland

[3] The Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland

[4] Center for Biomics, Erasmus University Medical Center, Rotterdam 3015 CN, the Netherlands

[5] Department of Development and Regeneration, Skeletal Biology and Engineering Research Center, University of Leuven, B-3000 Leuven, Belgium

[6] Department of Development and Regeneration, KU Leuven, B-3000 Leuven, Belgium

[7] Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis 68100, Greece

* *Shared first authors*

Published in

Methods Mol Biol. 2021;2351:165-79

4.1. Summary

Targeted chromatin capture (T2C) is a method derived from 3C and is used to study the 3D chromatin organization, interactomes and structural changes associated with gene regulation, progression through the cell cycle, and cell survival and development. Low-input targeted chromatin capture (low-T2C) is an optimized version of the T2C protocol for low numbers of cells/inputs. Here, we present the protocol for low-T2C, including all the experimental steps and bio-informatics tools in detail.

4.2. Introduction

How the genome is folded into the cell nucleus while maintaining its function in many critical molecular-cellular processes is still poorly understood. To address that issue, technological advances over the last decades led to the development of several chromosome conformation capture (3C) based methods (Decker *et al.*, 2002; Dostie *et al.*, 2006; Simonis *et al.*, 2006; Lieberman-Aiden *et al.*, 2009; Stadhouders *et al.*, 2013). These have been shown to be a powerful tool in the interrogation of chromatin architecture, genomic interactions and hence loop formation events. 3C-based techniques led to the discovery of topologically associating domains (TADs) and, together with other techniques, have shown the role of TADs in gene expression and genome organization in development and disease (Spielman *et al.*, 2018).

We have previously developed Targeted Chromatin Capture (T2C) (Kolovos *et al.*, 2014; Kolovos *et al.*, 2018), a method to study the 3D chromatin organization and compartmentalization, at high coverage, high capture efficiency and high signal-to-noise ratio at sub-kbp resolution (0.4 kbp) (Kolovos *et al.*, 2016). Briefly, chromatin is crosslinked and lysed nuclei are digested (1st restriction enzyme digestion), followed by ligation of the DNA-fragments located in close proximity, and de-crosslinked. Upon the 2nd restriction enzyme digestion, the chromatin is sonicated briefly, allowing a reliable filtering of the PRC duplicates (Kolovos *et al.*, 2014; Brant *et al.*, 2016; Kolovos *et al.*, 2016; Kolovos *et al.*, 2018). That approach can be easily standardized between different laboratories. Alternatively, the 2nd restriction enzyme digestion could be replaced by shearing of the chromatin followed by ligating with the linkers. Subsequent steps include the hybridization with the oligonucleotides, followed by the identification and quantification of ligated fragments by sequencing and bio-informatic analysis to reveal the spatial organization of the genome.

In contrast to other methods (Mifsud *et al.*, 2015; Schoenfelder *et al.*, 2015; Oudelaar *et al.*, 2020), T2C's high coverage makes data binning unnecessary, improving T2C's resolution. The resolution of T2C is determined by the chosen 1st restriction enzyme, with the use of a 5-cutter enzyme (e.g. *Apo*I) generally offering sub-kbp resolution (see also Kolovos *et al.,* 2018) for the advantages and disadvantages of selecting a 6-, 5- or 4-base cutter restriction enzyme). Therefore, T2C offers absolute restriction fragment resolution maps of the interactome of genomic regions spanning from some Mb up to whole chromosomes, without the limitations in the interpretation and analysis of the data caused by the binning of reads (Kolovos *et al.*, 2014; Brant *et al.*, 2016; Kolovos *et al.*, 2016; Kolovos *et al.*, 2018; Birkhoff *et al.*, 2020). Thus, T2C can identify (sub-)TADs and their boundaries at high resolution, while it requires low sequencing depth (1/10 of a sequencing lane), significantly reducing the costs when compared to Hi-C.

Also, T2C allows the investigation of native chromatin architecture, avoiding the bias introduced by crosslinking of the chromatin (Brant *et al.*, 2016). Hence, T2C is an affordable method to unveil the interactome and genome architecture, and can be applied in a clinical environment to investigate the genotype- phenotype correlation of disorders linked to structural variants (SVs) or point mutations affecting chromatin conformation (Kolovos *et al.*, 2014; Brant *et al.*, 2016; Kolovos *et al.*, 2016; Kolovos *et al.*, 2018; Birkhoff *et al.*, 2020).

One of the limitations of several 3C-based techniques is the requirement for a large amount of input material. Thus, all studies involving clinically derived material, embryonic tissues or other scarce cell population would have been very challenging if not impossible. To address this issue, we developed a low-T2C protocol which allows the interrogation of the chromatin architecture and interactome. The current protocol is optimized for 10^5 cells focusing at a ~7.4 Mb-long region of chromosome-17 in mouse erythroleukemia (MEL) cells at a median resolution of 0.26 kbp.

4.3. Materials

4.3.1. Cell preparation

- *Culturing Medium:* DMEM medium with 10% (v/v) FBS, add 50 ml of FBS to 450 ml of DMEM medium.
- *Formaldehyde:* 37% in H_2O, with 10-15% Methanol as stabilizer.
- *Glycine (2.5 M in PBS):* Add 9.38g of Glycine to a 50 ml Falcon tube and add 30 ml of PBS, mix the solution well, and then add PBS to a final volume of 50 ml. Filter and store the solution at 4 °C for up to 1 year.
- *Ice-cold PBS:* Place a bottle of phosphate buffered saline at 4 °C.
- *Nonidet P-40 substitute (10% (v/v), NP-40):* Add 5 ml of NP-40 to 30 ml of MilliQ H_2O, mix the solution well, and then add MilliQ H_2O up to a final volume of 50 ml. IGEPAL® CA-630 may also be used.
- *Sodium chloride (5 M, NaCl):* Add 14.6g of NaCl to a 50 ml Falcon tube and add 30 ml of MilliQ H_2O, mix the solution well, and then add MilliQ H_2O up to a final volume of 50 ml. Filter and store the solution at 4 °C for up to 1 year.
- *Tris-HCl (1 M, pH 8.0):* Add 60.57g of Trizma base to a 500 ml sterile bottle and add 200 ml of MilliQ H_2O, mix the solution well, and then adjust the pH to 8.0 with HCl. Using MilliQ H_2O, bring the final volume to 500 ml. Filter and store the solution at 4 °C for up to 1 year.
- *Complete protease inhibitor (Merck):* Dissolve one tablet in 1 ml of PBS to create a 50× working solution. Store it at −20 °C for up to 3 months, avoid repeated freeze–thaw cycles.
- *Lysis buffer:* Mix 50 µl of 1 M Tris-HCl pH 8.0 (final concentration is 10 mM Tris, pH 8.0), 10 µl of 5 M NaCl (final concentration is 10 mM NaCl), 100 µl of 10% (v/v) NP-40 (final concentration is 0.2% (v/v) NP-40), 100 µl of 50× complete protease inhibitor (50×) and 4.74 ml of MilliQ H_2O to reach a final volume of 5 ml. Keep the solution on ice and use it immediately.

4.3.2. Enzymatic digestion

- *1.2x restriction buffer:* Mix 6 µl of restriction buffer with 44 µl of MilliQ water, for each sample.

- *SDS (20% (w/v)):* Add 10 g of SDS to 30 ml of MilliQ H_2O and mix the solution well. Then add MilliQ H_2O up to a final volume of 50 ml. Store the solution at room temperature for up to 1 year.
- *Triton X-100 (20% (v/v)):* Add 10 ml of Triton X-100 to 30 ml of MilliQ H_2O, mix the solution well, and then add MilliQ H_2O up to a final volume of 50 ml.
- *Tris-HCl (10 mM, pH 7.5):* Add 0.06 g of Trizma base to a 50 ml Falcon tube and add 20 ml of MilliQ H_2O, mix the solution well, and then adjust the pH to 7.5 with HCl. Using MilliQ H_2O, bring to a final volume of 50 ml. Filter and store the solution at 4 °C for up to 1 year.
- *RNase-A (10 mg/ml):* Dissolve 10 mg of RNase-A in 1 ml of 10 mM Tris-HCl, pH 7.5. Heat the solution to 100 °C for 5 min and cool to room temperature (20-24 °C). Store the solution at −20 °C for up to 1 year.
- *Proteinase-K (10 mg/ml):* Dissolve 10 mg of Proteinase-K in 600 µl of MilliQ H_2O, and fill up to 1 ml with MilliQ H_2O. Store the solution at −20 °C for up to 1 year.

4.3.3. Ligation

- *T4 DNA ligase (high concentrated)* and *ligation buffer*
- *Ligation buffer (1.15×):* Mix 541.6 µl of MilliQ H_2O and 70.4 µl of T4 DNA ligation buffer. Prepare fresh ligation buffer and keep it on ice.

4.3.4. DNA purification

- Phenol/chloroform/iso-amylalcohol (25:24:1, (v/v/v))
- *Glycogen (20 mg/ml)*
- *100% ethanol (UltraPure)*
- *Sodium acetate (2 M, pH 5.6):* Dissolve 20.5 g of sodium acetate (anhydrous) in 70 ml of MilliQ water. Adjust pH to 5.6 by adding glacial acetic acid. Fill up to 100 ml with MilliQ water. Store the solution at room temperature for up to 1 year.
- *Ethanol (70% (v/v):* Mix 35 ml of 100% (v/v) ethanol with 15 ml of MilliQ H_2O. Store the solution at 4 °C for up to 1 year.

4.3.5. Equipment

- Cell strainer, 40 µm
- 2 ml Pasteur pipettes
- 1.5 ml low DNA binding tubes
- 15 ml centrifuge tubes
- 50 ml centrifuge tubes
- Filter tips (for P1000, P200, P20, P10 pipettes)
- 0.2 µm syringe filters
- 10 ml and 50 ml syringes
- 360° Tube Rotator
- Centrifuge with cooling system, and rotors dedicated to 1.5 ml and 15 ml tubes
- Thermomixer
- Water bath
- 1 mL Syringes, 25G needles

4.4. Methods

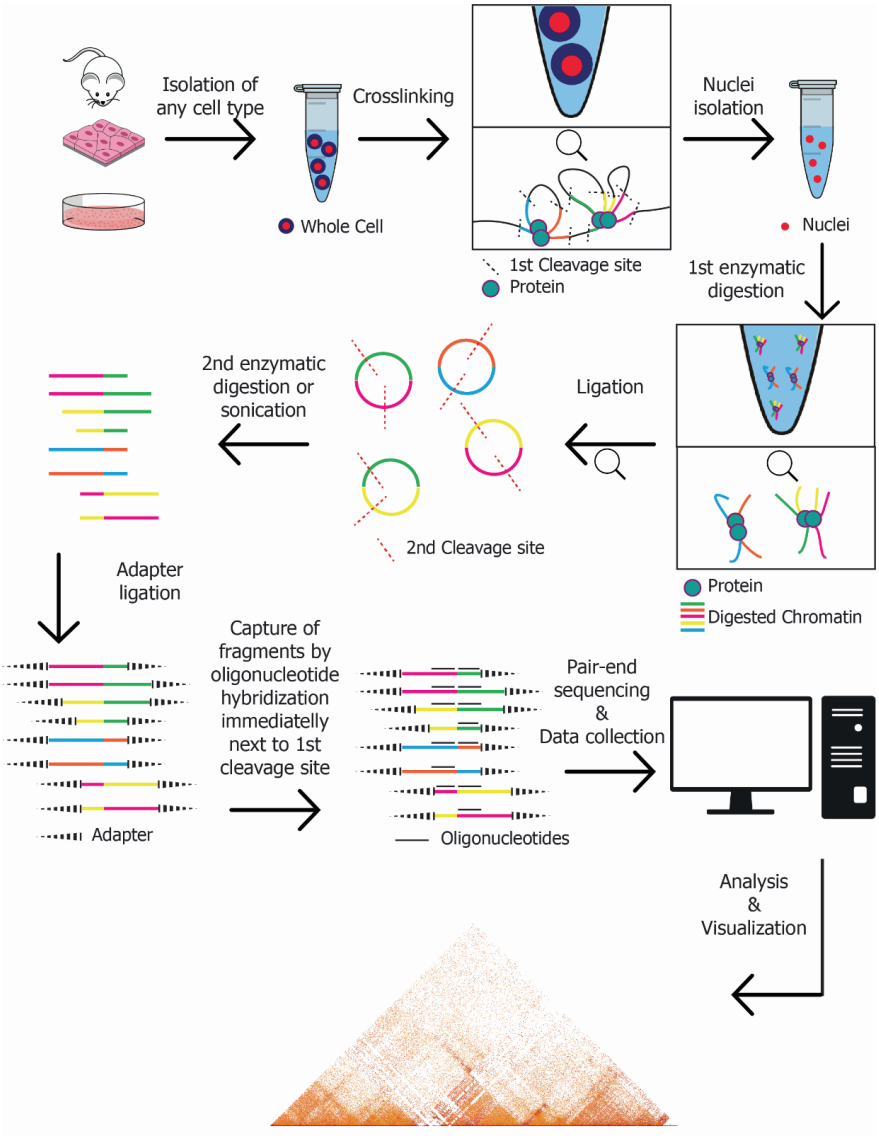


Figure 4.1. Overview of the entire T2C procedure, allowing to investigate spatial structure and proximities of the genome at sub-kb resolution, including for specific chromosome regions and key gene(s) of inetrest therein, at affordable sequencing cost.

Chromatin, after cross-linking of the cells, is extracted from nuclei and digested with chosen first restriction enzyme (1st enzymatic digestion). The obtained fragmented chromatin is then ligated under dilute conditions to promote intramolecular ligation, and then digested with the chosen second restriction enzyme (can also be sonicated). Such obtained fragments receive adapters by ligation (serving Illumina sequencing later) and the genomic sequences of interest are then captured by hybridization on oligo-arrays, and this enriched library is sequenced, and these data analyzed by bio-informatic tools and visualization of the genomic interactions, as described in Kolovos *et al.* (2014, 2018) and also as exemplified in **Fig. 4.5** below.

4.4.1. Crosslinking and nuclei isolation

1. Resuspend 10^5 of cells in 300 µl DMEM medium with 10% (v/v) FBS and transfer to 1.5 ml DNA low binding tube (*see* **Note 1**, section 4.5 below).
2. Add 8 µl of 37% (v/v) formaldehyde to each 1.5 ml tube (1% (v/v) final formaldehyde concentration), and incubate for 9 min at room temperature while rotating (*see* **Note 2**).
3. Add 17 µl of pre-cold 2.5 M Glycine to each tube and place them on ice. From this point, keep the tubes on ice.
4. Centrifuge the tubes for 8 min at 340 g at 4 °C and carefully discard the supernatant. The pellet may not be visible.
5. Wash the pellet by resuspending in 0.5 ml of ice-cold PBS, centrifuge the tubes for 8 min at 340 g at 4 °C. Repeat the wash one more time. **PAUSE POINT:** The pellet can be snap-frozen in liquid-N_2 and stored at -80 °C.
6. While the samples are spinning prepare fresh lysis buffer (*see* **Materials**, section 4.3 above) and keep it on ice (*see* **Note 3**).
7. Add 125 µl of lysis buffer to the pellet, and resuspend by gently pipetting up and down ten times (*see* **Note 4**).
8. Incubate the tubes on ice for 20 min.
9. Centrifuge the tubes for 5 min at 650 g at 4 °C and discard the supernatant. Pellet may be not visible.

4.4.2. First enzymatic digestion

10. Prepare 1.2x restriction buffer (*see* **Materials**) supplied with the *Apo*I restriction enzyme, and keep it on ice.
11. Resuspend the nuclei in 50 µl of freshly prepared 1.2x restriction buffer and place the tubes in a thermomixer for 10 min at 37 °C, while shaking at 900 rpm (*see* **Note 4**).
12. Add 1.5 µl of 10% (w/v) SDS and incubate samples for additional 50 min (*see* **Note 5**).
13. Add 5 µl of 20% (v/v) Triton X-100 and incubate for 1 h at 37 °C, while shaking at 900 rpm.
14. Collect 10 µl of suspension from each sample and transfer into new tubes. Store them at −20 °C (undigested controls, *see* **validation of digestion efficiency**, section 4.4.3 below).
15. Add 40 U of the *Apo*I restriction enzyme to the samples and incubate the tubes overnight in a thermomixer at 37 °C while shaking at 900 rpm (*see* **Note 6**).

4.4.3. Validation of digestion efficiency

16. Collect 10 µl of suspension from each sample and transfer into new tubes (digested controls). The rest of the sample should be placed at 4 °C.
17. Take the undigested control samples from Step 15 and thaw them on the bench.
18. Add 36 µl of 10 mM Tris, pH 7.5 to both undigested and digested samples and mix gently by flipping the tubes.
19. Add 4 µl of 10 mg/ml proteinase-K stock to each sample.
20. Place the tubes for 1 h in a Thermomixer at 65 °C while shaking at 900 rpm.
21. Cool down the samples by changing the temperature of the Thermomixer to 37 °C. Incubate the samples for 20 min at 37 °C.
22. Add 1.2 µl of 10 mg/ml RNase-A stock to each sample and incubate for 45 min at 37 °C in a thermomixer while shaking at 900 rpm.
23. Load each sample with the appropriate loading dye on a 1% (w/v) agarose gel (*see* **Note 7**) (**Fig. 4.2**).

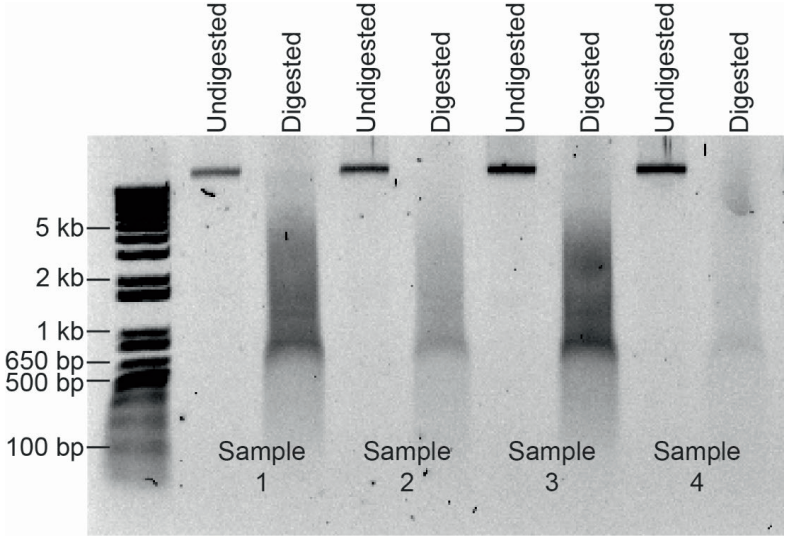


Figure 4.2. Quality control testing of digestion with *Apo*I.

1.0% agarose gel with undigested and *Apo*I-digested samples of four replicates. MEL cells were used as input material.

4.4.4. Ligation and de-crosslinking

24. Thaw the samples from step 16 and wait until they reach room temperature.
25. Add 4 µl of 20% (w/v) SDS and place the samples in a thermomixer for 20 min at 65 °C while shaking at 900 rpm.
26. Prepare 1.15× ligation buffer (*see* **Materials**).
27. Add 612 µl of 1.15× ligation buffer to each sample and mix the suspension by gently flipping the tubes.
28. Add 37.5 µl of 20% (v/v) Triton X-100 and incubate the tubes for 1 h in a thermoshaker at 37 °C, while shaking at 900 rpm (*see* **Note 8**).
29. Add 2 µl (10U) of high-concentrated T4 DNA-Ligase and incubate at 16 °C for 4 hours to overnight (*see* **Note 9**).
30. Add 3 µl of 10 mg/ml proteinase-K stock and incubate at 65 °C overnight (*see* **Note 10**).
31. Add 3 µl 10 mg/ml of RNase-A and incubate samples in a thermomixer for 30-45 min at 37 °C, while shaking at 900 rpm.

4.4.5. DNA purification after ligation

32. Remove the samples from the thermomixer and wait until they reach room temperature.
33. Add 700 µl of phenol/chloroform/isoamyl alcohol (25:24:1) and vortex for 1 min. The suspension should achieve a milky white color (*see* **Note 11**).
34. Centrifuge the samples for 5 min at 16.000 g at room temperature.
35. Transfer the upper phase (~700 µl) to a 15 ml tube and add the 700 µl of MilliQ H_2O, mix the solution by gently pipetting up and down four times.
36. Add 1 µl of glycogen, 150 µl of 2M Sodium Acetate pH 5.6 and 3.5 ml of 100% ethanol, mix the solution by flipping the tube.

37. Place the tubes on ice for 1 h, and move the samples to -80°C for an additional 2 h. PAUSE POINT: The samples may be stored in -80°C for up to 2 days.
38. Centrifuge the tubes for 45 min at 3200 g at 4 °C.
39. Discard the supernatant and wash the pellet by adding 1 ml of 70% ethanol (*see* **Note 12**).
40. Centrifuge the tubes for 15 min at 3200 g at 4 °C.
41. Discard the supernatant, and centrifuge the tubes for 1 min at 3200 g at 4 °C.
42. Remove supernatant and air-dry the pellet for ~20 min (*see* **Note 13**).
43. Resuspend the DNA in 15 µl of MilliQ H_2O and incubate the tubes for 15 min at 37 °C, mix the samples by pipetting gently up and down five times.
44. Transfer the DNA solution to new 1.5 ml low binding DNA tube. PAUSE POINT: The DNA suspension may be stored in -20 °C for up to 1 year.
45. To validate ligation efficiency, take 1 µl of sample and load it in a 1.0 % (w/v) agarose gel (**Fig. 4.3**).
46. To assess the amount of obtained DNA, use the Qubit fluorometer (*see* **Note 14**).

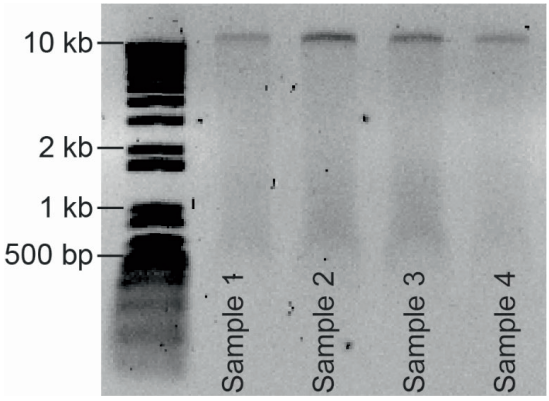


Figure 4.3. Quality control testing of ligation.

1.0% agarose gel with ligated samples of four replicates. MEL cells were used as Input material.

4.4.6. Second enzymatic digestion and DNA purification

47. Take the tubes from step 44 and add 31 µl of MilliQ H_2O, mix the suspension by flipping the tube.
48. Add 5 µl of 10x *Dpn*II restriction buffer, mix the suspension by flipping the tube.
49. Add 1 µl of *Dpn*II restriction enzyme (*see* **Note 15**).
50. Place the tubes in thermomixer and incubate overnight at 37 °C, while shaking at 500 rpm.
51. Add 450 µl of MilliQ H_2O and mix the suspension by flipping the tubes five times.
52. Add 500 µl of phenol/chloroform/isoamyl alcohol (25:24:1) and mix vigorously for 1 min or until the suspension achieve a milky white color (*see* **Note 11**).
53. Centrifuge the samples for 5 min at 16000 g at room temperature.
54. Transfer the upper phase to a new DNA low binding 1.5 ml tube and add 1 µl of glycogen.
55. Add 36 µl of 2M sodium acetate pH 5.6, and 850 ul of 100% ethanol.

56. Mix the solution gently, by inverting the tubes, and place the samples on ice for 1 h, then transfer the tubes for an additional 2 h to -80°C.
57. Centrifuge the tubes for 20 min at 15800g at 4 °C.
58. Discard the supernatant, and add 1 ml of 70% ethanol (*see* **Note 12**).
59. Centrifuge the tubes for 5 min at 15800 g at 4 °C.
60. Remove the supernatant, air-dry the pellet for ~15 min (*see* **Note 13**).
61. Add 20 µl of MilliQ H_2O and incubate suspension for 10-15 min in a thermomixer at 37 °C.
62. Load 1 µl of sample in a 1.5 % (w/v) agarose gel to validate the digestion efficiency (*see* **Note 16**) (**Fig. 4.4**).

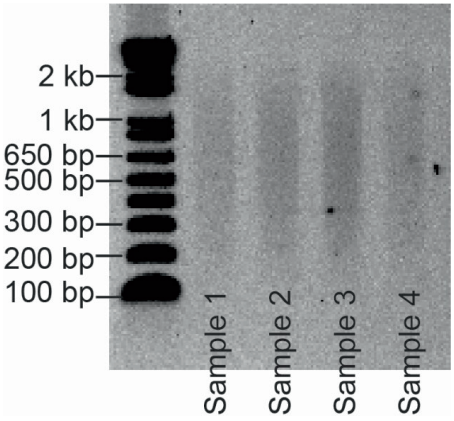


Figure 4.4. Quality control testing of digestion with *Dpn*II.

1.5% agarose gel with *Dpn*II-digested samples of four replicates. MEL cells were used as input material.

4.4.7. Library preparation and sequencing

63. The library preparation and sequencing were performed according to the standard T2C protocol (Kolovos *et al.*, 2018). Example of low-T2C analysis is depicted in **Fig. 4.5**.

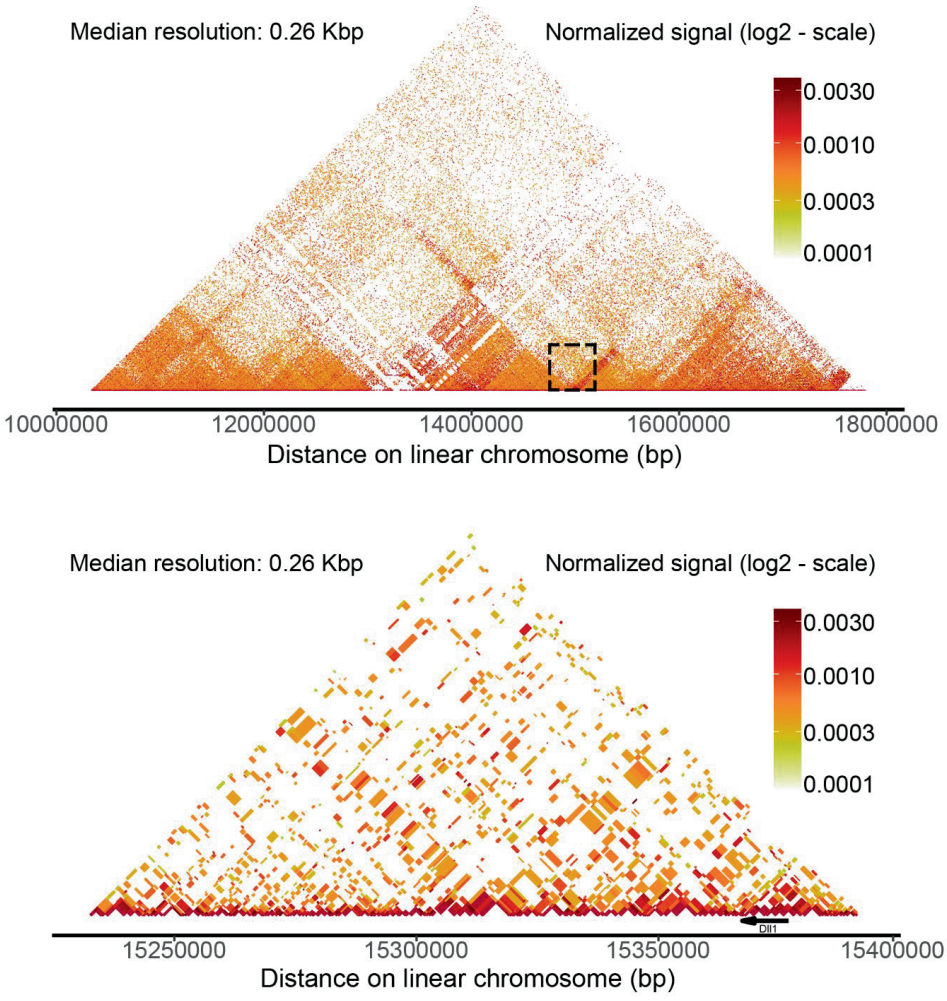


Figure 4.5. Upper panel: T2C interaction map for a ~7.4 Mb-long genomic region on chr-17. The T2C interaction map for the genomic region chr17:10337000-17800000 (version mm10) at a median resolution of 0.26 kbp, displays the TADs and sub-TADs together with their boundaries and intra- and inter-interactions. **Lower panel**: T2C interaction map for a ~0.16 Mb-long genomic region on chr-17 depicting the region from the dashed square box from upper panel. The T2C interaction map for the genomic region chr-17: 15232053 - 15392070 (mm10) at a median resolution of 0.26 kbp, unveils the interactome of the genomic region encompassing the *Dll1* gene.

4.5. Notes

1. When working with *in vivo* material, it is recommended to pass the dissociated cells through a 40-µm cell strainer to achieve a single cell suspension. If *in vivo* material is hard to dissociate and contains cell clusters pass the solution through a syringe equipped with a G25 needle five times. Validate the single cell suspension under a microscope. If clusters of cells are still present, gently pipette ten times and then check whether it has become a single cell suspension.
2. The 37% formaldehyde is toxic and should be used with great care inside a fume cabinet or a hood.
3. Stored lysis buffer may reduce cell lysis and affect the end result. Always use freshly prepared lysis buffer. Calculate the volume of the buffer, needed to be prepared, according to the number of samples in each experiment.
4. The pellet is usually sticky and may be difficult to resuspend. Ensure that the pellet is properly resuspended by flipping the tube a couple of times and by checking the clarity of the suspension under a light source.
5. During the incubation, check for the presence of clumps of nuclei and gently flip the tubes if clumps are visible.
6. It is recommended to use highly concentrated restriction enzymes (preferably 50 U/µl). The optimal temperature for digestion with *Apo*I is higher than 37 °C, however, it is recommended to be used at 37 °C to avoid partial de-crosslinking of the sample.
7. The undigested control should be visible as a single high molecular weight band (single band). The digested control should give a smear. Continue with the next steps, if digestion was efficient, or repeat steps 1-15, if digestion was unsuccessful.
8. It is recommended to additionally mix the suspension by flipping the tubes 3-5 times by every 10 min as it ensures the inactivation of SDS by Triton X-100. Improper inactivation of SDS may lead to failure of the ligation.
9. Use a thermomixer with cooling system or a water bath inside a cold room and adjust temperature to 16 °C.
10. Use a thermomixer without shaking or a water bath.
11. Phenol/chloroform/isoamyl alcohol (25:24:1) is toxic and should be used in a fume cabinet or under a hood.
12. Remove the supernatant gently and slow as the pellet can slide and get lost.
13. To prevent over-drying the pellets monitor the tubes every 5 min.
14. It is recommended to use the dsDNA HS Assay kit with 1µl of DNA solution as input.
15. It is recommended to use high-concentration restriction enzyme stocks. The fragment resolution of the T2C, as well as the quality of data, depends on the choice of the appropriate restriction enzyme. This protocol was optimized with the use of *Dpn*II as the second restriction enzyme. It can be replaced by shearing (Covaris) followed by ligating with the linkers.
16. The digested DNA should give a smear with majority of fragments below 1 kbp.

Acknowledgments

We would like to thank the members of Huylebroeck and Grosveld laboratories for the insightful comments and helpful discussions, and the members of the Center for Biomics at Erasmus Medical Center for their technical assistance. PT and KN are partially supported by the UMO-2015/19/B/NZ4/03184 grant.

Literature references

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, Nowosad K, van den Hout MCGN, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. Hum Mol Genet. 2020 Aug 29;29(15):2535-50.

Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, Grosveld FG, Marenduzzo D, Papantonis A. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. Mol Syst Biol. 2016 Dec 9;12(12):891.

Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002 Feb 15;295(5558):1306-11.

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006 Oct;16(10):1299-309.

Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, Imam AMA, van de Werken HJG, Wendt KS, Knoch TA, van IJcken WFJ, Grosveld F. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. Nat Protoc. 2018 Mar;13(3):459-77.

Kolovos P, Georgomanolis T, Koeferle A, Larkin JD, Brant L, Nikolicć M, Gusmao EG, Zirkel A, Knoch TA, van Ijcken WF, Cook PR, Costa IG, Grosveld FG, Papantonis A. Binding of nuclear factor κB to noncanonical consensus sites reveals its multimodal role during the early inflammatory response. Genome Res. 2016 Nov;26(11):1478-89.

Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, Wendt KS, van IJcken WF, Grosveld F, Knoch TA. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. Epigenetics Chromatin. 2014 Jun 16;7:10.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015 Jun;47(6):598-606.

Oudelaar AM, Beagrie RA, Gosden M, de Ornellas S, Georgiades E, Kerry J, Hidalgo D, Carrelha J, Shivalingam A, El-Sagheer AH, Telenius JM, Brown T, Buckle VJ, Socolovsky M, Higgs DR, Hughes JR. Dynamics of the 4D genome during *in vivo* lineage specification and differentiation. Nat Commun. 2020 Jun 1;11(1):2722.

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, Dimitrova E, Dimond A, Edelman LB, Elderkin S, Tabbada K, Darbo E, Andrews S, Herman B, Higgs A, LeProust E, Osborne CS, Mitchell JA, Luscombe NM, Fraser P. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 2015 Apr;25(4):582-97.

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006 Nov;38(11):1348-54.

Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018 Jul;19(7):453-67.

Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protoc. 2013 Mar;8(3):509-24.

Chapter 5

Chromatin architecture and *cis*-regulatory landscape of the *DACT2-SMOC2* locus in the developing synovial joint

Karol Nowosad [1,2,3], Ewa Hordyjewska-Kowalczyk [2], Aneta Malesa [2,4], Adrian Odrzywolski [5,6], Rutger W. W. Brouwer [1,7], Petros Kolovos [8], Ilias Boltsis [1], Judith C. Birkhoff [1], Wilfred F. J. van IJcken [1,7], Frank G. Grosveld [1], Andrea Conidi [1], Danny Huylebroeck [1], Przemko Tylzanowski [2,9,*]

[1] Department of Cell Biology, Erasmus University Medical Center, 3015 CN Rotterdam, the Netherlands;
[2] Department of Biomedical Sciences, Laboratory of Molecular Genetics, Medical University of Lublin, 20-093 Lublin, Poland;
[3] The Postgraduate School of Molecular Medicine, Medical University of Warsaw, 02-291 Warsaw, Poland;
[4] Department of Development and Regeneration, Stem Cell Institute, KU Leuven, 3000 Leuven, Belgium;
[5] Department of Biochemistry and Molecular Biology, Medical University of Lublin, 20-059 Lublin, Poland;
[6] Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium;
[7] Center for Biomics-Genomics, Erasmus University Medical Center, 3015 CN Rotterdam, The Netherlands;
[8] Department of Molecular Biology and Genetics, Democritus University of Thrace, 68100 Alexandroupolis, Greece;
[9] Department of Development and Regeneration, Skeletal Biology and Engineering Research Center, KU Leuven, 3000 Leuven, Belgium

Submitted for publication

Available as preprint:

https://www.biorxiv.org/content/10.1101/2022.10.06.511134v1

5.1. Summary

Synovial joints form in several steps, starting with the formation of an interzone, a condensation of mesenchymal cells at the sites of prospective joints. Despite the identification of multiple factors essential for formation of interzone, little is known about the regulation of their spatio-temporal gene expression during that process in limb development. Here, we investigated the *cis*-regulatory landscape of the Wnt-modulator encoding genes *DACT2* and *SMOC2,* both expressed in the forming joint interzone. Mechanically collected interzone and phalange samples, from chick embryos were found to express acknowledged marker genes (*GDF5* and *MATN1*), as well as *DACT2* and *SMOC2*. Using Targeted Chromatin Capture (T2C) we characterized the 3D chromatin structure of a ~3.45 Mb-long region encompassing *DACT2* and *SMOC2*, which revealed differences at sub-TAD level between interzones and phalange. We identified candidate enhancers (CEs) based on H3-histone marks (H3K427ac and H3K4me1) located in close proximity to the promoters of *DACT2* and *SMOC2*, and further documented these CEs in a zebrafish enhancer assay. Our approach yields new insight into the regulation, in dynamic chromatin context, of two Wnt-signaling modulatory genes during synovial joint induction.

5.2. Introduction

The developing vertebrate limb is a frequently used model system to study the genetic and molecular control mechanisms of tissue/organ induction and subsequent patterning. Limb skeletal elements are derived from lateral plate mesoderm, while the limb muscle components originate from cells of the dermomyotome, a part of the transient, segmented somite formed in the paraxial mesoderm. During limb outgrowth, the mesenchymal cells undergo condensation followed by chondrogenic differentiation, resulting in the formation of a transient cartilage scaffold. The concomitant tissue patterning along the three limb axes determines the location and shape of the future bones (Akiyama *et al.*, 2005; Bi *et al.*, 1999; Petit *et al.*, 2017). At the onset of chondrogenic differentiation, these cells express *COL2A1* (α(II)-collagen) and *MATN1* (Matrilin-1) (Hyde *et al.*, 2007), followed by *COL10A1* (α(X)-collagen), which is specific for hypertrophic chondrocytes (Zheng *et al.*, 2003). Eventually, this hypertrophic cartilage becomes vascularized and is replaced by bone (Mackie *et al.*, 2008). In contrast, the cells from the interzone region, located between the ends of future skeletal elements, are involved in the development of synovial joint structures, which include articular cartilage, menisci, ligaments and the synovium itself (Archer *et al.*, 2003; Pacifici *et al.*, 2006; Ray *et al.*, 2015; Chijimatsu and Saito, 2019). Interzone cells express acknowledged marker genes such as *GDF5* (Growth and Differentiation Factor-5), *ATX/ENPP2* (Autotaxin), *WNT9B* (Wnt Family Member 9B) and *ERG* (an ETS family transcription factor) (Dhordain *et al.*, 1995; Hartmann and Tabin, 2001; Ray *et al.*, 2015).

Significant efforts have been made to analyze the transcriptome of developing synovial joints (Jenner *et al.*, 2014; Pazin *et al.*, 2014; Feng *et al.*, 2019; Bian *et al.*, 2020; Feng *et al.*, 2019), but enhancer-driven regulation of gene expression during joint formation remains under-explored. Our laboratory studies the induction of the synovial joint. As part of these efforts, we study the role of a secreted Wnt/BMP signaling modulator, SMOC2, isolated from articular cartilage (Mommaerts *et al.*, 2014; Peeters *et al.*, 2018; Lu *et al.*, 2019; Long *et al.*, 2021). *SMOC2* mRNA is expressed in several tissues, including in the developing synovial joint of the E14.5

mouse embryo (Diez-Roux *et al.*, 2011). *DACT2* (Dishevelled binding Antagonist of β-catenin-2), one of the *SMOC2* neighboring genes on the same chromosome, is also expressed in the interzone region (Sensiate *et al.*, 2014). The co-localization of these two genes in the genome is conserved among human, mouse, chicken and zebrafish, with species-specific differences of the length of the intergenic region (from 75 to 150 kb). DACT2 negatively regulates Wnt-β-catenin signaling by disrupting the β-catenin:LEF1 complex in the nucleus (Wang *et al.*, 2015). DACT2 also modulates YAP/TAZ signaling by preventing nuclear accumulation of Yes-Associated Protein (YAP) (Tan *et al.*, 2017), a transcription factor (TF) involved in the negative regulation of the gene encoding the BMP-subgroup ligand GDF5 (Kania *et al.*, 2020). Since both *SMOC2* and *DACT2* are co-expressed during joint formation, and are genomically separated in a head-to-head configuration by an intergenic region, we hypothesized that they may share transcriptional regulatory elements.

Distant genomic enhancers co-control target gene expression in spatiotemporal manner, orchestrating cell-type specific pattern of genes (Osterwalder *et al.*, 2018). Enhancers co-regulate RNAPol2-based gene transcription by bridging TFs (and co-factors of the latter) with the promoter-proximal region of their target gene(s), likely via DNA-looping (Grubert *et al.*, 2020). Importantly, the organization of chromatin architecture into topologically associating domains (TADs), delineated by borders enriched for DNA-binding sites of the multiple zinc-finger protein CTCF, was shown to promote intra-TAD enhancer-promoter contacts and insulate inter-TAD interactions (Zuin *et al.*, 2014; Dixon *et al.*, 2016; Fudenberg *et al.*, 2016; Krefting *et al.*, 2018). Therefore, enhancers function mainly in *cis* and within TADs. Such enhancer activity also correlates with overall chromatin accessibility, established in part by nucleosome positioning and dynamics, and is secured by histone modifications (e.g., H3 acetylation (ac) and methylation (me), such as H3K27ac and H3K4me1 signatures) (Calo and Wysocka, 2013).

Chromosome conformation capture (3C) and its derivatives (4C, 5C, Hi-C) are powerful tools for investigating enhancer-promoter interactions in higher-order 3D chromatin context (Tolhuis *et al.*, 2002; Lieberman-Aiden *et al.*, 2009; Lupiáñez *et al.*, 2015; Mifsud *et al.*, 2015). These techniques however provide low to moderate resolution only, or require large or deep sequencing efforts. A recently developed technique, i.e. Targeted Chromatin Capture (T2C), a 3C-based variant, addresses these limitations and permits to obtain high-resolution data at affordable sequencing cost (Kolovos *et al.*, 2014; Kolovos *et al.*, 2018; Birkhoff *et al.*, 2020; Boltsis *et al.*, 2021). Another advantage of T2C is the targeted enrichment of short genomic regions of interest by hybridization with custom oligonucleotide probes, which increases signal-to-noise ratio.

Recently, we developed an atlas of candidate enhancers (CEs) active in interzone and adjacent phalange, and associated these CEs with genes upregulated in the interzone (Nowosad *et al.*, 2022). We also linked such CEs to genes known as crucial in synovial joint hypermobility and osteoarthritis, as well as phalange malformations (Nowosad *et al.*, 2022). This CE atlas is serving as resource for identifying and validating enhancer-controlled synovial joint and phalange formation, but also provided us a new starting point to study the regulation of *DACT2* and *SMOC2* expression at the onset of joint formation. Using a combination of integrative analysis of chromatin 3D architecture, CEs defined by enrichment of histone modifications (H3K27ac and H3K4me1), and an enhancer assay in zebrafish larvae, we report here the identification of multiple CEs associated with *DACT2*, one of which is shared with *SMOC2*, in the developing joint.

5.3. Results

5.3.1. Micro-dissection of interphalangeal joint interzones

To collect the samples for T2C, we dissected and separated joint interphalangeal interzones and the adjacent proximal part of phalange from digit-3 of the chick embryo hindlimb at HH32 (*see* Experimental procedures; Nowosad *et al.*, 2022). Validation of the purity of the dissected tissues was done by RT-qPCR analysis of acknowledged marker mRNAs for interzone (*GDF5*) and phalange (*MATN1*), and showed that both samples were successfully dissected and separated (**Fig. 5.1A**).

Previously, we carried out RNA-Sequencing (RNA-Seq) on such samples and showed significant upregulation of steady-state *DACT2* mRNA in dissected interzones as compared to phalanges, whereas *SMOC2* mRNA displayed high variation between biological replicates of interzone samples (Nowosad *et al.*, 2022). However, a trend towards upregulation in interzone tissue was observed (**Fig. S5.1;** see also Nowosad *et al.*, 2022).

RT-qPCR of the dissected tissues used in the present study showed significantly higher expression of both *DACT2* and *SMOC2* in dissected interzones as compared to phalanges (**Fig. 5.1B**). To document the *Smoc2* expression domain, we additionally performed whole-mount (**Fig. 5.1C**) and dual *in situ* hybridization (using also *Gdf5* anti-sense probe; **Fig. 5.1D**) in forelimb digits from E14.5 mouse embryos. The *Smoc2* expression was found to increase in the interzone, and partially overlapped with joint-specific *Gdf5* transcripts, but extended distally, beyond the *Gdf5* mRNA expression domain. In summary, *Smoc2* mRNA is expressed in the developing synovial joint.

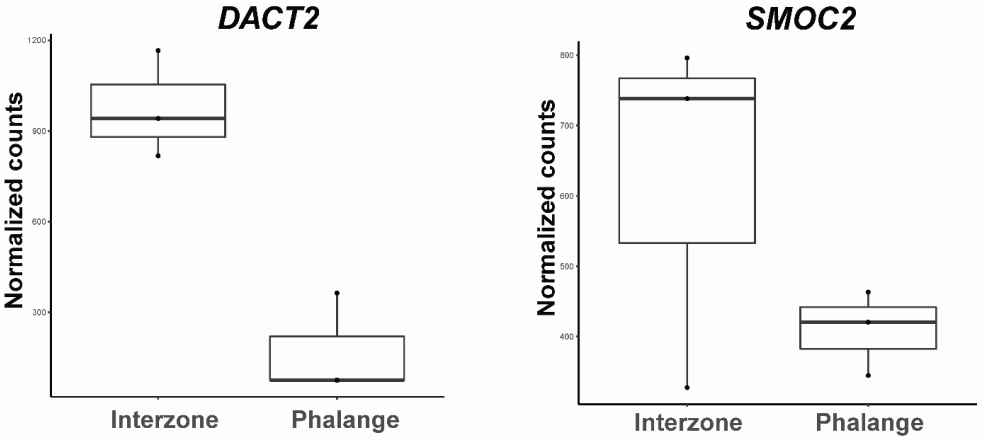


Figure S5.1. Normalized levels of expression of *DACT2* and *SMOC2* transcripts as determined by RNA-seq of interzone and phalange RNA samples, respectively.

The counts have been normalized to sequencing depth using DEseq2 (Love *et al.*, 2014).

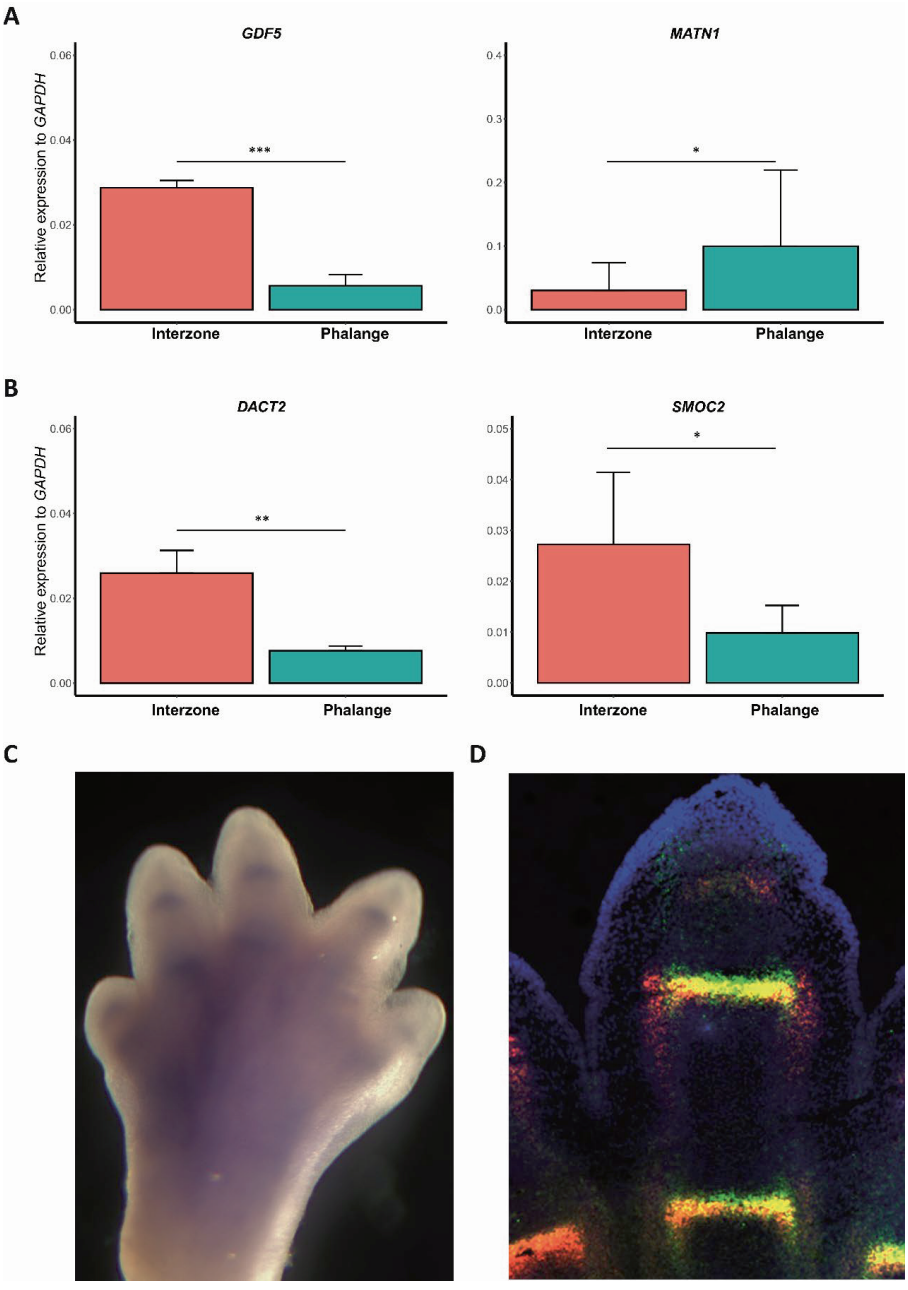


Figure 5.1. Expression of *DACT2*, *SMOC2 and* interzone (*GDF5*) and phalange (*MATN1*) marker genes.

(A) The *GDF5*, *MATN1*, and **(B)** *DACT2*, *SMOC2* steady-state mRNA levels, determined by RT-qPCR, after dissection of phalanges and interzones from chick embryo hindlimb digit-3 at stage HH32 (see Experimental procedures). All data was normalized to expression of *GAPDH*. The p-value is based on T-test. **(C)** Whole-mount *in situ* hybridization on the autopod of a E14.5 mouse forelimb. The blue staining depicts *Smoc2* mRNA expression. **(D)** Dual *in situ* hybridization in a digit of a E14.5 mouse forelimb. In red: *Gdf5* mRNA, green: *Smoc2* mRNA, yellow: overlap between *Gdf5* and *Smoc2* expression, blue: nuclear counterstain using DAPI.

5.3.2. Chromatin organization of the *DACT2–SMOC2* genomic region during joint formation

Following the respective tissue isolation, the interzone and phalange 3D chromatin structure within a ~3.45 Mb region (chicken chr3: 40,15-43,6 Mb) encompassing *DACT2* and *SMOC2* was determined by low-T2C, a protocol optimized for lower cell numbers (Boltsis *et al.*, 2021). First, we validated the efficiency of targeted fragment enrichment within the region of interest by quantification of paired reads mapped to both the whole genome and the target region. This showed that such reads for this target region were enriched in both tissues (**Table S5.1)**. Next, we removed the self-ligated/non-digested regions and calculated the *cis*/*trans* interactions (**Fig. S5.2A**) and fragment density distribution together with median resolution (**Fig. S5.2B**). The analysis of *cis* and *trans* interactions revealed that the T2C data contained much more *cis* interactions, further suggesting that these data are of good quality. The comparison of fragment length distributions showed that these did not change with tissue, while the median resolution (~570 bp) revealed that both interzone and phalange T2C datasets were generated at sub-kb resolution.

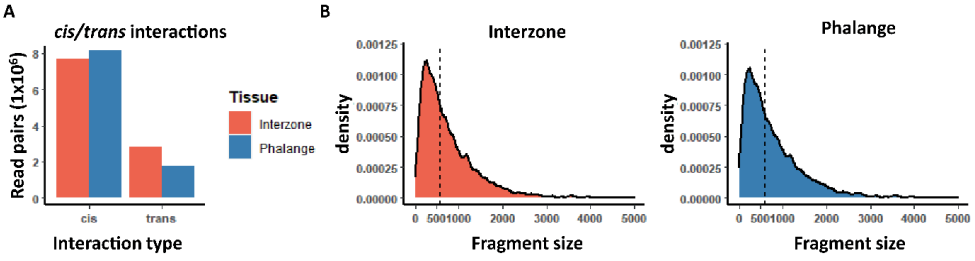


Figure S5.2. T2C Quality metrics.

(A) Number of *cis* and *trans* targeted read-pairs in the proximity matrix. **(B)** Density plot of fragment distribution within the *DACT2-SMOC2* 3.45 Mb-long genomic region. The dashed line represents median fragment size.

The resulting T2C interaction maps for the chicken *DACT2-SMOC2* region (**Fig. 5.2,** left panel) confirmed the global organization of the chicken genome into TADs, in line with a recent Hi-C study (Fishman *et al.*, 2019). The maps also confirmed the hypothesis that TADs are conserved among tissues, in this case between interzone and phalange. The quantification of unique fragment pair proximities within the region of interest showed that the interzone dataset contains a higher number of fragment pairs when compared to the phalange dataset (**Table S5.1**). This was reflected in more pronounced interzone-specific patterns of fragment proximities in the T2C interaction maps compared to those in phalange, which presented with a relatively more diffused pattern of such proximities (**Fig. 5.2**, with the entire *DACT2-SMOC2* 3.45 Mb-long region in the left panel, and a zoomed-in map for a 0.7 Mb segment in the right panel). To characterize the intra-TAD chromatin organization, we analyzed the chicken chr3:41.6-42.3 Mb zoomed-in region (~0.7 Mb) showing differences between interzone and phalange within this region (**Fig. 5.2**, right panel). The divergent pattern of the intra-TAD interaction between the two dissected tissues is most likely caused by DNA-looping of tissue-specific regulatory elements to their target gene/s, for instance *DACT2* and *SMOC2*, which is our focus for further studies here.

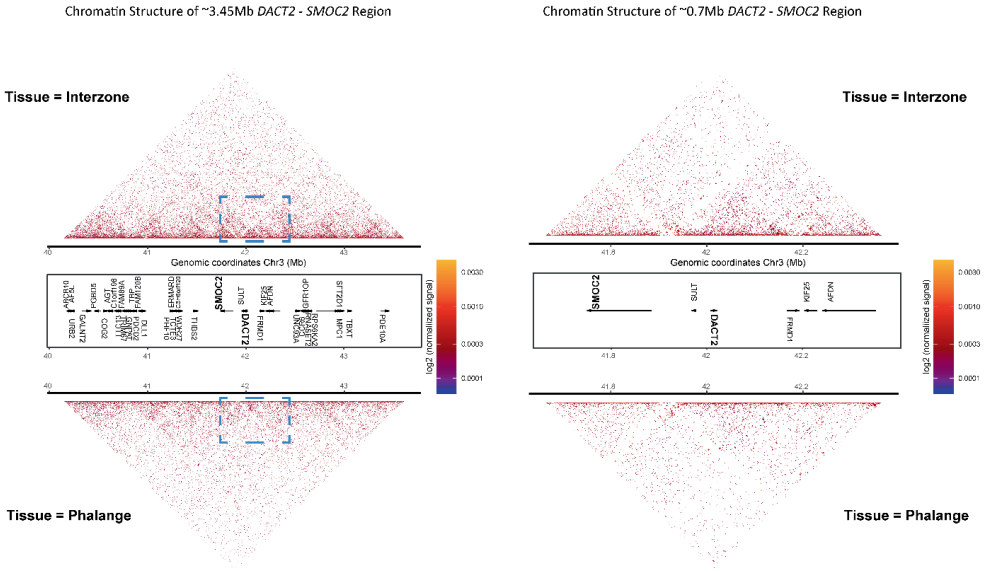


Figure 5.2. Characterization of the 3D chromatin structure of *DACT2-SMOC2* genomic region as determined by T2C.

(left panel) T2C-interaction map (restriction fragment resolution) for the ~3.45 Mb-long *DACT2-SMOC2* genomic region (chicken chr3: 40,15-43,6 Mb) for separated interzone tissue (upper triangles in left and right panel) and phalange (lower triangles). The dashed rectangle marks the zoomed region shown in the right panel B. **(right panel)** The reconstruction of the T2C map for the chicken chr3: 41,6-42,3 (~0.7 Mb-long) genomic region presents the chromatin architecture at inter-TAD level. Tissues are marked as in the left panel.

We extracted all genomic regions located in spatial proximity to the gene promoters (defined as a 5 kb-long segment, i.e. from -2.5/+2.5 kb flanking the transcriptional start site, TSS), and averaged the signal from T2C to the length of the promoter region by binning the data, using a bin size of 5 kb (**Table S5.1**). Subsequently, we quantified the number of *cis*-proximities and showed that the promoter of *DACT2* and *SMOC2* presented higher percentages of tissue-specific interactions as compared to common interactions, and this in both dissected tissues (**Fig. 5.3**). Interestingly, for both genes the ratio of tissue-specific vs. consensus *cis*-proximities was higher for the interzone than for the phalange (**Fig. 5.3**).

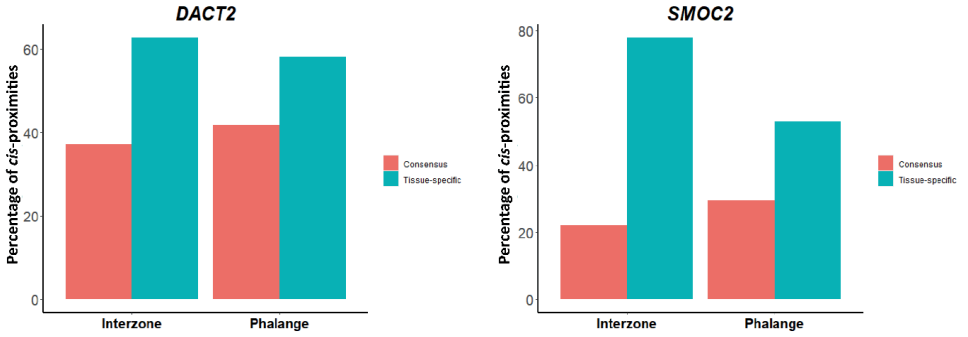


Figure 5.3. Quantification of fragments located in close proximity to *DACT2* and *SMOC2* promoters.

The genomic regions located in spatial proximity to the gene promoters (-2.5/+2.5 kb from TSS) were quantified for *DACT2* **(left)** and *SMOC2* **(right)**. Further, *cis*-proximities have been divided into separated groups (tissue specific, consensus for both tissues) and presented as a percentage of all interactions identified per tissue.

5.3.3. Identification of *DACT2* and *SMOC2* candidate enhancers

Next, to identify CEs in the ~3.45 Mb-long *DACT2-SMOC2* region, we screened our CEs atlas (Nowosad *et al.*, 2022) for those enhancers that are active in interzone and phalange. This analysis showed either low or no enrichment of *cis*-proximities between CEs and promoters of *DACT2* and *SMOC2*. However, this can be explained by the fact that our enhancer atlas contains only CEs that are conserved between chicken, human and mouse. Therefore, we had to apply a less stringent species conservation context for the analysis of CEs in the study here. We subsequently reanalyzed the original species-specific ChIP-Seq profiles of H3K27ac and H3K4me1 that were used to define the CEs in our atlas (Nowosad *et al.*, 2022). Using this strategy for the ~3.45 Mb-long *DACT2-SMOC2* region, we identified 53 interzone CEs and 35 phalange CEs, defined here as regions enriched for both H3 marks (**Table S5.1**). To further characterize chicken CEs, we carried out differential analysis of H3K27ac marks to investigate which of these CEs are located in differentially acetylated regions (DacRs). This revealed that the ~3.45 Mb-long *DACT2-SMOC2* region in chicken contains 59 DacRs, 45 of which were highly enriched in interzone as compared to phalange (**Table S5.1**). H3K27ac is a mark for active enhancers, and the majority of DacRs were enriched in interzone, suggesting that the *DACT2-SMOC2* region contains more differentially active enhancers in the interzone than phalange region. Importantly, intersection of CEs with DacRs identified that 33 CEs in interzone and 19 CEs in phalange, respectively, were located in DacRs (**Table S5.1**).

To identify CEs that regulate *DACT2*, we screened all interacting regions with *DACT2* promoter and intersected them from the aforementioned 88 CEs (53 interzone + 35 phalange). This permitted to select 7 CEs with a high T2C-score and enrichment for both H3 marks as well (**Fig. 5.4A**; *see* also **Fig. S5.3**; **Table S5.1**). Three of these (named CE1, CE3, CE5) were mapped in proximity to *DACT2* exclusively in interzone sample (**Fig. S5.3**, right panel), whereas the remaining other 4 (named CE2, CE4, CE6, CE7) were found close in 3D space to the *DACT2* promoter in both tissues (**Fig. 5.4B**; **Fig. S5.3**, both panels). Intersecting these selected CEs with DacRs revealed that CE1-4 and CE7 were located in DacRs, whereas CE5-6 did not present with significant differences in this acetylation level (**Fig. 5.5A**).

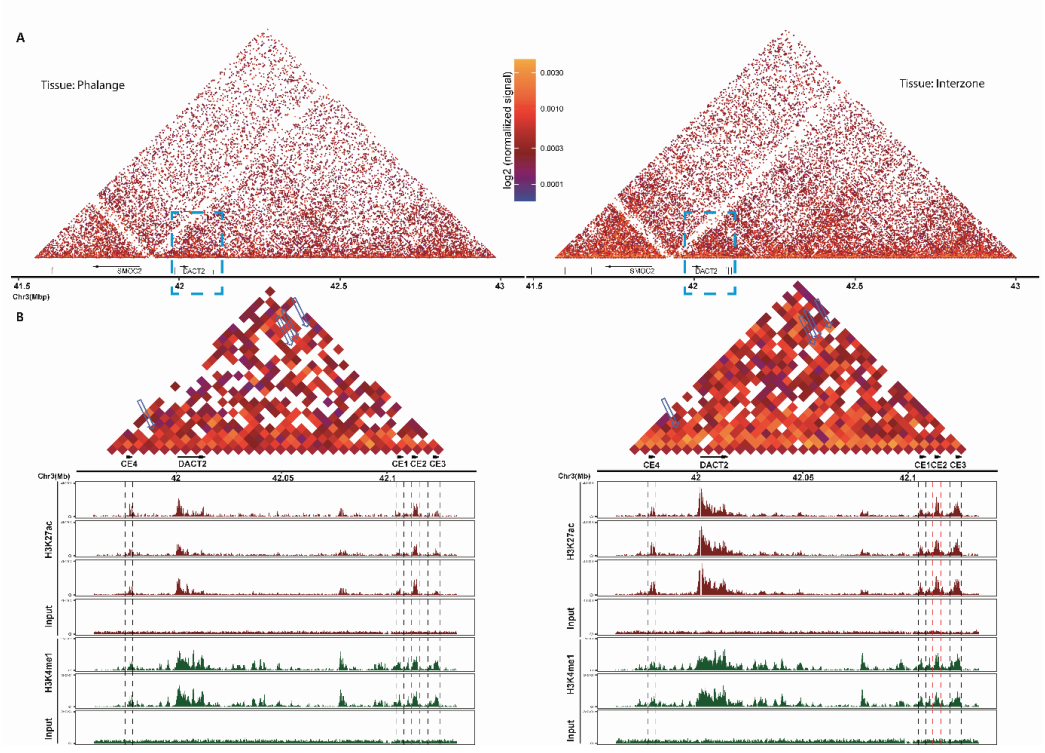


Figure 5.4. Identification of *DACT2* CEs.

(A) T2C interaction maps (bin size = 5 kb) encompassing identified candidate enhancers (CEs) highlighted by vertical lines. Dashed rectangles mark the zoomed regions visualized in the lower panel. **(B)** Zoom-in of the panel A for region encompassing *DACT2* and CE1, CE2, CE3 and CE4. *Cis*-proximities between these are indicated by blue arrowheads. Dashed lines show selected CEs regions enriched in histone modification. The H3K27ac tracks are highlighted in dark red. The H3K4me1 tracks are marked by dark green. ChIP-Seq inputs are marked by black.

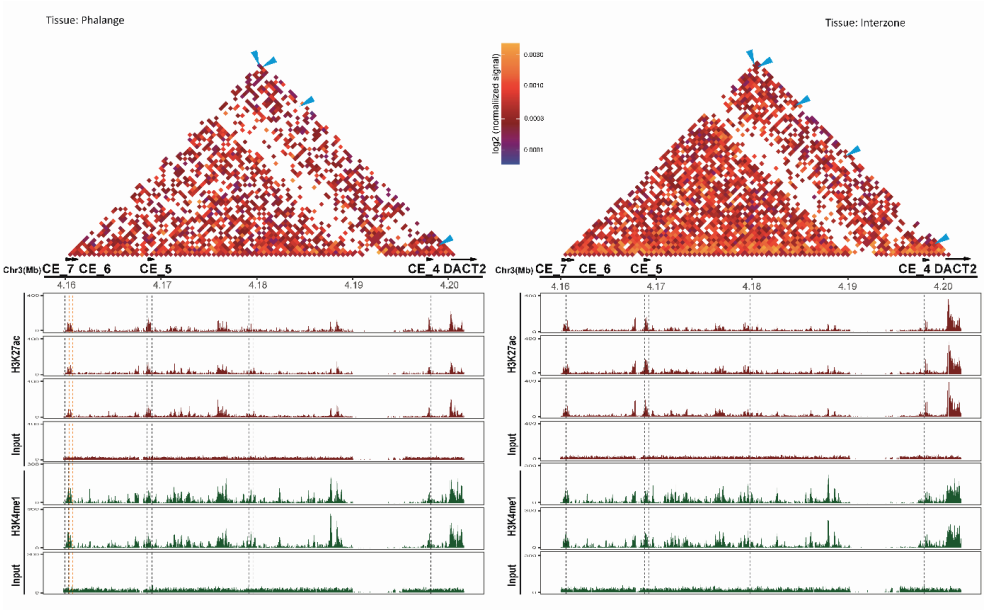


Figure S5.3. The *DACT2* CEs.

Zoomed T2C interaction maps (bin size = 5 kb) presenting proximities (blue arrows) between *DACT2* and CE4-CE7. Dashed lines mark regions of CEs. The H3K27ac peaks are highlighted in red. The H3K4me1 peaks are marked by green. ChIP-Seq inputs are marked by black.

Next, we analyzed whether the seven selected CEs change enhancer state from strongly active (i.e. enriched in H3K27ac and H3K4me1) to poised enhancer (enriched in H3K4me1 only). For this, we estimated the chromatin state signatures with ChromHMM tool which uses a multivariate hidden Markov model (HMM) based on ChIP-Seq data from both H3 marks. Four such chromatin state signatures have been identified as signature A (strongly active enhancer), B (active enhancer), C (poised enhancer) and D (no enhancer marks), respectively (**Fig. 5.5B**; **Table S5.1**).

This comparison further revealed that the major part of the CE1 region changed from strong to poised enhancer/no enhancer state, like the entire CE3 region did. In contrast, CE2 (located between CE1 and CE3) was identified as strongly active enhancer in interzone (Int) and phalange (Ph in **Fig. 5.5C**, top line). Further, CE4 switched from active to poised enhancer in the majority of its region, and CE5 from strongly active to poised enhancer only in a part of its region (**Fig. 5.5C**, 2nd and 3rd line from the top). A part of CE6 switched from poised enhancer in interzone to active enhancer in phalange; in contrast, CE7 presented with the same chromatin state in both tissues (**Fig. 5.5C**, bottom line). These data support the hypothesis that the chromatin state between interzone and phalange tissue during joint formation in the embryo is different and dynamic around the *DACT2* locus.

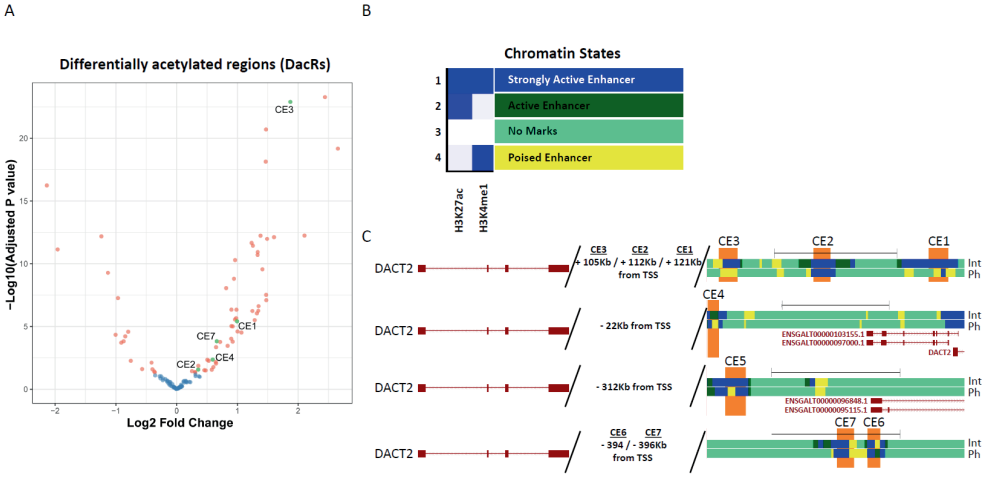


Figure 5.5. Further characterization of *DACT2* CEs.

(A) Volcano plot of differentially acetylated regions (DacRs). The volcano plot was generated using DiffBind based on the enrichment of H3K27ac in interzone and phalange samples. Blue dots represent DAR with FDR > 0.05, salmon-color dots mark significantly differentiated DAR (FDR <0.05). Green dots depict DacRs encompassing selected CEs. **(B)** A 4-chromatin state model obtained using interzone and phalange histone ChIP-Seq data. The chromatin state model was generated using ChromHMM based on all sample replicates, including input controls. The states were annotated using the model emission probabilities (**Table S5.7**) visualized by the intensity of blue color within the heatmap. (C) The UCSC genome browser view of interzone and phalange states around of selected CEs (marked by orange). Colors of chromatin states correspond to state annotation in panel B.

Similar investigation of CEs possibly involved in the regulation of the *SMOC2* locus revealed that CE7 is located in spatial proximity to the *SMOC2* promoter, suggesting that CE7 regulates both *SMOC2* and *DACT2*. Further analysis of our T2C data indeed revealed that both genes are brought in spatial proximity to CE7 (**Table S5.1**). For visualizing the *cis*-proximities between each gene and the putative common enhancer CE7, we generated virtual 4C tracks based on our T2C data, and combined this with H3K27ac and H3K4me1 enrichment analysis. This operation, using virtual 4C *DACT2/ SMOC2* promoters as viewpoints, confirmed the presence of *cis*-proximities between the promoter segment of both genes and CE7. Also, this 4C approach showed that both gene promoters are located in spatial proximity (**Fig. 5.6**).

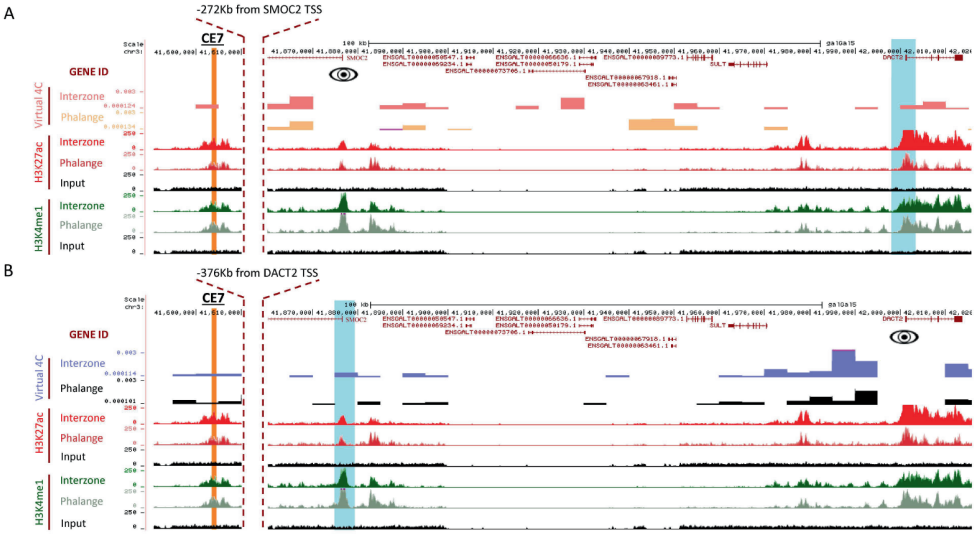


Figure 5.6. Characterization of virtual 4C for *DACT2* and *SMOC2*.

(A) The virtual 4C was generated based using the T2C data for interzone and phalange samples. The genomic region coordinates are indicated in the top lane, the height of peaks corresponds to signal value. The signal value was averaged using bin size = 5 kb. The virtual 4C for *SMOC2* as a viewpoint revealed the loop between *SMOC2* and CE7 (dark orange), and that and *SMOC2* and *DACT2* promoter (highlighted by blue) are in spatial proximity. The Virtual 4C track for interzone marked by salmon-color and phalange by orange. The H3K27ac enrichment tracks are marked by red, and H3K4me1 by green. Input control tracks are marked by black. **(B)** Characterization of virtual 4C tracks (blue for interzone and black for phalange) generated for *DACT2* as a viewpoint showed cis-proximity between *DACT2* and CE7 were (highlighted by dark orange). Also, in this case the *DACT2 and SMOC2* promoter (marked in blue) have been identified as located in spatial proximity. The H3 histone marks tracks colored as in panel A.

5.3.4. *In vivo* testing of candidate enhancers using a zebrafish enhancer assay

To test *in vivo* activity of selected *DACT2* CEs we applied a zebrafish enhancer assay using a Zebrafish Enhancer Detection (ZED) vector containing two expression cassettes (Bessa *et al.*, 2009). One of the vectors has a Green Fluorescent Protein (GFP) cDNA driven by the minimal *GATA2A* promoter, which can be activated by an enhancer placed upstream of it. A second cassette then serves as an integration control where Red fluorescent protein (DsRed) mRNA is placed under the control of Cardiac Actin Promoter, which is active in heart and somites. Embryos were injected at the 1-cell stage and monitored under a fluorescent microscope every 24 hours (*data not shown*). Most prominent production of GFP for each of our selected 7 CEs was observed 96 hours post-fertilization (hpf). The specific GFP signals presented in at least 30% of embryos analyzed were considered specific (**Fig. S5.4**).

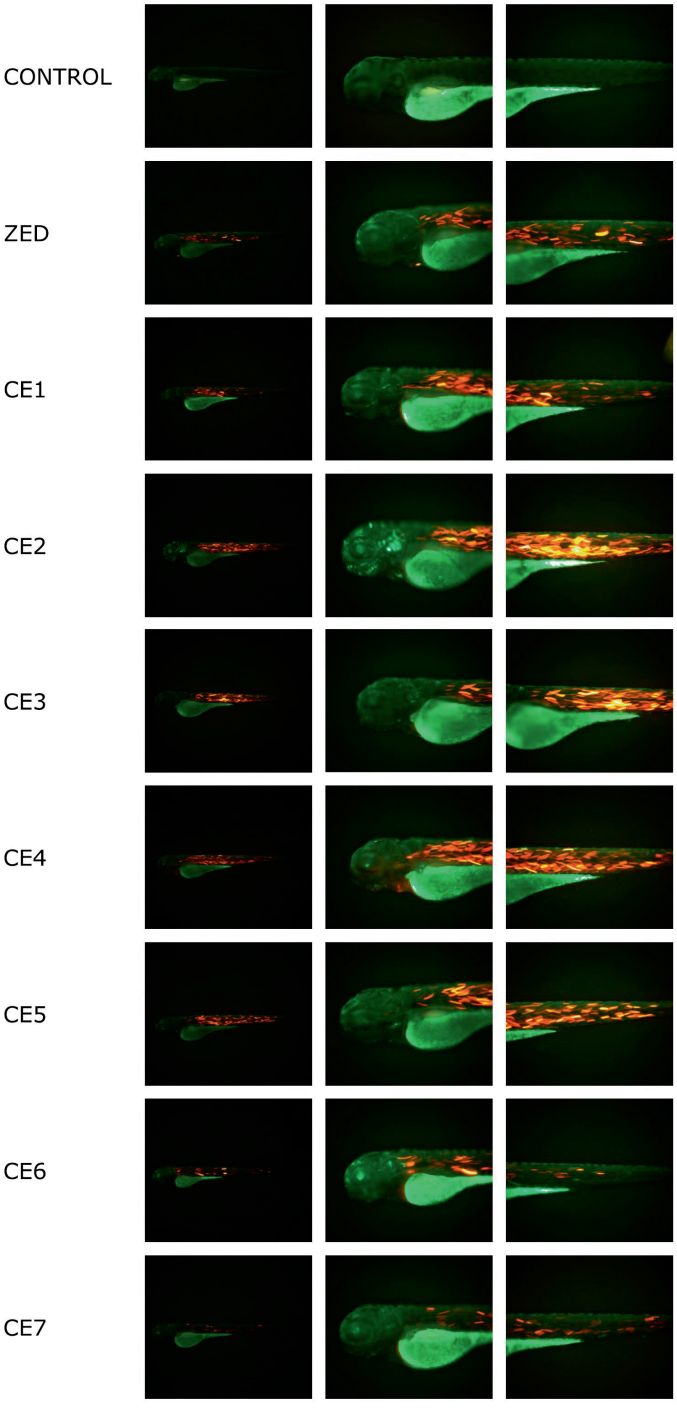


Figure S5.4. Screening of DACT2 CEs in zebrafish.

Merged fluorescence images showing GFP and RFP presence in transgenic enhancer reporter assays in zebrafish larvae at 96 hpf (lateral view). Tested are the enhancers CE1 to CE7. DsRed indicates integration control.

We found the most prominent activity of the selected CEs in the regions of forebrain, midbrain, otolith and jaw cartilages, where *DACT2* is also expressed in zebrafish (Schubert *et al.*, 2014) (**Fig. 5.7A**). We noticed that CE1 and CE2 were active in the branchial arches of the developing larvae, partially sharing regulatory information, although with varying fluorescence intensities (**Fig. 5.7D**). Further, the low intensity GFP signal for CE3 was observed around the otic vesicle (**Fig. 5.7D**). CE6 was active in the forebrain and around the otic vesicle (**Fig. S5.5**), however its signal in the forebrain region sporadically overlapped with ZED auto-fluorescence. So, the regulatory role of CEs in this specific area cannot be fully determined, although the detected CE signal was invariably stronger than in the group injected with empty (without CE) vector. Contrary, we could never detect signals from CE4, CE5 and CE7. Importantly, the localization of GFP for CEs1-3 and CE6 in developing larvae resembled the pattern of RNA *in situ* hybridization of *DACT2* in zebrafish larvae (Schubert *et al.*, 2014). Overall, these results show overlap between the *DACT2* expression and its CEs when tested *in vivo*, suggesting the involvement of the analyzed CEs in the formation of those structures.

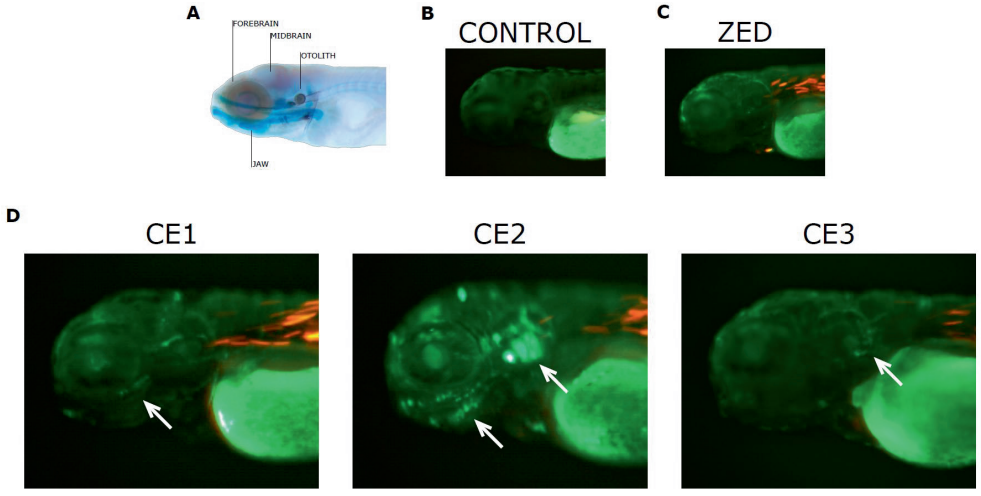


Figure 5.7. CEs activity in developing zebrafish larvae.

(A) Zebrafish larvae at 120 hpf stained with Alcian Blue (cartilage) and Alizarin Red (bone) to reveal the skeletal structures. **(B)** Uninjected zebrafish larve at 96 hpf observed in the green channel, revealing the autofluorescence. The autofluorescence in the red channel was not detected (*data not shown*). **(C)** Empty vector injected larvae at 96 hpf. **(D)** Representative merged fluorescent images showing the activity of enhancer CE1, CE2 or CE3-driven GFP and DsRed integration control. The arrow in CE1 points to the developing jaw, in CE2 to the developing jaw and otic vesicle and in CE3 to the otic vesicle only.

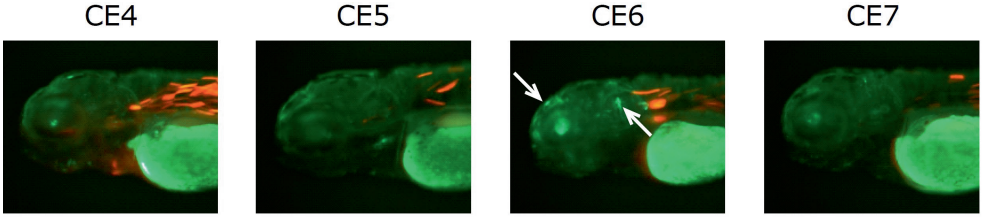


Figure S5.5. Screening of DACT2 CE4-CE7 in zebrafish.

Representative images showing activity of tested CE4-CE7 (GFP). DsRed indicates integration control.

5.4. Discussion

Limb development, including patterning and cell differentiation, is subject to precise and dynamic spatiotemporal transcriptional control. There are many examples describing the intricate regulatory circuits for limb-relevant genes such as *SHH*, genes of the *HOXD* cluster, and *GDF5* (Lettice *et al.*, 2003; Lee *et al.*, 2006; Chen *et al.*, 2016; Lettice *et al.* 2017). The deregulation of such genes, caused for instance by mutation in enhancers or by chromosomal rearrangements, leads to limb deformities (Spielmann *et al.*, 2012; Lohan *et al.*, 2014; Tayebi *et al.*, 2014; Bunyan *et al.*, 2016). In this study, we investigated the role of chromatin architecture and enhancers in the expression of two genes expressed during joint development, *DACT2* and *SMOC2*. We opted for the physical dissection of interzone tissue instead of one based on the isolation of *GDF5*-positive (+) cells, because not all cells during early stages of joint formation are convincingly *GDF5+* (Bian *et al.*, 2020). Dissected interzones presented significantly higher expression of *GDF5*, *DACT2* and *SMOC2* as compared to adjacent phalange, consistent with publicly available (Diez-Roux *et al.*, 2011; Sensiate *et al.*, 2014; Chen *et al.*, 2016) as well as our own *in situ* hybridization studies.

To explore the potential transcriptional co-regulation of both genes, we analyzed the chicken chr3:40.15-43.6 genomic region encompassing the *DACT2-SMOC2* loci, focusing on the chromatin architectural dynamics for joint interzone and the adjacent phalange. Based on T2C experiments, we detected DNA-loops within the TADs, which were predominantly tissue-specific. This is in agreement with Hi-C data showing that TADs tend to be conserved between tissues and cell types, in contrast to intra-TAD interactions, which within different cell types can be dynamic (Dixon *et al.*, 2016; Fishman *et al.*, 2019). However, to characterize the dynamics changes of DNA-loops for joint interzone and the adjacent phalange with higher accuracy and reproducibility this analysis should be further supported by additional T2C experiments with more biological replicates.

Multiple studies point out the role of enhancers in co-shaping chromatin architecture (Sanyal *et al.*, 2012; Dixon *et al.*, 2016; Espinola *et al.*, 2021). Using our recent data from ChIP-Seq for H3K27ac and H3K4me1 (Nowosad *et al.*, 2022) we identified 88 CEs, i.e. 53 in interzone and 35 in phalange, all CEs being located in the aforementioned *DACT2-SMOC2* region. Further, by then performing the analysis of H3K27ac marks, we showed that 45 (out of a total of 59) DacRs were significantly enriched in interzone. These differences in numbers of CEs as well as overall acetylation level partially explain the dynamics of 3D organization within the *DACT2-SMOC2* region, which presents higher frequency of *cis*-proximities in interzone than in adjacent phalange.

By integration of the data on 3D chromatin structure and H3 signatures, we selected seven candidate enhancers (CE1 to CE7) located close to *DACT2*. Multiple studies have shown that the enhancers' activities are associated with their states (Cui *et al.*, 2009; Creyghton *et al.*, 2010; Koenecke *et al.*, 2017). Thus, we used ChromHMM to characterize the states around these selected CEs, and annotated them according to the Roadmap Epigenomics Program (Roadmap Epigenomics Consortium *et al.*, 2015). We discovered interesting correlations, in particular the switching from strongly active to poised enhancer, and the interzone specific *cis*-proximity within the *DACT2* locus. Specifically, the chromatin state changes within the entire CE3, whereas CE1 and CE5 present a switch over a section of the enhancer region. Virtual 4C data revealed that CE7 presents *cis*-proximity with both *DACT2* and *SMOC2*, suggesting that this enhancer supports expression of more than one target gene.

Previous studies have shown that enhancer elements without homologous sequence in zebrafish can act as active enhancers in zebrafish (Fisher *et al.*, 2006; Hirsch *et al.*, 2018; Bar Yaacov *et al.*, 2019). This is supported by our findings, but not all of the CEs were detectably active in the zebrafish assay, suggesting that some of the CEs can be chicken-specific. However, we were able to show that CEs1-3 and CE6 present activity during overall zebrafish development, which correlates with the known expression domain of *DACT2* in zebrafish larvae. Further studies in developing chicken embryos themselves will be required for detailed characterization of all *DACT2* enhancers in the interzone region.

Bullet points

- DACT2 and SMOC2 are differentially expressed between interzone and phalange.
- DACT2 CEs switch chromatin state between interzone and phalange.
- Differences in numbers of active CEs, as well as overall acetylation level between interzone and phalange, partially explain the dynamics of 3D organization within the *DACT2-SMOC2* region.

Data availability

The T2C data is available under the gene expression omnibus (GEO) accession number GSE210108.

Funding

This work was supported by the Polish National Science Centre (UMO-2015/19/B/NZ4/03184) and primary and local BIG initiative funding to the Department of Cell Biology at Erasmus University Medical Center.

Ethical statement

The experiments with mouse and chicken early embryos were performed in accordance with the relevant guidelines as applied and approved by the Ethical Committee at the Medical University in Lublin, where this work was performed, and also comply with the European regulations (directive 2010/63/EU).

Conflict of interest

All authors have no conflict of interest to declare.

Acknowledgments

We would like to thank the members of the Center for Biomics-Genomics at the Erasmus Medical for technical support and the Experimental Medicine Center at the Medical University of Lublin for taking care the mice and zebrafish.

5.5. Experimental procedures

Tissue collection

The experiments with mouse and chicken early embryos were performed in accordance with the relevant guidelines as applied and approved by the Ethical Committee at the Medical University in Lublin, where this work was performed, and also comply with the European regulations (directive 2010/63/EU).

Fertilized eggs (White Leghorn) were incubated at 38.5°C and 70% humidity for 7.5 days in a Grumbach BSS420 CTD7 incubator. The Hamilton Hamburger (HH) developmental stage was evaluated under Zeiss Stereo Discovery V8 microscope equipped with 0.63x Plan Apo S Objective Lens. Selected embryos at HH32 were sacrificed for tissue microdissection. The joint interzones and phalanges were collected from 2nd and 3rd hindlimb digit using a Dumont No.5 forceps (tip dimensions: 0.005 x 0.025 mm).

CD1 mice were group-housed in conventional cages conforming to local and (inter)national Animal Welfare Guidelines. Pregnant mice were sacrificed after 14 or 14.5 days by cervical dislocation. Subsequently, E14 or E14.5 mouse embryos were collected for RNA *in situ* hybridization.

RT-qPCR

The Syngen Tissue RNA Kit was used for total RNA extraction, followed by treatment with DNase, using QIAGEN RNase-Free DNase Set. The RNA in biological triplicates for interzone and phalange was used for cDNA synthesis with Invitrogen™ SuperScript™ IV Reverse Transcriptase and Oligo(dT) primer, following the manufacturer's suggestions. The qPCR was performed with LightCycler® 480 Instrument II using PowerUp™ SYBR™ Green Master Mix II. Sequences of primers are listed in **Table S5.1**. Gene expression was normalized to expression of GAPDH. The T-test was used to analyze the expression of candidate genes.

ChIP-Seq data analysis, candidate enhancer (CEs) identification and characterization of chromatin states

The ChIPseq data for H3K27ac and H4K4me1 was used from Nowosad *et al.,* (2022). The quality of raw fastq files was validated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, reads were mapped to the *Gallus gallus* 5.0 reference genome using Bowtie 2 with default parameters (Langmead and Salzberg, 2013) followed by removal of PCR-duplicates using Picard (http://broadinstitute.github.io/picard/). The peaks were called using MACS2 with input as a control (Zhang *et al.*, 2008). Default parameters and significance level threshold FDR <0.05 were used for MACS2 peak calling.

The differentially acetylated regions (DacRs) were identified based on the H3K27ac ChIP-Seq data using DiffBind tool (https://bioconductor.org/packages/DiffBind/) with default setting, except dba.count(summits = 1000). The log2FoldChange for identified DacRs were calculated using DiffBind binding affinity analysis following the default settings. Subsequently, the DacRs were filtered for FDR <0.05 and coordinates encompassing the *SMOC2-DACT2* region (chr3:40,15-43,6 Mb). The enrichment tracks for H3K4me1 and H3K27ac ChIP-Seq data were generated using deepTools2 (Ramirez *et al.*, 2016). Briefly, bamCoverage with kilobase per million mapped reads (RPKM) normalization was used. The enrichment tracks were visualized using in-house R script.

Candidate enhancers (CEs) were identified using an in-house made script. Specifically, the consensus peaks called by MACS2 were defined by merging replicates using BEDTools. Next, the H3K27ac regions were intersected from H3K4me1 peaks using BEDTools intersect (Quinlan and Hall, 2010), and promoter regions (-2.5/+2.5 kb from TSS) were removed from the dataset using GenomicRanges (Lawrence *et al.*, 2013).

RNA-Seq data analysis for evaluation of *DACT2* and *SMOC2* expression

The RNA-Seq data for interzone and phalange was used from Nowosad *et al.* (2022) and analyzed as previously described (s). Briefly, after quality check of raw data, the reads were mapped to the *Gallus gallus* 5.0 reference genome using STAR with default parameters (Dobin *et al.*, 2013). Next, the reads were calculated per gene using featureCounts (Liao *et al.*, 2014). The differential data analysis has been performed with DEseq2 (Love *et al.*, 2014), and boxplots have been prepared based on DEseq2 normalized reads counts.

Targeted chromatin capture (T2C)

T2C protocols were further adapted from Boltsis *et al.* (2021). For each sample, 100 interzones or transient cartilages were dissected, respectively pooled, to obtain 10^6 cells per tissue. The preparation of single cell suspension and chromatin crosslinking was performed as described in the ChIP protocol. Nuclei were extracted during 20 min by incubating the cells in ice-cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl,

0.5% IGEPAL® CA-630 and complete protease inhibitors) at 4°C. Isolated nuclei were washed twice by resuspension in 500 µl of PBS, followed by slow-spin centrifugation at 340 g, at 4 °C. Next, the nuclei in the pellet were resuspended in freshly prepared 1.2x restriction buffer, followed by addition of 10% SDS to a final concentration of 1.6% SDS and 1 hour of incubation at 37°C, while shaking at 900 rpm. SDS was quenched by addition of 20% Triton X-100 (final concentration: 1%) and incubation for 1 hour at 37°C while shaking at 900 rpm. The chromatin was digested with *Apo*I (New England Biolabs) (40 U/sample) for 16 hours at 37°C, while shaking at 900 rpm. The digested chromatin was ligated with T4 DNA Ligase (10 U/sample) at 16°C for 16 hours. The next day, chromatin was incubated with 3 µl of Proteinase K (10 mg/ml) for 1 hour at 65°C, and 3 µl of RNAse-A (10 mg/ml) for 45 min at 37°C, followed by DNA purification using Phenol:Chloroform according to the manufacturer's instruction. The religated DNA was digested with *Dpn*II (50 U/sample) at 37°C for 16 hours, while shaking at 500 rpm, and purified by Phenol:Chloroform prior to T2C library preparation.

T2C library preparation

The protocol was again further adapted from Kolovos and co-workers (Kolovos *et al.*, 2018) with previously applied modifications (Birkhoff *et al.*, 2020, Boltsis *et al.,* 2021). For the joint interzone and adjacent phalange sample, a T2C library was prepared using 350 and 175 ng of linearized chromatin, respectively. The samples were re-buffered to 10 mM Tris-HCl, pH 8 by a standard AMPure XP (Agencourt) bead clean-up procedure. The chromatin was sheared to 250-400 bp-sized fragments by a S220 Covaris (Covaris Inc.). The concentration was determined by Quant-it high sensitivity (ThermoFisher Scientific). For each sample, 100 ng of sheared chromatin was end-repaired and A-tailed using the Kapa hyper prep kit (Roche) according to the manufacturer's instructions. SeqCap library adaptors were ligated followed by AMPure bead clean-up. The pre-capture library was amplified by PCR using KAPA HiFi hotstart readymix for 9 cycles. The amplified pre-capture library was purified by bead clean-up and quantified by Bioanalyzer DNA1000 assay (Agilent) according to the manufacturer's instructions.

A *DACT2 – SMOC2* locus *Gallus gallus* 5.0-based design was ordered at NimbleGen (Roche) with baits located between 40,154,526 and 43,603,576 of Chr 3. A pooled hybridization mixture was prepared with 1 µg pre-capture library of each sample, 1 mM HE-index-oligo, 1 mM HE universal oligo, COT human DNA, AMPure XP reagent and added to 4.5 µl of pre-ordered baits and subsequently hybridized for 16 hours at 47°C. Post-hybridization, the samples were washed according to the instructions in the Nimblegen SeqcapEZ Hypercap workflow (Roche), the chromatin captured using capture beads. The captured library was amplified by PCR using Kapa HiFi mix and purified by AMPure XP beads. The captured library was quantified by Nanodrop spectrophotometer and the quality was assessed using a Bioanalyzer DNA1000 assay. Finally, the captured T2C libraries were denatured and sequenced on an Illumina HiSeq2500 sequencer with a custom recipe of 6 dark cycles, followed by paired end 101 sequencing with single index using the rapid v2 chemistry according to manufacturer's instructions (Illumina) to a depth of approximately 30M clusters per sample.

T2C data analysis

The analysis was performed using the pipeline described by Kolovos *et al.* (2018). Specifically, the quality of raw fastq files was evaluated using FASTQC. The reads were trimmed for adapters using AdapterTrimmer (https://github.com/erasmus-center-for-biomics/AdapterTrimmer) and mapped to the *Gallus gallus* 5.0 reference genome with the BWA aligner and the BWA-backtrack method. Alignments were subsequently annotated with the restriction fragments in which they were located. The proximity matrix was then constructed from the mapped primary alignments with their mapped primary mates. Further analyses and filtering based on the proximity matrix were performed in the R environment for statistical computing. Samples were normalized using array normalization. The virtual 4C tracks were generated using in-house R script.

Cloning and zebrafish transgenic enhancer assay and luciferase assay

Primers were designed to amplify CEs from chicken DNA (**Table S5.1**). Forward primers contain CACC flanking sequence complementary to the GTGG overhang sequence of the pENTR™/D-TOPO® vector. The CEs were PCR amplified using touchdown PCR protocol with Phusion™ High-Fidelity DNA Polymerase. The PCR products were cloned into pENTR™/D-TOPO® vector using pENTR™ Directional TOPO® Cloning Kit, followed by recombination to Zebrafish Enhancer Detection (ZED) vector (Bessa *et al.*, 2009). The ZED vector contains green fluorescent protein GFP under GATA minimal promoter and red fluorescent protein (DsRed2) under cardiac actin promoter.

The wild-type strain zebrafish larvae were handled according to the welfare regulations and standard protocols approved by the Ethical Committee at the Medical University in Lublin, which comply with the European regulations (directive 2010/63/EU). The CEs/ZED and empty ZED vector alone were injected using standard procedures under the Stemi 508 stereo microscope (ZEISS) equipped with a microinjection

unit. Specifically, the single-cell stage eggs were used for injection of 1nl mixture containing 1:1 ratio of ZED (50 ng/µl) along with Tol2 transposase mRNA (40 ng/µl) to facilitate genomic integration (Fisher *et al.*, 2006). For statistical significance, the 130 to 200 embryos were injected with each construct, and for each construct a full experiment was conducted at least twice to ensure reproducibility and high number of embryos injected. Eggs that were unfertilized or damaged during the injection process were removed after approximately 10 hours post-injection, as their presence could negatively affect the survival of remaining embryos. Larvae were then maintained in standard conditions (28.5°C) and the efficiency of the genomic integration was validated based on the DsRed2 expression pattern analyzed after 72 hours. The 1-phenyl 2-thiourea (PTU) at the standard concentration of 0.003% (200 µM) was added to the E3 medium to delay pigmentation and therefore more accurately visualize the fluorescent signals. The specimens without DsRed2 expression in muscles and heart (positive control of construct integration) were excluded from further examination. Next, the larvae were screen for expression of GFP after 96h. If GFP expression pattern was present in at least 30% of specimens, it was considered as enhancer-specific. Larvae were screened for expression patterns under the SteREO Discovery.V8 fluorescence stereomicroscope (ZEISS, Germany) and images were captured using the ZEN2 software (ZEISS, Germany).

Supplementary tables

Table S5.1 compiles many subtables and is available as:

- e-copy on request to the author and/or promoters of this PhD thesis.
- downloadable as supplementary data from: https://github.com/karolnowosad/Chromatin-architecture-and-cis-regulatory-landscape-of-the-DACT2-SMOC2-locus-in-the-developing-synov

Literature references

Akiyama H, Kim JE, Nakashima K, Balmes G, Iwai N, Deng JM, Zhang Z, Martin JF, Behringer RR, Nakamura T, de Crombrugghe B. Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. Proc Natl Acad Sci USA. 2005 Oct 11;102(41):14665-70.

Archer CW, Dowthwaite GP, Francis-West P. Development of synovial joints. Birth Defects Res C Embryo Today. 2003 May;69(2):144-55.

Bar Yaacov R, Eshel R, Farhi E, Shemuluvich F, Kaplan T, Birnbaum RY. Functional characterization of the ZEB2 regulatory landscape. Hum Mol Genet. 2019 May 1;28(9):1487-97.

Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F, Gómez-Skarmeta JL. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev Dyn. 2009 Sep;238(9):2409-17.

Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B. Sox9 is required for cartilage formation. Nat Genet. 1999 May;22(1):85-9.

Bian Q, Cheng YH, Wilson JP, Su EY, Kim DW, Wang H, Yoo S, Blackshaw S, Cahan P. A single cell transcriptional atlas of early synovial joint development. Development. 2020 Jul 20;147(14):dev185777.

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, Nowosad K, van den Hout MCGN, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. Hum Mol Genet. 2020 Aug 29;29(15):2535-50.

Boltsis I, Nowosad K, Brouwer RWW, Tylzanowski P, van IJcken WFJ, Huylebroeck D, Grosveld F, Kolovos P. Low Input Targeted Chromatin Capture (Low-T2C). Methods Mol Biol. 2021;2351:165-79.

Bunyan DJ, Baffico M, Capone L, Vannelli S, Iughetti L, Schmitt S, Taylor EJ, Herridge AA, Shears D, Forabosco A, Coviello DA. Duplications upstream and downstream of SHOX identified as novel causes of Leri-Weill dyschondrosteosis or idiopathic short stature. Am J Med Genet A. 2016 Apr;170A(4):949-57.

Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013 Mar 7;49(5):825-37.

Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, Shoulders, Elbows, Knees, and Toes: Modular Gdf5 Enhancers Control Different Joints in the Vertebrate Skeleton. PLoS Genet. 2016 Nov 30;12(11):e1006454.

Chijimatsu R, Saito T. Mechanisms of synovial joint and articular cartilage development. Cell Mol Life Sci. 2019 Oct;76(20):3939-52.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA. 2010 Dec 14;107(50):21931-6.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell. 2009 Jan 9;4(1):80-93.

Dhordain P, Dewitte F, Desbiens X, Stehelin D, Duterque-Coquillaud M. Mesodermal expression of the chicken erg gene associated with precartilaginous condensation and cartilage differentiation. Mech Dev. 1995 Mar;50(1):17-28.

Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, Lin-Marq N, Koch M, Bilio M, Cantiello I, Verde R, De Masi C, Bianchi SA, Cicchini J, Perroud E, Mehmeti S, Dagand E, Schrinner S, Nürnberger A, Schmidt K, Metz K, Zwingmann C, Brieske N, Springer C, Hernandez AM, Herzog S, Grabbe F, Sieverding C, Fischer B, Schrader K, Brockmeyer M, Dettmer S, Helbig C, Alunni V, Battaini MA, Mura C, Henrichsen CN, Garcia-Lopez R, Echevarria D, Puelles E, Garcia-Calero E, Kruse S, Uhr M, Kauck C, Feng G, Milyaev N, Ong CK, Kumar L, Lam M, Semple CA, Gyenesei A, Mundlos S, Radelof U, Lehrach H, Sarmientos P, Reymond A, Davidson DR, Dollé P, Antonarakis SE, Yaspo ML, Martinez S, Baldock RA, Eichele G, Ballabio A. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol. 2011 Jan 18;9(1):e1000582.

Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol Cell. 2016 Jun 2;62(5):668-80.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21.

Espinola SM, Götz M, Bellec M, Messina O, Fiche JB, Houbron C, Dejean M, Reim I, Cardozo Gizzi AM, Lagha M, Nollmann M. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. Nat Genet. 2021 Apr;53(4):477-86.

Feng C, Chan WCW, Lam Y, Wang X, Chen P, Niu B, Ng VCW, Yeo JC, Stricker S, Cheah KSE, Koch M, Mundlos S, Ng HH, Chan D. Lgr5 and Col22a1 Mark Progenitor Cells in the Lineage toward Juvenile Articular Chondrocytes. Stem Cell Reports. 2019 Oct 8;13(4):713-29.

Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science. 2006 Apr 14;312(5771):276-9.

Fishman V, Battulin N, Nuriddinov M, Maslova A, Zlotina A, Strunov A, Chervyakova D, Korablev A, Serov O, Krasikova A. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res. 2019 Jan 25;47(2):648-65.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016 May 31;15(9):2038-49.

Grubert F, Srivas R, Spacek DV, Kasowski M, Ruiz-Velasco M, Sinnott-Armstrong N, Greenside P, Narasimha A, Liu Q, Geller B, Sanghi A, Kulik M, Sa S, Rabinovitch M, Kundaje A, Dalton S, Zaugg JB, Snyder M. Landscape of cohesin-mediated chromatin loops in the human genome. Nature. 2020 Jul;583(7818):737-43.

Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. Cell. 2001 Feb 9;104(3):341-51.

Hirsch N, Eshel R, Bar Yaacov R, Shahar T, Shmulevich F, Dahan I, Levaot N, Kaplan T, Lupiáñez DG, Birnbaum RY. Unraveling the transcriptional regulation of TWIST1 in limb development. PLoS Genet. 2018 Oct 29;14(10):e1007738.

Hyde G, Dover S, Aszodi A, Wallis GA, Boot-Handford RP. Lineage tracing using matrilin-1 gene expression reveals that articular chondrocytes exist as the joint interzone forms. Dev Biol. 2007 Apr 15;304(2):825-33.

Jenner F, IJpma A, Cleary M, Heijsman D, Narcisi R, van der Spek PJ, Kremer A, van Weeren R, Brama P, van Osch GJ. Differential gene expression of the intermediate and outer interzone layers of developing articular cartilage in murine embryos. Stem Cells Dev. 2014 Aug 15;23(16):1883-98.

Kania K, Colella F, Riemen AHK, Wang H, Howard KA, Aigner T, Dell'Accio F, Capellini TD, Roelofs AJ, De Bari C. Regulation of Gdf5 expression in joint remodelling, repair and osteoarthritis. Sci Rep. 2020 Jan 13;10(1):157.

Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J. Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res. 2017 Jan;27(1):64-74.

Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, Imam AMA, van de Werken HJG, Wendt KS, Knoch TA, van IJcken WFJ, Grosveld F. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. Nat Protoc. 2018 Mar;13(3):459-77.

Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, Wendt KS, van IJcken WF, Grosveld F, Knoch TA. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. Epigenetics Chromatin. 2014 Jun 16;7:10.

Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. BMC Biol. 2018 Aug 7;16(1):87.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.

Lee AP, Koh EG, Tay A, Brenner S, Venkatesh B. Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters. Proc Natl Acad Sci USA. 2006 May 2;103(18):6994-9.

Lettice LA, Devenney P, De Angelis C, Hill RE. The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. Cell Rep. 2017 Aug 8;20(6):1396-408.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003 Jul 15;12(14):1725-35.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923-30.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93.

Lohan S, Spielmann M, Doelken SC, Flöttmann R, Muhammad F, Baig SM, Wajid M, Hülsemann W, Habenicht R, Kjaer KW, Patil SJ, Girisha KM, Abarca-Barriga HH, Mundlos S, Klopocki E. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. Clin Genet. 2014 Oct;86(4):318-25.

Long F, Shi H, Li P, Guo S, Ma Y, Wei S, Li Y, Gao F, Gao S, Wang M, Duan R, Wang X, Yang K, Sun W, Li X, Li J, Liu Q. A SMOC2 variant inhibits BMP signaling by competitively binding to BMPR1B and causes growth plate defects. Bone. 2021 Jan;142:115686.

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

Lu H, Ju DD, Yang GD, Zhu LY, Yang XM, Li J, Song WW, Wang JH, Zhang CC, Zhang ZG, Zhang R. Targeting cancer stem cell signature gene SMOC-2 Overcomes chemoresistance and inhibits cell proliferation of endometrial carcinoma. EBioMedicine. 2019 Feb;40:276-89.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 21;161(5):1012-25.

Mackie EJ, Ahmed YA, Tatarczuch L, Chen KS, Mirams M. Endochondral ossification: how cartilage is converted into bone in the developing skeleton. Int J Biochem Cell Biol. 2008;40(1):46-62.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015 Jun;47(6):598-606.

Mommaerts H, Esguerra CV, Hartmann U, Luyten FP, Tylzanowski P. Smoc2 modulates embryonic myelopoiesis during zebrafish development. Dev Dyn. 2014 Nov;243(11):1375-90.

Nowosad K, Brouwer RWW, Odrzywolski A, Korporaal AL, Gielniewski B, Wojtaś B, van IJcken WFJ, Grosveld F, Huylebroeck D, Tylzanowski P. Identification of candidate enhancers controlling the transcriptome during the formation of interphalangeal joints. Sci Rep. 2022 Jul 27;12(1):12835.

Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, Kato M, Garvin TH, Pham QT, Harrington AN, Akiyama JA, Afzal V, Lopez-Rios J, Dickel DE, Visel A, Pennacchio LA. Enhancer redundancy provides phenotypic robustness in mammalian development. Nature. 2018 Feb 8;554(7691):239-43.

Pacifici M, Koyama E, Shibukawa Y, Wu C, Tamamura Y, Enomoto-Iwamoto M, Iwamoto M. Cellular and molecular mechanisms of synovial joint and articular cartilage formation. Ann N Y Acad Sci. 2006 Apr;1068:74-86.

Pazin DE, Gamer LW, Capelo LP, Cox KA, Rosen V. Gene signature of the embryonic meniscus. J Orthop Res. 2014 Jan;32(1):46-53.

Peeters T, Monteagudo S, Tylzanowski P, Luyten FP, Lories R, Cailotto F. SMOC2 inhibits calcification of osteoprogenitor and endothelial cells. PLoS One. 2018 Jun 13;13(6):e0198104.

Petit F, Sears KE, Ahituv N. Limb development: a paradigm of gene regulation. Nat Rev Genet. 2017 Apr;18(4):245-58.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W160-5.

Ray A, Singh PN, Sohaskey ML, Harland RM, Bandyopadhyay A. Precise spatial restriction of BMP signaling is essential for articular cartilage differentiation. Development. 2015 Mar 15;142(6):1169-79.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30.

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012 Sep 6;489(7414):109-13.

Schubert FR, Sobreira DR, Janousek RG, Alvares LE, Dietrich S. Dact genes are chordate specific regulators at the intersection of Wnt and Tgf-β signaling pathways. BMC Evol Biol. 2014 Aug 6;14:157.

Sensiate LA, Sobreira DR, Da Veiga FC, Peterlini DJ, Pedrosa AV, Rirsch T, Joazeiro PP, Schubert FR, Collares-Buzato CB, Xavier-Neto J, Dietrich S, Alvares LE. Dact gene expression profiles suggest a role for this gene family in integrating Wnt and TGF-β signaling pathways during chicken limb development. Dev Dyn. 2014 Mar;243(3):428-39.

Spielmann M, Brancati F, Krawitz PM, Robinson PN, Ibrahim DM, Franke M, Hecht J, Lohan S, Dathe K, Nardone AM, Ferrari P, Landi A, Wittler L, Timmermann B, Chan D, Mennen U, Klopocki E, Mundlos S. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am J Hum Genet. 2012 Oct 5;91(4):629-35.

Su JR, Kuai JH, Li YQ. Smoc2 potentiates proliferation of hepatocellular carcinoma cells via promotion of cell cycle progression. World J Gastroenterol. 2016 Dec 7;22(45):10053-10063.

Tan Y, Li QM, Huang N, Cheng S, Zhao GJ, Chen H, Chen S, Tang ZH, Zhang WQ, Huang Q, Cheng Y. Upregulation of DACT2 suppresses proliferation and enhances apoptosis of glioma cell via inactivation of YAP signaling pathway. Cell Death Dis. 2017 Aug 10;8(8):e2981.

Tayebi N, Jamsheer A, Flöttmann R, Sowinska-Seidler A, Doelken SC, Oehl-Jaschkowitz B, Hülsemann W, Habenicht R, Klopocki E, Mundlos S, Spielmann M. Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. Orphanet J Rare Dis. 2014 Jul 29;9:108.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell. 2002 Dec;10(6):1453-65.

Wang S, Dong Y, Zhang Y, Wang X, Xu L, Yang S, Li X, Dong H, Xu L, Su L, Ng SS, Chang Z, Sung JJ, Zhang X, Yu J. DACT2 is a functional tumor suppressor through inhibiting Wnt/β-catenin pathway and associated with poor survival in colon cancer. Oncogene. 2015 May 14;34(20):2575-85.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

Zheng Q, Zhou G, Morello R, Chen Y, Garcia-Rojas X, Lee B. Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression *in vivo*. J Cell Biol. 2003 Sep 1;162(5):833-42.

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci USA. 2014 Jan 21;111(3):996-1001.

Chapter 6

Targeted Chromatin Conformation (T2C) analysis identifies novel distal neural enhancers of *ZEB2* in pluripotent stem cell differentiation

Judith C. Birkhoff [1], Rutger W.W. Brouwer [1,2], Petros Kolovos [3], Anne L. Korporaal [1], Ana Bermejo-Santos [1], Ilias Boltsis [1], **Karol Nowosad** [1,4], Mirjam C.G.N. van den Hout [1,2], Frank Grosveld [1], Wilfred F.J. van IJcken [1,2], Danny Huylebroeck [1,5,#], Andrea Conidi [1,#]

[1] Department of Cell Biology, Erasmus University Medical Center, Rotterdam, 3015 CN, the Netherlands
[2] Center for Biomics, Erasmus University Medical Center, Rotterdam, 3015 CN, the Netherlands
[3] Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, 68100, Greece
[4] Department of Biochemistry and Molecular Biology, Medical University of Lublin, 20-093 Lublin, Poland
[5] Department of Development and Regeneration, KU Leuven, Leuven, B-3000, Belgium

[#] *Shared senior authors*

Published (with exception of section 6.3.6) in ***Hum Mol Genet.*** 2020 Aug 29;29(15):2535-50

6.1. Summary

The transcription factor (TF) ZEB2 controls embryonic and adult cell fate decisions and cellular maturation in many stem/progenitor cell types. Defects in these processes in specific cell types underlie several aspects of Mowat-Wilson syndrome (MOWS), which is caused by *ZEB2* haplo-insufficiency. Human *ZEB2*, like mouse *Zeb2,* is located on chromosome 2 downstream of a ±3.5 Mb-long gene-desert, lacking any protein-coding gene. Using temporal Targeted Chromatin Capture (T2C), we show major chromatin structural changes based on mapping in-*cis* proximities between the *ZEB2* promoter and this gene desert during neural differentiation of human induced pluripotent cells (iPSCs), including at early neuroprogenitor cell (NPC)/rosette state, where *ZEB2* mRNA levels increase significantly. Combining T2C with histone-3 acetylation mapping, we identified three novel candidate enhancers about 500 kb upstream of the *ZEB2* transcription start site (TSS). Functional luciferase-based assays in heterologous cells and NPCs reveal co-operation between these three enhancers. This study is the first to document in-*cis* regulatory elements (REs) located in *ZEB2*'s gene desert. The results further show the usability of T2C for future studies of *ZEB2* REs in differentiation and maturation of multiple cell types, and the molecular characterization of newly identified MOWS patients that lack mutations in *ZEB2* protein-coding exons.

6.2. Introduction

Genome-wide identification of regulatory elements (REs), using the mapping of epigenetic modifications and TF binding sites coupled to chromatin conformation capture analyses at different resolutions (3C and beyond), strongly indicate the occurrence of DNA-looping between promoters and distal enhancers in controlling cell fate and differentiation. These transcriptionally functional DNA-loops occur within large interaction-domains, the so-called topologically associating domains (TADs), which themselves are genome-structural loops connected by linker regions (Visel *et al.*, 2009; Nora *et al.*, 2012; de Laat and Duboule, 2013; Long *et al.*, 2016; Polychronopoulos *et al.*, 2017). The intra-TAD DNA-loops formed by in-*cis* promoter-RE proximities lead to precise cell type/stage-specific regulation of gene expression (Thurman *et al.*, 2012; Murakawa *et al.*, 2016; Chatterjee and Ahituv, 2017; Allahyar *et al.*, 2018). REs are often evolutionary highly conserved and usually flank genes shown to control development and cell differentiation, but also metabolic pathways (Ovrachenko *et al.*, 2005). For example, in pluripotent stem cells (PSCs) structural "hubs" can be observed wherein the core pluripotency factors Sox2, Nanog and Oct4 control genes involved in the maintenance of pluripotency (e.g., *Tcf3*, *Smarcad1*) (de Wit *et al.*, 2013). In these PSCs, the transcriptionally inactive regions are less organized, but these regions become more organized during differentiation to somatic cell types. It is also generally accepted that reprogramming of somatic cells, based on transduction of Oct4, Sox2, Klf4 and c-Myc (OSKM) expressable cDNA, results in an almost complete reorganization of the genome architecture, which then becomes similar to that of PSCs.

Zeb2 is a TF critical for vertebrate embryogenesis, including the development of the central and peripheral nervous system (CNS, PNS) (Verschueren *et al.*, 1999; Hegarty *et al.*, 2015). In differentiating mouse embryonic stem cells (mESCs), Zeb2 is needed for the exit from primed

pluripotency and general as well as neural differentiation (Stryjewska *et al.*, 2017). In oligodendrocyte precursor cells (OPCs) during embryonic CNS myelinogenesis and adult Schwann Cell function in PNS (re)myelination, Zeb2 plays a dual transcriptional regulatory role, i.e. directly repressing genes involved in inhibition of differentiation, while directly activating (other) genes promoting cell differentiation and maturation. In doing so, Zeb2 generates the necessary anti-BMP(-Smad) and anti-Wnt, and in the PNS also the anti-Notch and anti-Sox2 activities needed for normal progression of commitment, differentiation and maturation in this glial cell lineage (Conidi *et al.*, 2011; Weng *et al.*, 2012; Quintes *et al.*, 2016; Wu *et al.*, 2016). In humans, *ZEB2* haplo-insufficiency causes the rare Mowat-Wilson Syndrome (MOWS, OMIM #235730). Patients exhibit severe intellectual disability, epilepsy and/or seizures, Hirschsprung disease, and other anomalies including typical craniofacial defects (Garavelli *et al.*, 2003; Zweier *et al.*, 2005; Ivanovski *et al.*, 2018). Analysis of the spectrum of the *de novo* mutant *ZEB2* alleles in a more recent cohort of 87 patients indicated for the first time that the severity of MOWS may correlate with the type of mutation (Ivanovski *et al.*, 2018). *De novo* deletions in the *ZEB2* gene that involve protein-coding exons or cause protein C-terminal truncation due to mutation into a stop codon, as well as even larger genomic deletions, cause severe defects, whereas the few known missense mutations (1.5% of about 320 exon-sequenced MOWS patients thus far) present with a milder form of the syndrome (Garavelli *et al.*, 2009; Ivanovski *et al.*, 2018).

Additional work done in mouse models, including rescuing conventional or conditional *Zeb2*-KO backgrounds via introduction of a *Zeb2* cDNA (provided as heterozygous or homozygous transgene), strongly suggests that proper control of Zeb2 amounts, including via steady-state mRNA levels, is critical for normal Zeb2 functions, as observed from the graded phenotypic severities of the *Zeb2*-KO and/or transgene combinations (Tatari *et al.*, 2014; Goossens *et al.*, 2015; Scott *et al.*, 2016; Scott *et al.*, 2018). In this respect, little is however known about the precise and temporal transcriptional control of *ZEB2*, and only few studies have thus far focused on candidate TFs that bind to the *ZEB2* proximal promoter (Katoh and Katoh, 2009; Qiao *et al.*, 2015). Recently, different enhancers were identified, mainly by documenting evolutionary conserved *ZEB2* containing and flanking regions, followed by validation in zebrafish, rodent models, and/or *in vitro* cellular models. In addition, in a transgenic rat model, *Zeb2* is regulated in a tissue and time specific manner by an enhancer located 1.2 Mb upstream of the TSS (El-kasti *et al.*, 2012). In the subpallium of the developing mouse brain, two enhancers flanking the *Zeb2* locus have been proposed to be activated by the TF Dlx2 (McKinsey *et al.*, 2013). More recently, combining publicly available databases of chromatin interaction and mapped histone signatures, again followed by validation in zebrafish, eight enhancers were identified in intergenic, intronic and exonic sequences of *ZEB2* (Bar Yaacov *et al.*, 2019); these enhancers are active in mid-/hindbrain regions, trigeminal ganglia, notochord or the whole brain.

Human *ZEB2* is located on chr2:145141942-145277958 (genome release GRCh37/hg19), downstream of a 3.3 Mb-long region lacking protein-coding sequences, which encodes several non-coding RNAs. The sequence of this region is locally highly conserved between different species, despite the differences in length of this region (chimp: 3.4 Mb; mouse: 3.7 Mb; chicken: 1 Mb; *Xenopus*: 1 Mb). This gene desert is located between *ZEB2* and *ACVR2A*, which encodes for the Activin type-IIA receptor, a component of the TGFβ/BMP signaling system (Walton *et al.*, 2012). In a separate study, a map of non-coding elements involved in human cortical neurogenesis was obtained by combining chromatin accessibility and mRNA profiling data (de la Torre-Ubieta *et al.*, 2018). Several non-protein coding elements are in proximity of *ZEB2*, including the long non-coding RNA (lncRNA) *LINC01412*, which maps roughly 2 kb upstream

of the *ZEB2* TSS. In a genome-wide association study (GWAS) of more than 2,400 cases of aortic valve stenosis (a pathology that hits about 5% of MOWS patients), this region harbors single-nucleotide polymorphisms (SNPs) in the non-coding RNA *TEX41*, located about 150 kb upstream of the *ZEB2* TSS, that directly interact with *LINC01412* and the *ZEB2* proximal promoter region (Helgadottir *et al.*, 2018; Ivanovski *et al.*, 2018). A schematic overview of the *ZEB2* locus, with published enhancers and chromatin interactions that co-regulate *ZEB2*, is depicted in **Fig. S6.1**.

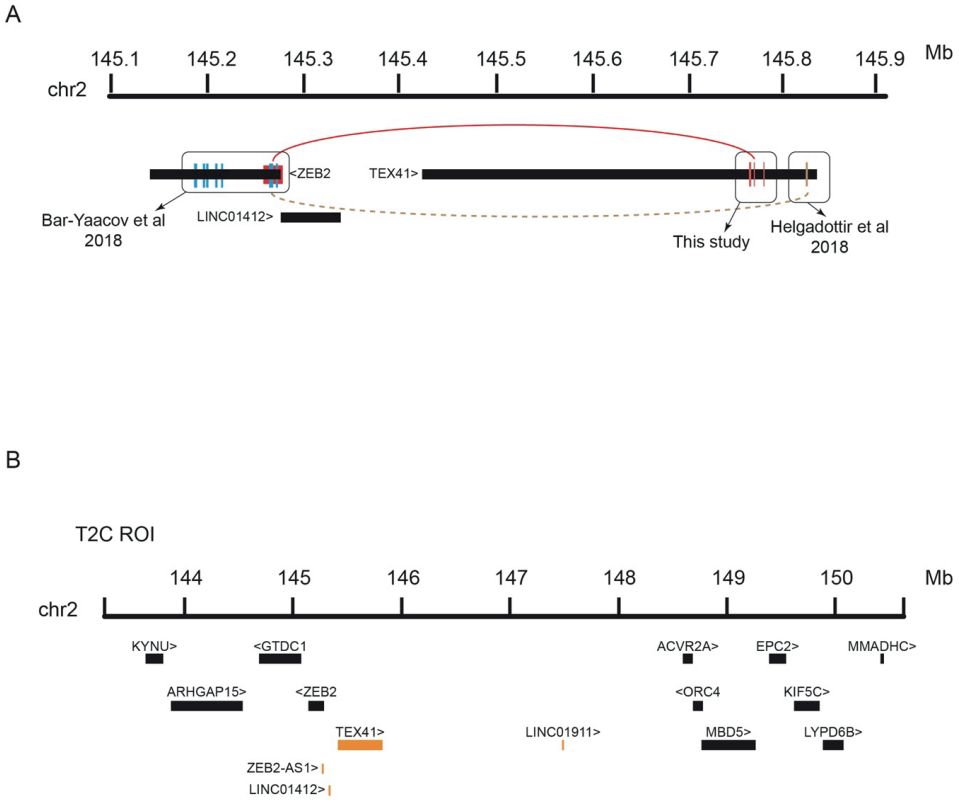


Figure S6.1. Schematic overview of published enhancers and chromatin architecture changes and overview of T2C Region of Interest (ROI) in this study.
A. Known identified *ZEB2* intragenic enhancers as described in Bar-Yaacov *et al.* (2018) and of predicted looping between SNPs located in *TEX41* (a non-coding RNA) and *ZEB2* (Helgadottir *et al.*, 2018). **B.** T2C ROI considered for this study with protein-encoding genes (black boxes) and non-coding RNAs (orange boxes). <, >: indicates the direction of transcription.

Altogether, and also considering its listing as a super-enhancer top-gene (Hnisz *et al.*, 2013; Fufa *et al.*, 2019), the gene desert upstream of the human *ZEB2* becomes a priority for identifying candidate and/or pathologic *ZEB2* REs. Given the critical role of *Zeb2* during exit from primed pluripotency, its dynamic regulation during neural and general differentiation of mESCs, and applied rescues in knockout stem cells with inserted *Zeb2* expressible cDNA (Stryjewska *et al.*, 2017), we decided to study chromatin conformation dynamics of the human *ZEB2* locus during

neural differentiation of iPSCs. Several chromosome conformation capture techniques (3C, 4C, 5C, Hi-C, ChIA-PET) have been developed to investigate and characterize spatial genomic organization by chromatin interactions (Dekker *et al.*, 2002; Gondor *et al.*, 2008; Lieberman-Aiden *et al.*, 2009; Sexton *et al.*, 2012; Stadhouders *et al.*, 2013). These techniques are mostly expensive, require extensive primer design and have a resolution of tens of kb. Recently, Targeted Chromatin Capture (T2C) was shown to virtually provide high resolution (in the order of few kb or even less than 1 kb) and combine this with high coverage and low sequencing efforts, hence at a contained cost (Kolovos *et al.*, 2014). We therefore aimed to study the chromatin dynamics of the *ZEB2* locus, during iPSCs neural differentiation, considering the whole gene desert and both its flanking regions, for a total genomic region of 7.4 Mb in length (coordinates chr2: 143270465-150642631; GRCh37/hg19 genome reference). **Fig. S6.1B** shows this region of interest (ROI) used in this T2C study. By correlating chromatin architecture reconstruction via T2C at an average resolution below 1 kb with H3K27ac marks, RNA-profiling at selected time points of cellular neural differentiation, and further taking into consideration locus sequence conservation in vertebrates, we identified three novel candidate *ZEB2* enhancers. Our work demonstrates, for the first time, the dynamic regulation of *ZEB2* expression by distal REs that loop to the *ZEB2* promoter during cell differentiation. These studies are expected to open the road to improved and/or expanded genetic and additional functional characterization of those MOWS patients for whom no mutation affecting the protein-encoding sequence of *ZEB2* can be identified.

6.3. Results

6.3.1. Transcriptomic profile of neural differentiating human iPSCs

RNA-sequencing (RNA-seq) of undifferentiated cells (D0), early NPCs/neural rosettes (D6) and late NPCs (D15) was carried out (**Fig. 6.1A**; *see* section 6.5). Principal component analysis (PCA) shows the clustering of the samples based mainly on the time point of differentiation (PC1, **Fig. 6.1B**). *ZEB2* mRNA, as well as the transcripts of the second ZEB-family member *ZEB1*, is upregulated already at D6 of differentiation (**Fig. 6.1C**; **Fig. S6.2A**). The acknowledged ZEB1/2 direct target gene *CDH1*, encoding for the epithelial cell specific homotypic cell-cell adhesion protein E-cadherin, was concomitantly downregulated already at D6 (**Fig. 6.1**; **Fig. S6.2A**). Its expression inversely correlated in the bulk cell cultures with that of the N-cadherin encoding gene *CDH2* (**Fig. 6.1C**; **Fig. S6.2A**). Expression of genes encoding the core components of the pluripotency network, such as *NANOG*, *POU5F1* (*OCT4*), *SOX2,* and *ZFP42* (*REX1*), is downregulated upon differentiation (**Fig. 6.1C**; **Fig. S6.2B**). *SOX2* is however also critical to neurogenesis, and its expression - even though decreased during differentiation - remains high when compared to other pluripotency genes (Suh *et al.*, 2007) (**Fig. S6.2B**). Conversely, the expression of neuroectoderm marker genes, such as *GBX2*, *PAX6*, *SIX1* and *SOX1* increased (**Fig. 6.1C**; **Fig. S6.2C**). Expression of mesendodermal genes *EOMES*, *GATA4*, *SOX17* and *TBXT* (*T*, *BRA*) was in this cell culture protocol not significant (**Fig. S6.2D**). Of the mesendodermal markers, only *EOMES* is upregulated at D6 and D15 compared to D0, most likely due to its proposed role in neurodevelopment (Baala *et al.*, 2007; Mihalas and Hevner, 2017). We have also performed staining for pluripotency marker proteins OCT4 and SOX2, for NPC markers NESTIN, TUJ1, PAX6 and NCAM, and for ZEB2, at the considered time points of differentiation (**Fig. 6.1D**). Taken together, the neural differentiation protocol of the initial human iPSCs was highly efficient.

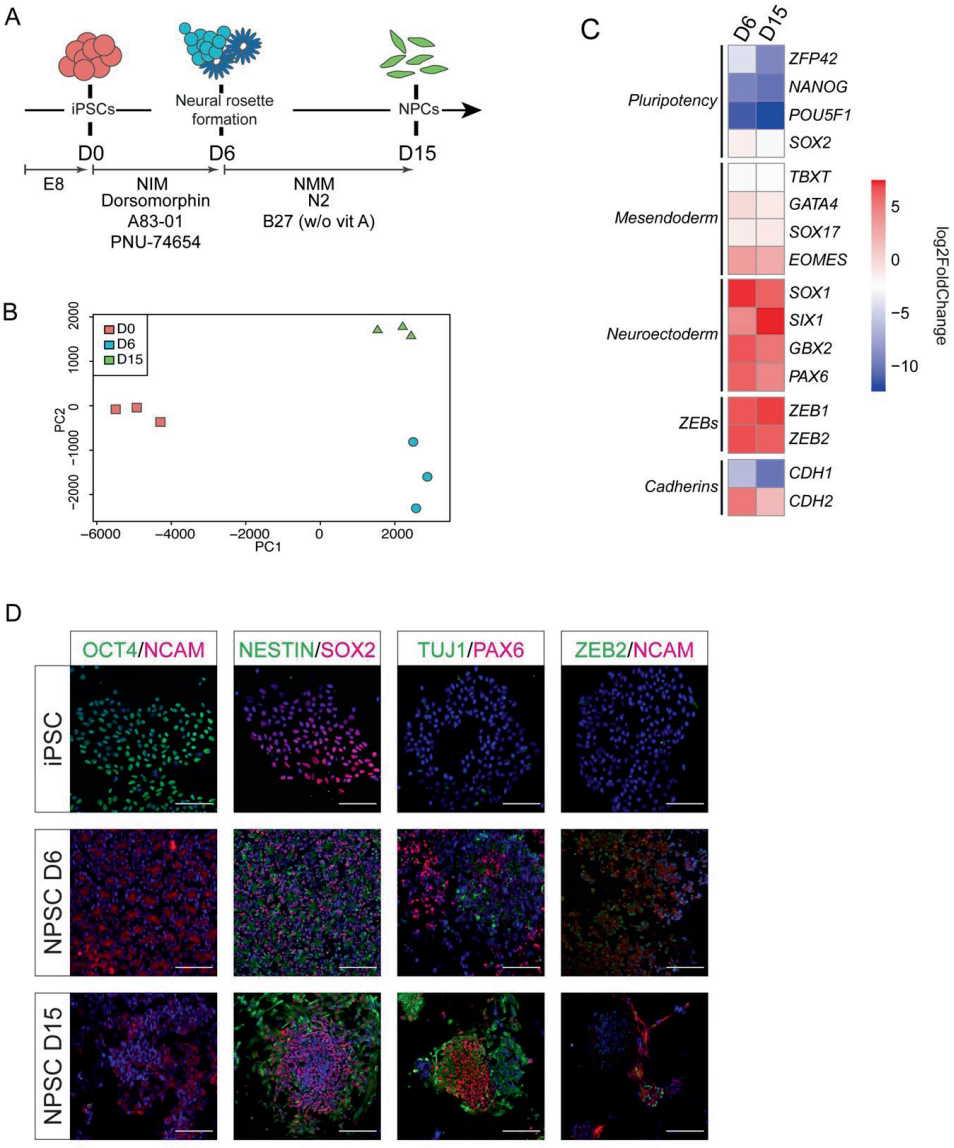


Figure 6.1. Gene expression profiling of differentiating iPSCs.

A. Schematic overview of the differentiation protocol used, including specific media and inhibitors (*see* Materials and Methods). *Abbreviations used*: EB: embryoid body; NIM: Neural Induction Medium, NMM: Neural Maturation Medium. D (day) 0: undifferentiated state, D6: early NPC/neural rosette formation, D15: NPCs. w/o vit A: no vitamin A added to B27. **B.** PCA plot showing clustering of biological repeat (n=3) RNA-seq samples based on time of differentiation. **C.** log2Fold Change heatmap of selected marker genes confirming progression from pluripotency to efficient neural differentiation. **D.** Immunofluorescence staining at selected timepoints of differentiation for pluripotency (OCT4, SOX2) and NPC markers (NCAM, NESTIN, TUJ1, PAX6) and ZEB2. Scale bar = 100µm.

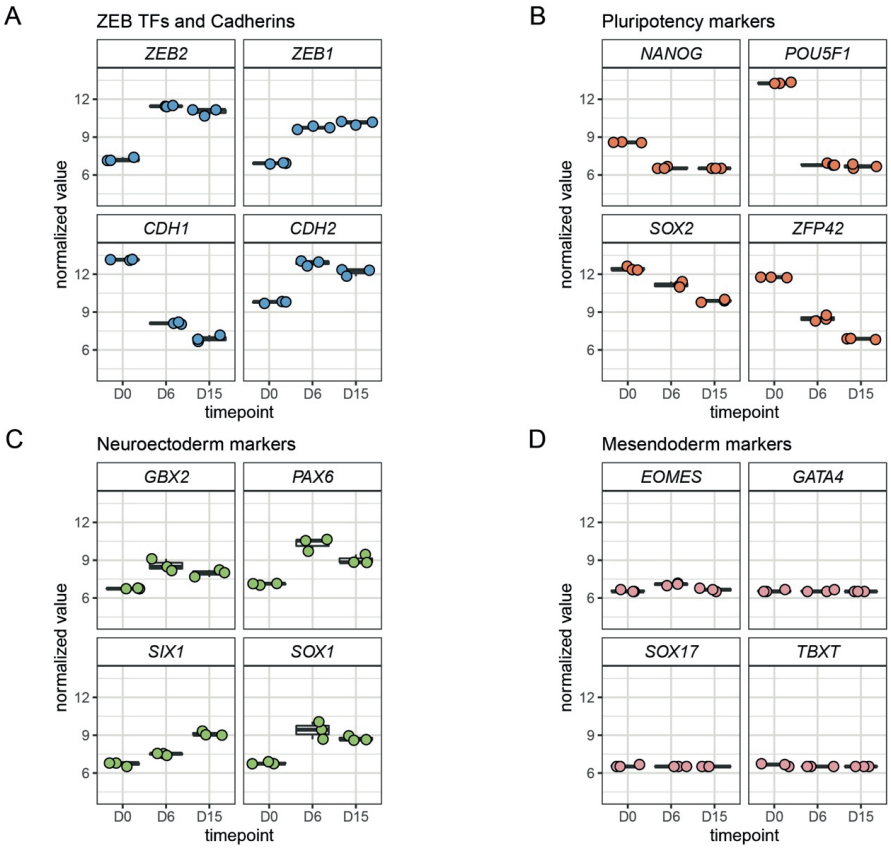


Figure S6.2. Normalized steady-state RNA values of *ZEB2* and selected marker genes at the considered time points of neural differentiation.

The same genes considered for differential expression in **Fig. 6.1** are assessed here for their temporal dynamics.

6.3.2. Chromatin dynamics of *ZEB2* locus during neural differentiation

We studied the regulation of the *ZEB2* locus by temporal T2C using the iPSC system at times documented above. First, we calculated the average size and density distribution of the fragments generated by *Apo*I cut in the considered area. Density distribution and frequency of the fragment size plots show a mean fragment size (dashed line in each panel of **Fig. S6.3A**) of ~500 bp for *Apo*I, which did not change during differentiation. Reconstruction of high-resolution T2C maps at the different time points, plotting the single fragments obtained from the digestion with *Apo*I, resulted however in a very sparse and unclear figure (*data not shown*). Therefore, we opted for reducing the resolution of our T2C maps by binning the signals to a resolution of 20 kb, resulting in an easier graphical interpretation (**Fig. 6.2**). On the other hand, the *Apo*I fragment resolution becomes very useful when zooming in on relatively small regions (such as these shown in **Fig. S6.4**), depicting the *Apo*I fragment proximity interactions on the *ZEB2* gene *per se* (chr2:145,141,942-145,277,958, hg19) (*see* **Supplementary File S6.1**).

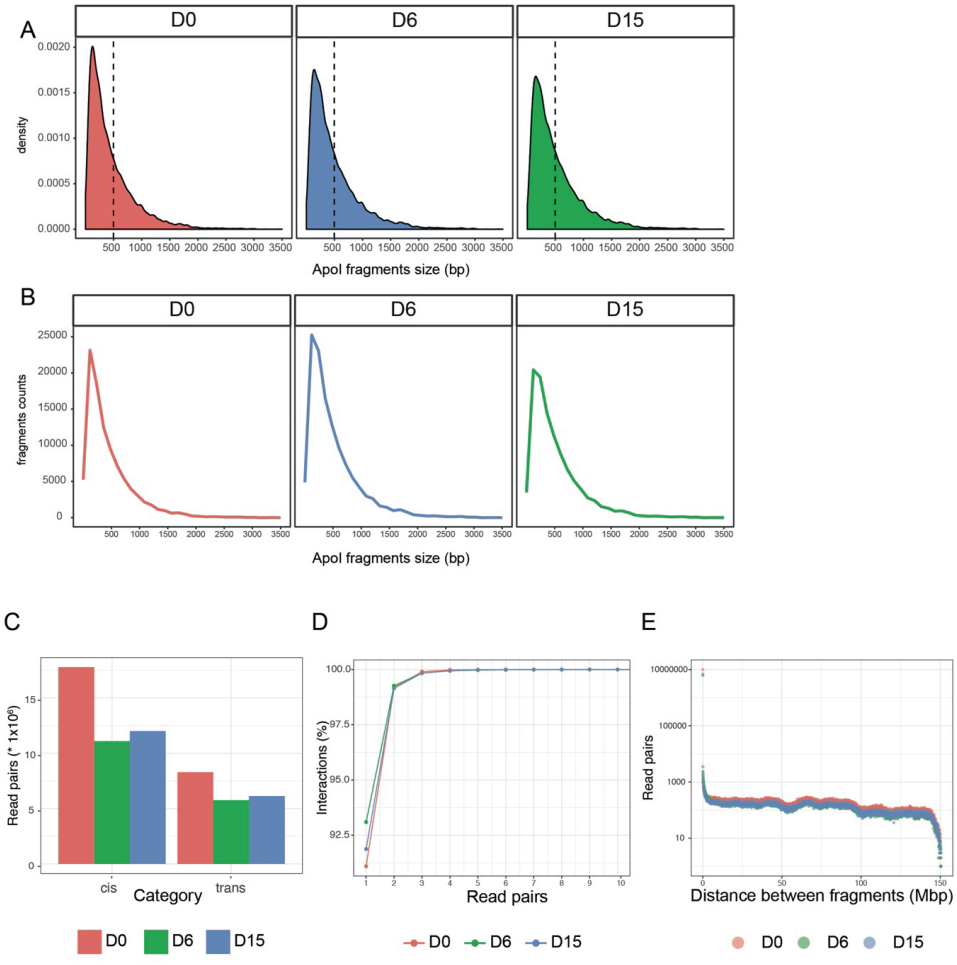


Figure S6.3. Distribution and frequency of fragments generated by *Apo*I in the T2C region of interest and T2C quality controls.

A. Density plot of fragment distribution according to the size of the obtained fragments. **B.** Fragment counts versus the size of the individual fragments generated by *Apo*I. **C.** Number of *cis* and *trans* targeted read-pairs in the proximity matrix. **D.** Proportion of targeted interaction over 1 kb apart, as compared with their score in read pairs. **E.** Cumulative number of read-pairs versus the distance of binned (per 50 kb) targeted interactions.

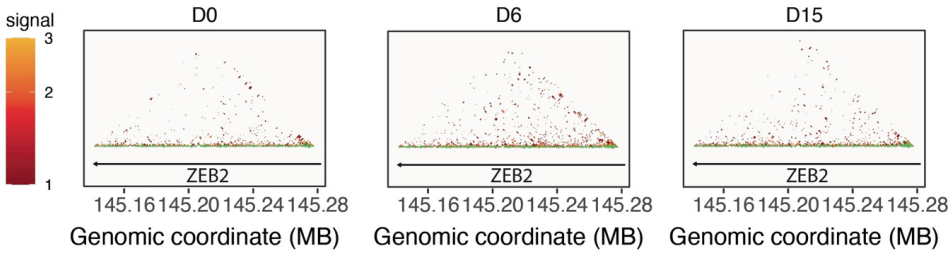
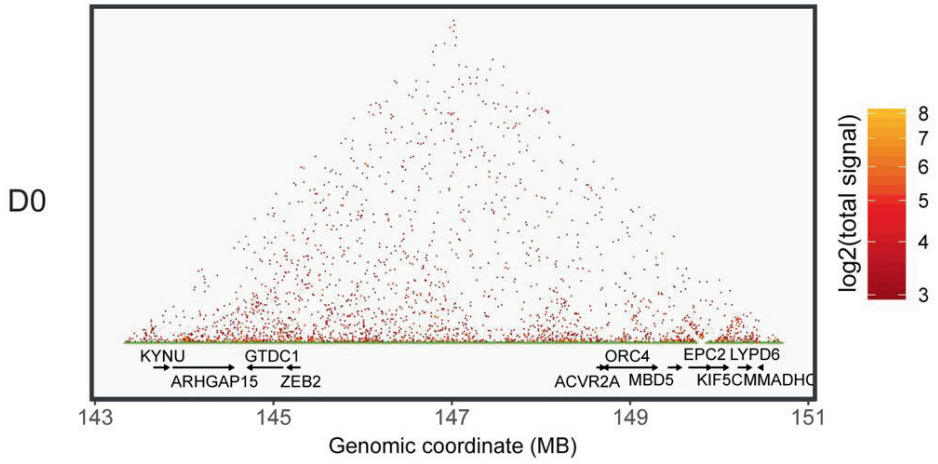


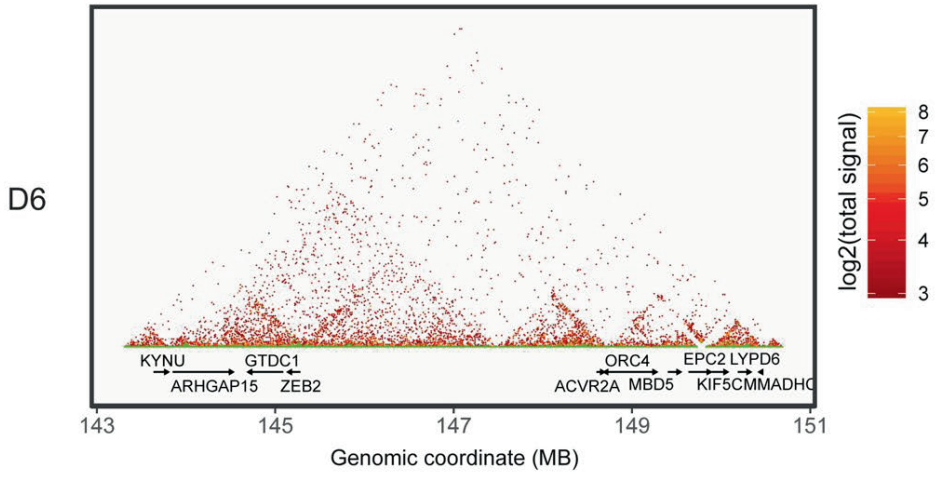
Figure S6.4. ***Apo*I fragments-based high resolution T2C maps of *ZEB2* gene chromatin dynamics during differentiation.**

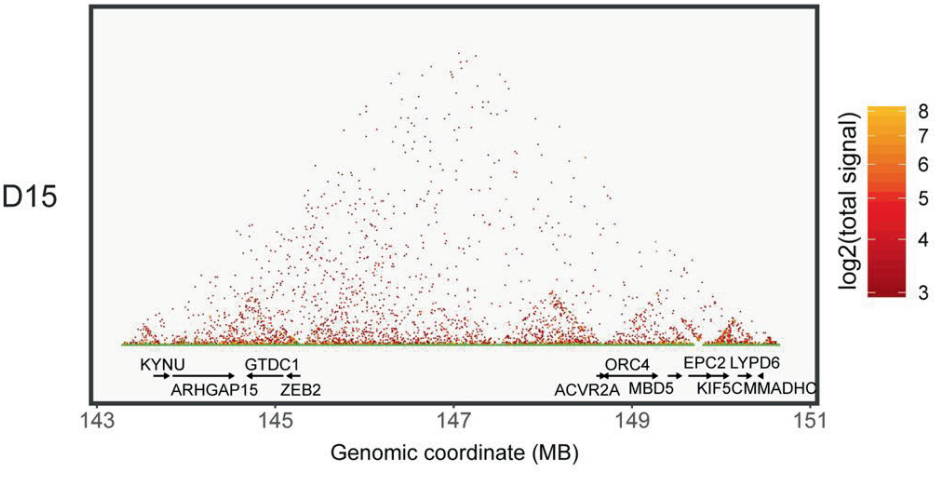
T2C maps can be reconstructed at *Apo*I fragments high resolution for local regions of ~100-150 kb as in the case of *ZEB2* (136 kb) confirming the existence of several intragenic loopings.

Figure 6.2. (*Next page*). ***ZEB2* locus dynamics during neural differentiation.**

Reconstruction of in-*cis* proximity interactions on the human chr2:143:151 region, including the *ZEB2* locus and the other annotated genes, in pluripotent (top panel, D0) and neural differentiating (D6, D15) iPSCs. To improve graphical clarity the signals were binned to get a resolution of 20 kb.


D0
log2(total signal)
8
7
6
5
4
3
KYNU
ARHGAP15
GTDC1
ZEB2
ORC4
ACVR2A
MBD5
EPC2
KIF5C
LYPD6
MMADHC
143
145
147
149
151
Genomic coordinate (MB)


D6
log2(total signal)
8
7
6
5
4
3
KYNU
ARHGAP15
GTDC1
ZEB2
ORC4
ACVR2A
MBD5
EPC2
KIF5C
LYPD6
MMADHC
143
145
147
149
151
Genomic coordinate (MB)


D15
log2(total signal)
8
7
6
5
4
3
KYNU
ARHGAP15
GTDC1
ZEB2
ORC4
ACVR2A
MBD5
EPC2
KIF5C
LYPD6
MMADHC
143
145
147
149
151
Genomic coordinate (MB)

As shown in **Fig. 6.2**, undifferentiated iPSCs revealed a diffuse pattern of proximity interactions with few appreciable TADs downstream of *ZEB2*, as well as along the gene desert and around the *ACVR2A* locus. At D6 (early NPCs), the TADs became more pronounced, with a long ~4 Mb TAD (chr2:143-147 Mb) being strongly defined and encompassing at least three sub-TADs. One, which we define as TAD1 (**Figs. 6.2 and 6.3B**), bridged *ZEB2* to *ARHGAP15*; another (named TAD2) did so between *ZEB2* and a region located upstream its TSS in the gene desert (i.e. around chr2:146 Mb) (**Figs. 6.2 and 6.3C**); and one (TAD3) between chr2:146 and chr2:147 Mb (**Figs. 6.2 and 6.3D**). The major chromatin conformation change observed at D6 was concomitant with high *ZEB2* mRNA in early NPCs, while the less pronounced TADs at undifferentiated state were associated with low *ZEB2* expression (**Figs. 6.1 and 6.2**). At D15 (NPCs), the 4-Mb TAD seemed less defined, whereas the three sub-TADs were still defined, even though their proximity signals were reduced, suggesting a slight loosening of the chromatin architecture (**Fig. 6.2**).

We also propose that several other TADs relate to the other genes, which are located in the same broad chromosomal region (*KYNU*, *ARHGAP15*, *GTDC1*, *ACVR2A*, *ORC4*, *MBD5*, *EPC2*, *KIF5C*, *LYPD6B,* and *MMADHC*) (**Fig. 6.2**). We therefore checked in our RNA-seq data whether transcription of these genes correlated with the formation of these TADs in the gene desert. In general, expression of these genes did not change during neural differentiation (**Fig. S6.5A**).

The sub-TAD TAD1 (chr2:144-145) bridges *ZEB2* with *ARHGAP15*, which encodes for a Rho-GTPase Activating Protein, is known to direct expression in both excitatory and inhibitory neurons of the adult hippocampus and midbrain (Zamboni *et al.*, 2016; Thul *et al.*, 2016; Human Protein Atlas; https://www.proteinatlas.org) (**Fig. 6.2**). *ARHGAP15* mRNA expression was slightly upregulated at D6 in our RNA-seq data (**Fig. S6.5A**). **Figure 6.3 B,C** shows that *ZEB2* sequences are not only in proximity with *ARHGAP15*, but also that - at D6 - *ZEB2* bridged with the 5'-region of *GTDC1*. According to the Human Protein Atlas database, highest levels of GTDC1 are observed in the cerebral cortex (Thul *et al.*, 2017).

Figures 6.3A,C and S.6.4 also clearly show a local DNA-looping of the *ZEB2* gene itself, which formed at D6 and then disappeared at D15, supporting therefore the results from Bar-Yaacov *et al.* (2019) who identified *ZEB2* intragenic enhancers. *ACVR2A*, which flanks the gene desert, did not show a significant change in expression, even though TADs were forming and involved its coding sequence (**Fig. 6.2A**). *ACVR2A* is highly expressed in skin and skeletal muscle rather than brain regions, where it does not show any significant regional expression (Thul *et al.*, 2017; The Human Protein Atlas).

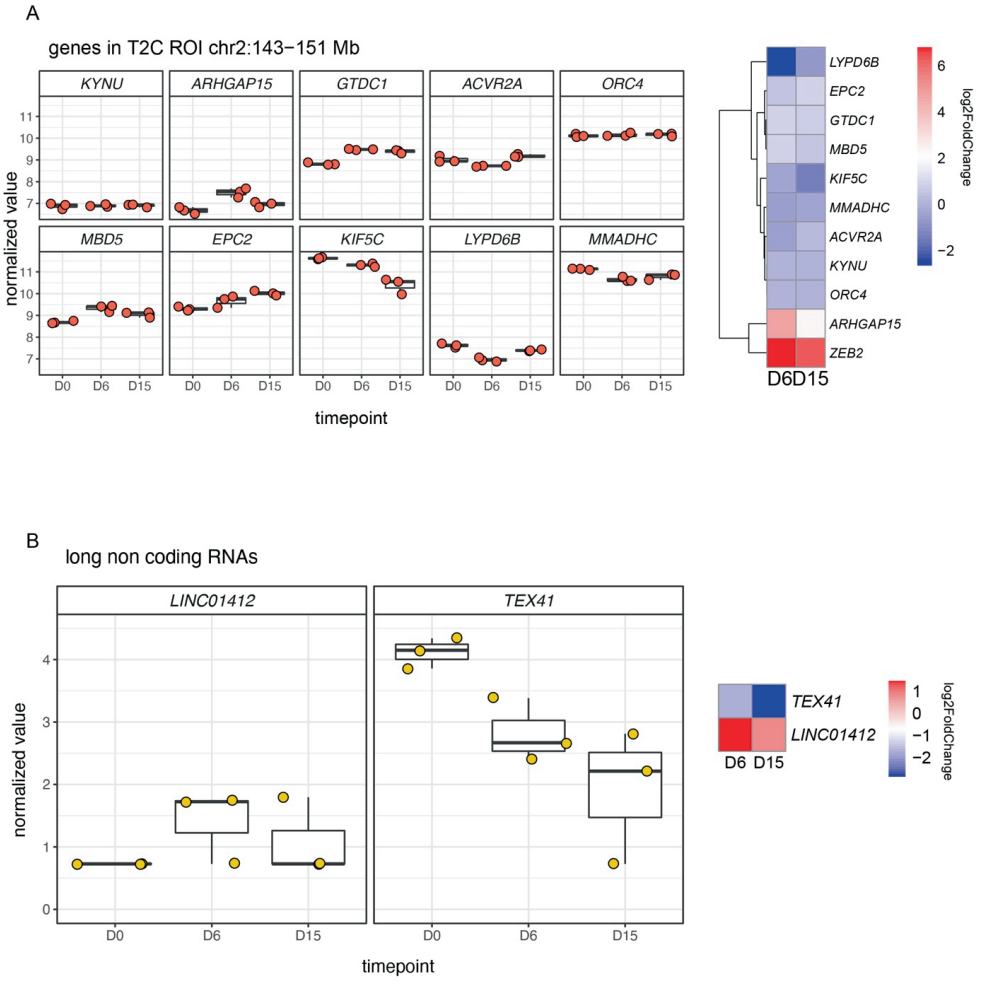


Figure S6.5. mRNA expression of annotated genes and long non-coding RNAs located in the considered T2C region of interest (ROI).

A. Normalized values and log2 Fold Change heatmap of genes. **B.** Non-coding RNAs located in the T2C ROI.

6.3.3. T2C mapping and H3K27Ac marks identify three novel candidate enhancers for *ZEB2*

When zooming-in at D6 on the sub-TAD TAD2, formed by *ZEB2* and chr2:146, we noticed long-range proximity defining a sub-TAD that bridges ~chr2:145,260,000 (*ZEB2*) to ~chr2:145,780,000 (upstream gene desert) (blue arrow in **Fig. 6.3C**; *see* **Supplementary File S6.2**). The coordinates for the mapped proximity are chr2:145,260,000-145,280,000 (region A) and chr2:145,760,000-145,780,000 (region B). In the hg19 release, region A corresponds to the first 20 kb of *ZEB2*, including the promoter and (the non-coding) exon1 and (protein-encoding) exon2, while region B is located in an intron of the lncRNA *TEX41*, in the upstream gene desert (**Fig. 6.3C,E**). Another strong interaction signal forming a loop of 295 kb was also present (with coordinates chr2:145,305,000-145,310,000 and chr2:145,600,000-145,605,000) in the same area and formed another small TAD (black arrow in **Fig. 6.3C**). The coordinates for this proximity fall in one intron of *LINC01412* and one intron of *TEX41*. Both *LINC01412* and *TEX41* are very low expressed during neural differentiation of iPSCs (**Fig. S6.5B**). H3K27ac marks are not present for this 295 kb proximity interaction (*data not shown*), whereas they are at the border of the sub-TAD formed by region A and B (**Fig. 6.3E**).

We therefore focused on the possible REs between region A and B that are associated with active enhancer marks (**Fig. 6.3E**). Combining H3K27ac marks and conservation tracks, region B can be divided in three clusters that possibly represent three novel candidate enhancers, named Enh1 (chr2:145,764,483-145,765,504), Enh2 (chr2:145,769,677-145,770,210) and Enh3 (chr2:145,779,965-145,780,193) (**Fig. 6.3E**). Taken together, these data suggest a time-regulated DNA-looping with the aim of bringing the three enhancers and the *ZEB2* promoter in close proximity. Furthermore, this looping might specifically regulate *ZEB2* promoter, and gene transcription, during neural differentiation.

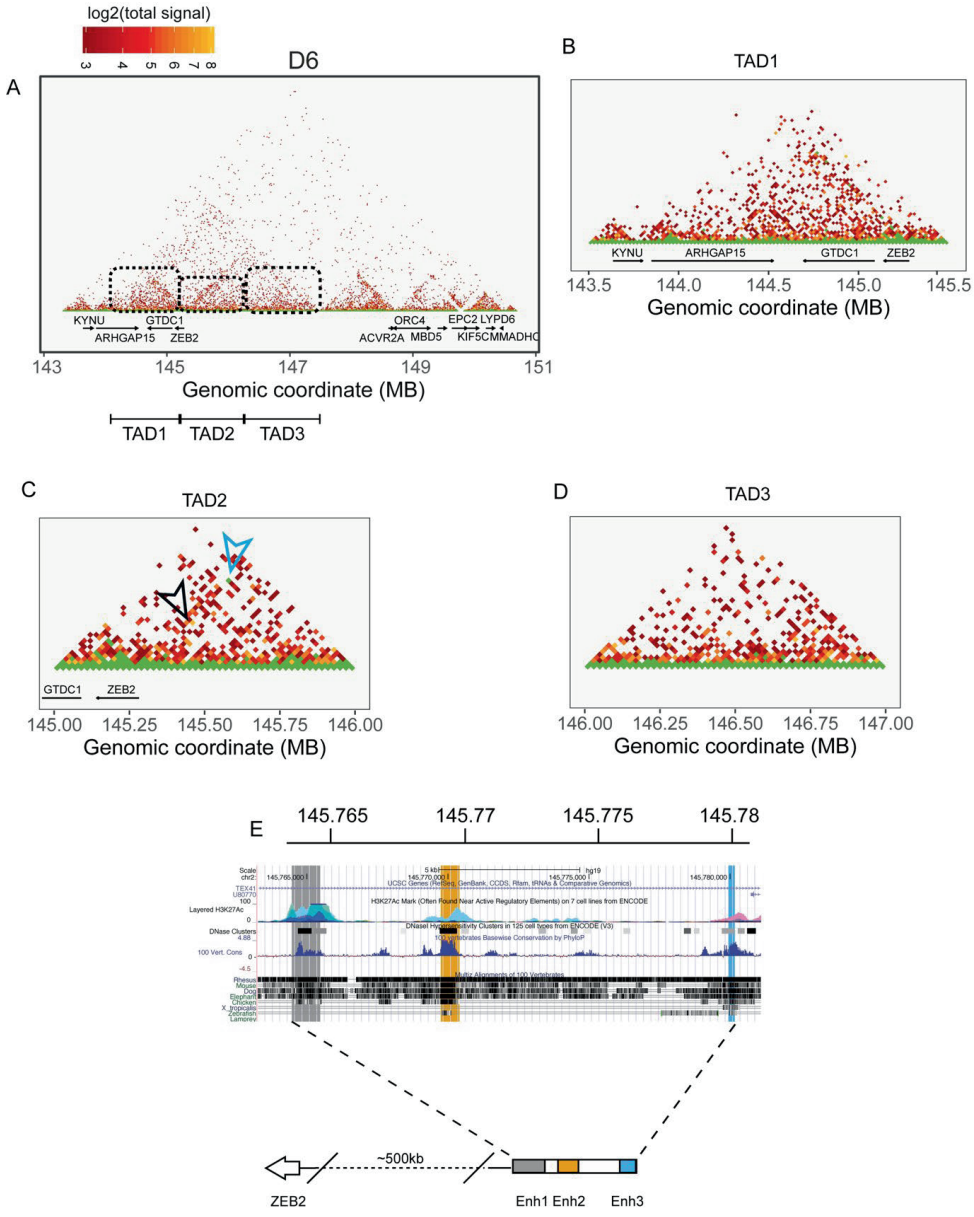


Figure 6.3. The main 143-147 Mb TAD defined at D6 is composed of three sub-TADs.

At D6 (**panel A**), the major chromatin conformation encompassing the *ZEB2* locus and upstream region, is evident, displaying more intragenic proximity interactions and also three well-defined TADs located between *ARHGAP15* and *ZEB2* (TAD1, **panel B**), between 145-146 Mb (TAD2, **panel C**) and a third one located in the *ZEB2* upstream gene desert between 146-147 Mb (**panel D**). At D6 the strongest signal is observed in a TAD structure defined by region coordinates 145,260,000-145,280,000 to 145,760,000-145,780,000. Zooming-in on the coordinates chr2:145,760,000-145-780,000 allows the definition of three possible well conserved enhancers defined by H3K27Ac (**panel E**).

6.3.4. The three novel candidate enhancers act on *ZEB2* promoter-based transcription

To functionally investigate the candidate enhancer regions identified by T2C in iPSCs at D6 of neural differentiation, we cloned these enhancers in combination with the *ZEB2* proximal promoter (chr2:145,277,927-145,278,000) in a luciferase-reporter based vector and transfected these respective constructs in iPSCs at the different time points of differentiation (**Fig. 6.4A,B**). A similar basal level of luciferase activation is seen at both at D0 (undifferentiated cells) and at D15 (mature NPCs), whereas at D6 the activation peaks to about 8-fold higher values (**Fig. 6.4B**), indicating that the three enhancers, tested away from their normal location, are bound by one or more transcriptional regulators specifically produced and active at this cell state.

We also produced combinatorial versions of the enhancers (**Fig. 6.4A**) and transfected the entire panel of enhancers-promoter combinations to heterologous HEK293T cells (**Fig. 6.4C**). The presence of all three enhancers and the promoter had the strongest effect on the vector-based luciferase activity. Enh2 in combination with either Enh1 or Enh3 also induced luciferase activity, albeit at a lower level. Remarkably, Enh2 by itself was not able to induce luciferase, indicating a co-operative effect of the three enhancers with Enh2, which enhances the stimulatory effect of the other two enhancers. The activity of Enh1 and Enh3 appeared to be additive, but only if Enh2 was present. We conclude that the three enhancers co-operate, including in the neural lineage, in driving *ZEB2* expression, and Enh2+Enh3 are required for sustained *ZEB2* transcription.

6.3.5. *In-silico* motif analysis predicts novel, remote-acting TF candidates for *ZEB2* transcription regulation

Next, we performed an *in-silico* prediction analysis of TF-binding elements present in the three enhancers. We used JASPAR database for human TF motif profiles, considering a >90% confidence score (http://jaspar.genereg.net). Among the many and different motifs defined as such, we found collective enrichment for ETS1, FOXD2, HOXB2, LHX1 and 9, OTX2, SOX10 and 15, and YY1 in the enhancers (**Fig. 6.5A**). In particular ETS1 seems a candidate for binding to Enh1 and Enh2, whereas Enh3 has just SOX15 and HOXB2 passing the applied 90% confidence threshold. **Figure 6.5B** shows the log2 Fold Change (log2FC) and **Fig. S6.6** the normalized values of the mRNAs for these TFs, as determined via our RNA-seq. *FOXD2* and *HOXB2* are the top upregulated genes among the possible TFs involved, while *SOX10* and *ETS1* show a very moderate upregulation (**Fig. 6.5B**). The other possible TFs are overall downregulated during differentiation (**Fig. 6.5A**).

To verify that FOXD2 and/or HOXB2 regulate the enhancers' activity, we transfected NPCs with shRNAs directed against these two TFs and also for *SOX10*, for which a crosstalk with *ZEB2* is already known (Van de Putte *et al.*, 2003; Stanchina *et al.*, 2006; Stanchina *et al.*, 2010; Van de Putte *et al.*, 2017; Watanabe *et al.*, 2017) (**Fig. 6.5C**). The knockdown (KD) of HOXB2 and SOX10 resulted in increased luciferase activity, while FOXD2 KD has no significant effect on the enhancers. Therefore, we propose that upstream HOXB2 and SOX10 might be necessary for or contributing to transcriptional repression of *ZEB2*.

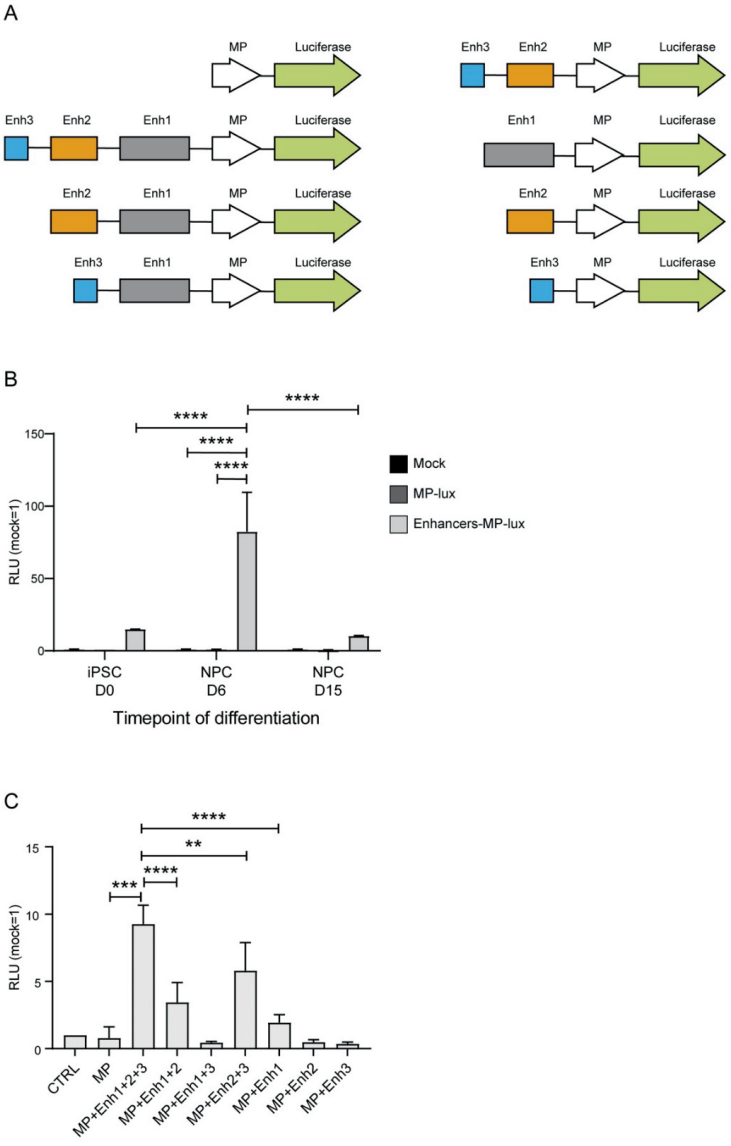


Figure 6.4. The three novel identified enhancers co-operate to drive upregulated activity of the minimal *ZEB2* promoter in a cell state/time specific manner.

A. Schematic overview of combinatorial luciferase reporter-based constructs used in this study. **B.** Luciferase assay performed in differentiating iPSCs transfected with a luciferase-based construct containing the three identified enhancers together with the minimal promoter of ZEB2 shows differential luciferase activation according the cell state A basal level of activation can be observed already at undifferentiated, D0, state. Similar level is also detected at mature NPC state, D15, while the highest expression is at D6, neural rosette/ early NPC state. **C.** Luciferase reporter assay of the whole panel of constructs, transiently transfected to heterologous HEK293T cells (*see* also section 5.5). As for NPCs, the three enhancers positively co-operate. MP: Minimal promoter of ZEB2. Error bars represent standard deviation of three independent biological replicate. Statistical significance was calculated with GraphPad Prism using a multiple comparison one-way Anova test. Asterisks represent p-values: * $p<0.05$; ** $p<0.005$; *** $p<0.0005$; **** $p<0.0001$.

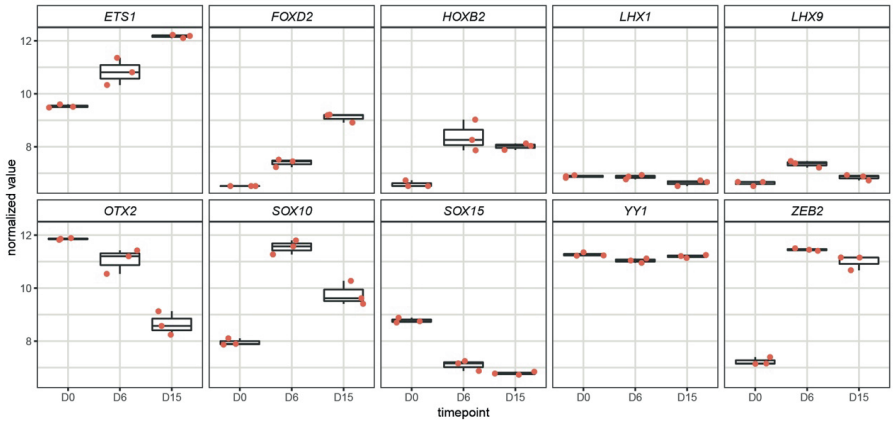
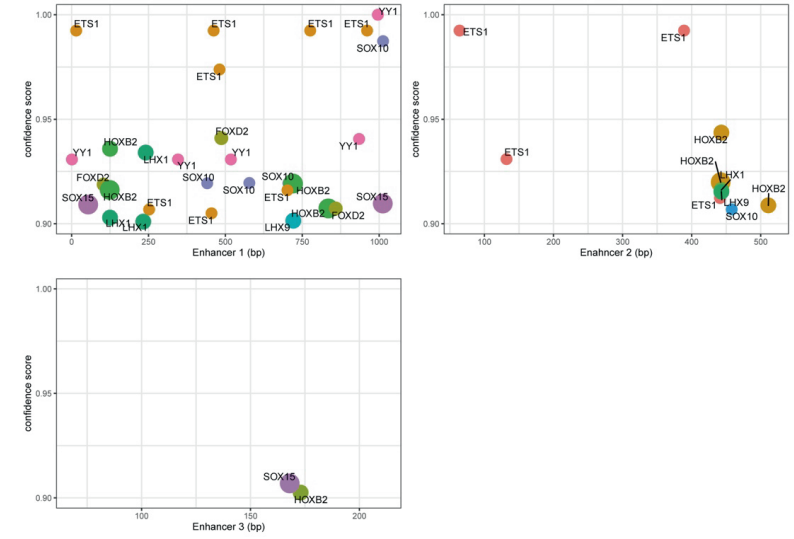


Figure S6.6. mRNA expression of transcription factors for which a consensus motif has been identified in the newly identified enhancers.



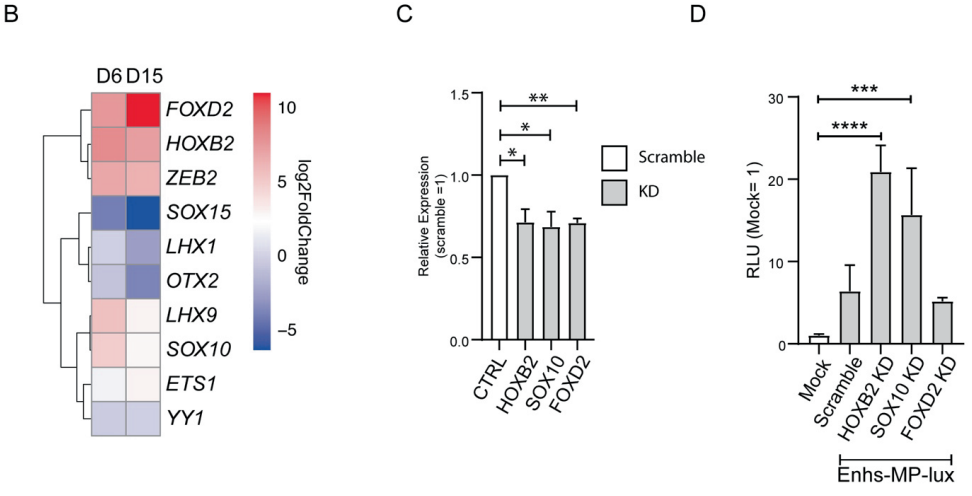


Figure 6.5. HOXB2 and SOX10 can bind the newly identified enhancers to regulate the *ZEB2* minimal promoter activity.

A. Localization of TF motifs in the sequence of the three enhancers using a confidence score >90%. **B.** Differential expression, during iPSCs neural differentiation, of TFs for which a binding motif has been found in the enhancers' sequences. FOXD2, HOXB2 are upregulated at D6 and D15 and might represent potential candidates able to bind the enhancers. **C.** mRNA levels of HOXB2, SOX10 and FOXD2 after shRNA mediated KD in NPCs. **D.** KD of HOXB2 and SOX10, but not of FOXD2, results in increased luciferase activation mediated by the three novel enhancers and the ZEB2 minimal promoter. Error bars represent standard deviation of three independent biological replicate. Statistical significance was calculated with GraphPad Prism using a multiple comparison one-way Anova test. Asterisks represent p-values: *** $p<0.0005$; **** $p<0.0001$.

6.3.6. *Annex*: T2C mapping for *Zeb2* locus in mouse ESCs

T2C was also applied during RA-induced neural differentiation of mouse ESCs (*see* **Fig. S6.7A**). Zeb2 mRNA levels increased during differentiation, with a first significant upregulation already at D4 (**Fig. S6.7B**), with *Zeb1* following similar dynamics. Neural differentiation was verified through NPC marker *Pax6*. As expected, pluripotency-associated genes were downregulated upon differentiation. T2C analysis in multipotent progenitors (D0), epiblast progenitors (D4), early neural progenitors (D6) and neural progenitors (D8) was then performed. In undifferentiated mESCs (D0; **Fig. S6.7C**) few 1 Mb-sized TADs are formed centered on *Zeb2* and *Acvr2a*. Interestingly, at D4 major chromatin rearrangements could be observed, with long-range proximities involving *Zeb2* and regions within the upstream gene desert. At early and late neural commitment stage (D6 and D8, respectively) barely any strong interaction could be observed. This data suggests that also in mouse ESCs major chromatin conformation changes coincide with the onset of detectable *Zeb2* expression. Furthermore, they confirm the importance of REs in the gene desert as also observed for *ZEB2* dynamic expression during NPC differentiation in human iPSCs.

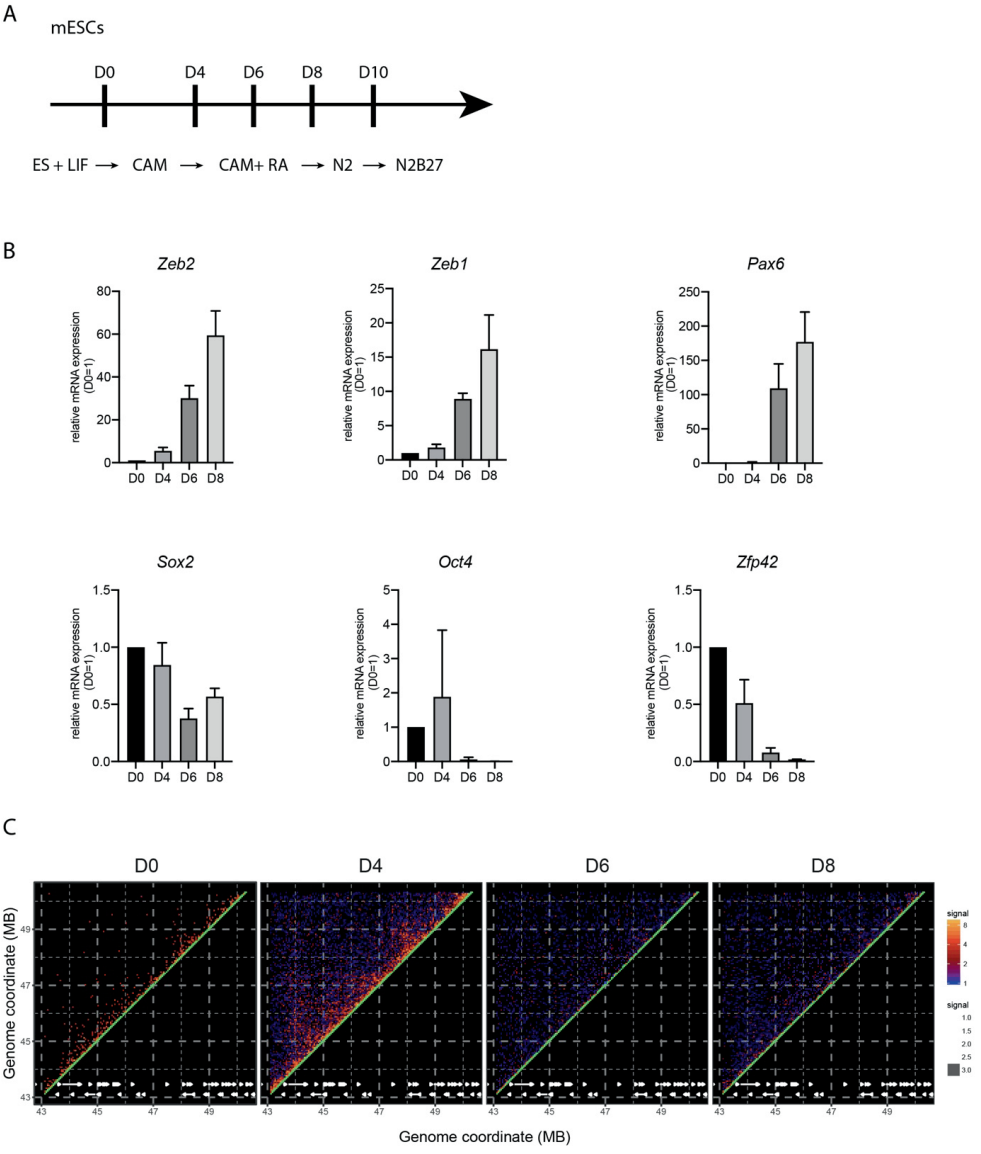


Figure S6.7. ***Zeb2*** **locus dynamics during neural differentiation of mouse ESCs.**

A. Schematic overview of the differentiation protocol, including specific media and inhibitors. *Abbreviations used:* CAM, cellular aggregate medium; RA: retinoic acid; N2: N2-supplemented medium; N2B27: N2 and B27 supplemented medium. **B.** mRNA expression of *Zeb2* and *Zeb1*, NPC marker Pax6 and pluripotency markers *Sox2*, *Oct4*, and also *Zfp42*. **C.** Reconstruction of in-*cis* proximities on the mouse chr2:43-50 (mm10 genome) region, including the *Zeb2* locus, with the intron/exon structured indicated by the bright dashed white lines.
D0: undifferentiated state; D4: epiblast progenitors; D6: early neuro-progenitors; D8: neuroprogenitors.

6.4. Discussion

Highly dynamic chromatin architecture accompanies stem cell differentiation, each change being reflected in state/stage-specific gene-signatures, often of direct relevance to specific lineage commitment and progression. Many genes involved in developmental processes that need to be temporally and/or spatially regulated are located in-*cis* of long gene deserts, such as *Shh*, *HoxD* and *Sox9* (Lettice *et al.*, 2003; Lee *et al.*, 2006; Gordon *et al.*, 2009). These gene deserts contain several REs, either enhancers or silencers, which in a number of cases have been found to regulate the expression of the aforementioned gene(s) in a time, cell-type and/or tissue and location specific manner (Lettice *et al.*, 2003; Montavon *et al.*, 2011; Mead *et al.*, 2013). Here, we have started to add *ZEB2* as another acknowledged and developmental/disease relevant locus located downstream of a 3.3 Mb-long gene desert, with at least in the *ZEB2*-proximal 500 kb the identification of three hitherto unknown enhancers that co-operate in neural differentiation. These results demonstrate for the first time the value of T2C for studies of locus-specific transcription in the context of chromatin conformation and concomitant DNA-looping. T2C has already been used to study global chromatin conformation and interactome at high-resolution (sub-kbp) and high coverage, with low sequencing efforts and at affordable cost (Kolovos *et al.*, 2014; Brant *et al.*, 2016; Knoch *et al.*, 2016; Kolovos *et al.*, 2016). In addition, for *ZEB2* itself, functional studies and dynamics in cell differentiation (including in vertebrate models, but also in MOWS patients) T2C and hence DNA-looping in the *ZEB2* locus can now be added to other types of functional study, including identification of ZEB2 direct target genes (ChIP-sequencing) and co-operation with other partners (ZEB2 interactome; *see* also this PhD thesis) (Stryjewska *et al.*, 2017; Wu *et al.*, 2016; van Grunsven *et al.*, 2007; Verstappen *et al.*, 2008; Conidi *et al.*, 2013).

We have previously shown, both *in vivo* (often in the mouse) and *in vitro*, that detectable *Zeb2* temporal expression directly correlates with cell state and behavior (e.g., differentiation, maturation, migration, epithelial-mesenchymal transition) (Tatari *et al.*, 2014; Goossens *et al.*, 2015; Scott *et al.*, 2016; Scott *et al.*, 2018). In addition, by varying the dosage of *Zeb2,* in (rescued) knockout mice or via Zeb2 transgene-based (over)production in wild-type mice, the concept of precise dosage has become relevant to normal Zeb2 needs or functions, but also developmental defect, but also pathology (van de Putte *et al.*, 2003; Maruhashi *et al.*, 2005; van den Berghe *et al.*, 2013; Goossens *et al.*, 2017; Van de Putte *et al.*, 2017). Hence, studies of mechanisms that regulate *ZEB2* mRNA levels, as well as still needed studies documenting miR-based ZEB2 control and ZEB2 protein (in)stability, become increasingly relevant to the field.

In this study, for the first time for the *ZEB2* locus, we report genomic architecture dynamics and identify three novel enhancers, located about 500 kb upstream of the *ZEB2* TSS, regulating transcription of *ZEB2* during neural differentiation. We initially assess the expression profile of our iPSC line subjected to neural differentiation, and show that *ZEB2* is highly expressed at D6, corresponding to early NPCs. At this state, iPSCs have silenced almost completely their pluripotency gene signature and activated lineage-specific markers. This is in line with the observation by Chng and co-workers, who studied neuroectodermal differentiation of human ESCs, where double inhibition of Activin and BMP signaling results in increase of *ZEB2* mRNA levels up to 6 days of differentiation (Chng *et al.*, 2010). In mouse ESCs, the levels of *Zeb2* mRNA rise at early NPC stage, after which they remain high (Stryjewska *et al.*, 2017). Similar to these ESCs, *ZEB2* expression is still sustained at NPC state of human iPSCs, and their NPC state is amenable to T2C analysis.

For our target region of interest, we designed probes to span roughly 7.4 Mb of chr2, i.e. 143270465-150642631 (hg19 genome reference). These probes cover *KYNU*, *GTDC1*, *ZEB2*, *ACVR2A*, *ORC4*, *MBD5*, *EPC2*, *KIF5C*, *LYPD6B,* and *MMADHC* protein-encoding genes, as well as the two, rather long lncRNA genes *LINC01412* and *TEX41*, located about 2.6 kb and 160 kb upstream of the *ZEB2* TSS, respectively. Our reconstruction of the in-*cis* dynamics of this region of chr2 shows that at iPSC state the majority of the detectable proximities are short-range without a clear TAD structure, indicating a closed conformation in which the *ZEB2* gene is tightly packed and not accessible. As neural differentiation proceeds, the chromatin organization reveals well-defined, distinct TADs, and hence several proximities are mapped, and *ZEB2* is significantly upregulated. The main sub-TAD involves a loop between the *ZEB2* promoter and a segment of ~500 kb upstream of its TSS. Based on histone-3 marks and evolutionary conservation, we identify via T2C three REs, i.e. Enh1, Enh2, and Enh3, which are each active enhancers.

Previous studies, nearly exclusively done using computational analysis of publicly available databases, have shown the existence of different enhancers located upstream, downstream and intergenic/intronic to *Zeb2/ZEB2* (McKinsey *et al.*, 2013; Bar Yacoov *et al.*, 2019). McKinsey *et al.* (2013) and co-workers propose for their mouse models that two *Zeb2* REs are regulated by Dlx1/2, more specifically in GABAergic interneurons of the embryonic ventral forebrain. One of these REs is an enhancer located about 1.4 Mb upstream of the *Zeb2* TSS. Bar Yaacov *et al.* (2019) identified eight new possible brain-specific human enhancers, albeit with intergenic or downstream location with regard to *ZEB2*. In our T2C approach, we also observe, at early NPC state, a major chromatin conformational change in the region containing *ZEB2,* corroborating the finding by Bar-Yaacov and colleagues on the regulation of *ZEB2* by intragenic enhancers.

The test of our three novel DNA-looping segments for enhancer activity after transient transfection at each considered timepoint of differentiation yields that, concomitantly with the major chromatin remodeling observed at D6, the highest levels of the luciferase reporter are observed at the same timepoint/cell state, while a basal, comparable activation is seen at D0 (undifferentiated state) and D15 (late NPCs). Transfection of a series of combinatorial enhancers indicates that the three enhancers work synergistically and that Enh2 and Enh3 alone do not exert the same effect as Enh1 does, but they seem to act together. On the basis of the relative roles for Enh1 and Enh3, we propose that these two enhancers co-operate in a time and/or tissue specific manner, but only when enhancer 2 is present.

Of the many TFs that can potentially bind these novel enhancers, HOXB2 and SOX10 are interesting candidates. While SOX10 has been demonstrated to associate as protein with ZEB2 (Stanchina *et al.*, 2010; Watanabe *et al.*, 2017), nothing is known about a possible interaction between ZEB2 and HOXB2. Knockdown of HOXB2 and SOX10, and assessment of the luciferase reporter activity in NPCs, reveal a role for both TFs as transcriptional repressor of *ZEB2*: lower levels of HOXB2 or SOX10 result in increased expression of the luciferase reporter. HOXB2 is crucial for proper hindbrain formation and regulation of oligodendrogenesis in mice, both processes involving the proper formation of rhombomere 3 (Davenne *et al.*, 1999; Miguez *et al.*, 2012). Hindbrain-specific enhancers of *ZEB* have been identified by Bar-Yaacov *et al.* (2019), suggesting that disruption of these enhancers might affect *ZEB2* expression in the hindbrain and ultimately proper hindbrain organization. In mice, OPC Hoxb2 activates *Olig2* transcription, which is critical as upstream activating TF for *Zeb2* (Weng *et al.*, 2012). Once the levels of Zeb2, as a result of *Zeb2* mRNA upregulation, are sufficiently high, Zeb2 TF acts in a dual mode: it activates genes promoting myelinogenesis in the embryonic CNS, whereas it represses other acknowledged genes

that inhibit it. We add yet another potential mechanism underlying precise *ZEB2* transcriptional control and propose that HOXB2 may be needed for or contribute to repression of the newly identified enhancers and therefore also subsequent precise human *ZEB2* activation and/or upregulation.

In addition to the growing number of control mechanisms for this critical gene in various stages of development, adult homeostasis, and differentiation and/or maturation of many cell types, this will prompt the field to continue to document similarities or differences between human and e.g., mouse models, further assess these controls both *in vivo* and *in vitro*, for example also involving NPCs, and study fine-tuning of *ZEB2* mRNA levels at multiple levels, including locus-specific chromatin conformational changes. Here we find that the major impact of our newly identified enhancers is to have *ZEB2* expression peak at D6 of differentiation, whereas at D15 (late NPCs) its steady-state transcription is again lower. The chromatin conformation change we observed at D6 might be required for a transcriptional boost to increase *ZEB2* mRNA levels, such that they become sufficient for the cells to further proceed with differentiation. At D15, TFs, such as HOXB2, which are produced in a time-specific manner and in our system parallel *ZEB2* expression, then intervene to occupy the enhancers and cause or contribute to negative regulation of *ZEB2*. It is also not excluded that ZEB2 and HOXB2 (and SOX10) co-control each other's loci both by feedback and feedforward regulations, which remains to be investigated in our iPSCs. Further studies are required to identify other TFs acting as negative and/or positive regulators of the neural enhancers we have identified. Interestingly, the expression levels of the other genes located in our *ZEB2* region of interest do not change significantly during neural differentiation, suggesting that of the many genes that flank the gene desert, only *ZEB2* - as the encoded TF is crucial for proper embryonic development - needs precise and dynamic regulation.

Of the whole spectrum of mono-allelic mutations described in MOWS patients, 20% are composed of large gene deletions (including cytogenetically detectable deletions) that very often are described to affect a significant part or the entire *ZEB2*, but sometimes also neighboring, such as downstream located genes (Garavelli *et al.*, 2009). Such patients have very severe phenotypes that encompass various, classical MOWS defects, but could very well have other associated defects due to loss of function of (one or more of) the other genes. Recently, a patient without a mutation in the *ZEB2* protein-coding exon sequences, but with clear MOWS, was identified (Baxter *et al.*, 2017). Sequencing identified a 69 kb-long duplication, located in chr2:145,218,807–145,287,401, containing exons 1 and 2, as well as intron 1 and (part of) intron 2 of *ZEB2* (Aksoy *et al.*, 2017). In a screening for copy number variants in genes related to Hirschsprung Disease, four patients have been described with *ZEB2* duplications of part of exon 1 and all of exon 2, ranging from 1.42 to 1.99 kb (Jiang *et al.*, 2011). Of these patients, three presented mutations in the *RET* gene and, in two of these 3, also a *SOX2* duplication. In line with our results obtained here, these duplications might hamper the formation of the TAD structure we identified here and therefore have an effect on *ZEB2* transcription, including blocking (candidate) TF-binding promoted DNA-looping needed for *ZEB2* upregulation.

We also found a sub-TAD formed between the whole *ZEB2* gene and the 5'-downstream gene *GTDC1* (encoding Glycosyltransferase Like Domain Containing 1). Aksoy *et al.* (2017) have described the sequencing, via DNA paired-end tag, of a patient affected by global developmental delay, language impairments and intellectual disability. In this patient, they found a *de novo* t(2;8) translocation affecting *GTDC1* on chr2, while no annotated gene was involved on chr8. On chr2, the breakpoint is located in intron 5 of *GTDC1*. This translocation might therefore result

in an impaired TAD formation between *GTDC1* and *ZEB2*, which might affect the expression of both genes.

Our work clearly indicates that T2C-based functional studies and the identification of novel REs would be beneficial not only for a better understanding of the connection between RE-containing developmental and disease loci, such as of *ZEB2*, and genome structural organization, but also to clinical geneticists who will systematically intensify gene desert sequencing on top of exon-sequencing in diagnostics in the future, and analyze REs in cell-based functional studies.

6.5. Experimental procedures

Induced pluripotent stem cells (iPSCs)

The WTC iPSC line was obtained from Dr. Bruce Conklin (The J. David Gladstone Institutes, San Francisco, CA, USA). These cells were cultured feeder-cells free on 6-well-plates coated with 1% Geltrex in Essential-8 (E8) Basal Medium (ThermoFisher Scientific), and the medium was changed daily. For their neural induction, a modified version of the protocol by Singec and co-workers was used (Singec *et al.*, 2016). In brief, 70%-confluent cells were changed to Neural Induction Medium (NIM): DMEM/F12, supplemented with 20% knock-out serum replacement, 1 mM non-essential amino acids (NEAA) and 0.1 mM β-EtSH (all from ThermoFisher Scientific), and a mix of inhibitors (abbreviated as DAP), i.e. 2 µM dorsomorphin (BMP-inhibitor, Tocris Bioscience), 2 µM A83-01 (TGFβ-inhibitor, Tocris Bioscience) and 2 µM PNU-74654 (WNT-inhibitor, Sigma). After 6 days (D6) the medium was replaced with Neural Maturation Medium (NMM): DMEM/F12, 1mM NEAA, 0.1 mM β-EtSH, 1xN2 and 1x B27 without VitA (both ThermoFisher Scientific), and the medium was changed every day until D15 (NPC-state). A schematic overview of the culturing conditions and differentiation protocol is depicted in **Fig. 6.1A**.

RNA extraction and RNA-sequencing

RNA was extracted using TRIZOL (Sigma) according to the manufacturer's instruction, and purified using standard extraction and purification by phenol:chloroform and precipitation by isopropanol.

Total RNA for triplicates of three timepoints were checked for quality on an Agilent Technologies 2100 Bioanalyzer using a RNA nano assay. All samples had RIN value greater than 9.10. Triplicate RNA-Seq libraries were prepared according to the Illumina TruSeq stranded mRNA protocol (www.illumina.com). Briefly, 200 ng of total RNA was purified using poly-T oligo-attached magnetic beads to end up with poly-A containing mRNA. The poly-A tailed mRNA was fragmented, and cDNA was synthesized using SuperScript II and random primers in the presence of Actinomycin D. The cDNA fragments were end repaired, purified with AMPure XP beads, A-tailed using Klenow exo-enzyme in the presence of dATP. Paired end adapters with dual index (Illumina) were ligated to the A-tailed cDNA fragments and purified using AMPure XP beads. The resulting adapter-modified cDNA fragments were enriched by PCR using Phusion polymerase as followed: 30 s at 98°C, 15 cycles of (10 s at 98°C, 30 s at 60°C, 30 s at 72°C), 5 min at 72°C. PCR products were purified using AMPure XP beads and eluted in 30 µl of resuspension buffer. One microliter was loaded on an Agilent Technologies 2100 Bioanalyzer using a DNA1000 assay to determine the library concentration and to check the quality. Cluster generation was performed according to the Illumina TruSeq SR Rapid Cluster kit v2 (cBot) Reagents Preparation Guide (www.illumina.com). Briefly, 18 RNA-Seq libraries were pooled together to get a stock of 10 nM. One microliter of the 10 nM stock was denaturated with NaOH, diluted to 6 pM and hybridized onto the flowcell. The hybridized products were sequentially amplified, linearized and end-blocked according to the Illumina Single Read Multiplex Sequencing user guide. After hybridization of the sequencing primer, sequencing-by-synthesis was performed using the HiSeq 2500 with a single read 50-cycle protocol followed by dual index sequencing.

Illumina reads were mapped against the GRCh38 human reference using HiSat2 (Kim *et al.*, 2015). Gene expression values were called using htseq-count (version 0.11.2) and Ensembl release 96 and transcript annotation (Anders *et al.*, 2015). Sample QC and differential expression analysis have been performed in R environment for statistical computing (version 3.6.2, https://www.R-project.org/), using DESeq2 (version 1.20.0;) with the ashr log fold shrinkage methodology (http://bioconductor.org/packages/release/bioc/html/DESeq2.html) and tidyverse (version 1.3.0; https://github.com/tidyverse/tidyverse) (Love *et al.*, 2014; Stephens, 2017; R Core team, 2018).

Indirect immunofluorescence

hiPCs were plated on Geltrex® coated chamber slides. When confluent the cells were differentiated as described before on the slides and harvested on the selected tim points (D0, D6 and D15). The cells were washed three times with PBS (Sigma) and fixed for 15 minutes at room temperature (RT, 24°C) with 4% paraformaldehyde (PFA) and washed again thrice 5 minutes with PBS. The cells were then permeabilized with 100% ice-cold Methanol at -20°C for 10 minutes, washed with PBS thrice for 5 minutes and then blocked for 1 hour at RT in blocking buffer (5% normal goat or donkey serum (Jackson Immunoresearch), 0,3% Triton-X in PBS). Cells were incubated overnight (O/N) at 4°C with the primary antibodies (Table 1) in antibody dilution buffer (1% BSA, 0.3% Triton-X in PBS) in a humidity chamber. The next day the cells were washed three times with PBS for 5 minutes and incubated with the corresponding fluorescent secondary antibody in antibody dilution buffer for 1,5 hours at RT in the dark. After washing the cells three

times 5 minutes with PBS the cells were mounted with Mowiol (Sigma) containing DAPI (1:1000, Sigma-Aldrich) and dried O/N in the dark. Images were acquired with a Leica SP5 confocal microscope.

Targeted Chromatin Capture (T2C)

The T2C protocol was adapted from Kolovos *et al.* (2018). Cells were collected at D0, at D6 and D15 of neural differentiation, using Accutase® (ThermoFisher Scientific) and passed through a 40-µm cell strainer (BD Falcon). $2.5x10^6$ cells were used for each time point; the cells were cross-linked using 1% formaldehyde at Room Temperature (RT, 24°C) for 10 min and quenched with 0.125 M Glycine. Subsequently, cells were lysed using cold lysis buffer containing 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.5% NP-40 and Complete Protease Inhibitors (Roche). Chromatin was digested with the *Apo*I restriction enzyme (New Englad Biolabs) (400 U/sample) overnight in a thermomixer (VWR) at 37°C at 900 rpm. The digested products were purified via Phenol:Chloroform. The diluted DNA fragments were ligated with T4 DNA Ligase High Concentration (100 U/sample; ThermoFisher Scientific) overnight at 16°C, and then 30 µl of 10 mg Proteinase K/ml (300 µg) were added and incubated at 65°C for 4 h, followed by 30 µl of 10 mg RNAse A/ml (300 µg) for 1h at 37°C, before proceeding to further purification with Phenol:Chloroform. In total, 6 µg of the resulting chromatin were then linearized using the frequent 4bp-cutting enzyme *Dpn*II (New England Biolabs) (1U/µg of DNA) overnight at 37°C, whilst shaking in a thermomixer at 400 rpm. The day after, the material was precipitated by sodium acetate/ethanol before proceeding to T2C library preparation.

For each samples a T2C Library was prepared using 250 ng of linearized chromatin. The samples were re-buffered to 10 mM Tris-HCl, pH8 by a standard AMPure XP (Agencourt) bead clean-up procedure. The chromatin was sheared to 300-400-bp fragments by a S220 Covaris (Covaris Inc.). Concentration was determined by Quant-it high sensitivity (ThermoFisher Scientific). For each sample 100 ng of sheared chromatin was end-repaired and A-tailed using the Kapa hyper prep kit (Roche) according to the manufacturer's instructions. SeqCap library adaptors were ligated followed by AMPure bead clean up. The pre-capture library was amplified by PCR using KAPA HiFi hotstart readymix for 9 cycles. The amplified pre-capture library was purified by bead clean up and quantified by Bioanalyzer DNA1000 assay (Agilent) according to the manufacturer's instructions. A *ZEB2* hg19 based design was ordered at NimbleGen (Roche) with baits located between 143270465 and 150642631 of chromosome 2. A hybridization mixture per sample was prepared with 2 µg pre-capture library, 1 mM HE-index-oligo, 1 mM HE universal oligo, COT Human DNA, AMPure XP reagent and added to 4.5 µl of pre-ordered baits and subsequently hybridized for 16 h at 47°C. Post hybridization the samples were washed according to the instructions in the Nimblegen SeqcapEZ Hypercap workflow (Roche), the chromatin captured using capture beads. The captured library was amplified by PCR using Kapa HiFi mix and purified by AMPure XP beads. The captured library was quantified by Nanodrop spectrophotometer and the quality was assessed using a Bioanalyzer DNA1000 assay. Finally, the captured T2C libraries were denatured and sequenced on an Illumina HiSeq2500 sequencer as described for RNA-seq but with a custom recipe of 6 dark cycles, followed by paired end 101 sequencing with single index using the rapid v2 chemistry according to manufacturer's instructions (Illumina). T2C analysis was performed as described in Kolovos *et al.* (2014). In short, reads were aligned to the human GRCh37 reference genome with the BWA aligner and the BWA-backtrack method. Alignments were annotated with the restriction fragments in which they were located. The proximity matrix was then constructed from the mapped primary alignments of with their mapped primary mates. Further analyses and filtering based on the proximity matrix was performed in the R environment for statistical computing.

Cloning and luciferase assay, and shRNA experiments

Clonable DNA-fragments encoding the candidate putative enhancers and the minimal promoter of human *ZEB2* were produced as a single gBlock (IDT), which was then inserted in the luciferase-based pGL4.10 vector (Promega). HEK293T cells, cultured in high glucose (4.5 g/L) DMEM supplemented with 10% FBS, were co-transfected with 1.5 µg of luciferase construct, containing different combinations of enhancers with the ZEB2 minimal promoter, and 50 ng Renilla-based vector, using Lipofectamine-2000 in a 1:1.5 ratio. Empty pGL4.10 was used as negative control. NPCs were transfected with Amaxa Nucleofector II, using the Kit V (Lonza) and transfection program A-33. 4.5 µg of luciferase-construct were transfected with 50 ng of Renilla-encoding vector to 400,000 cells grown in a 12-well plate. After 24 h the cells were lysed in 1x Passive Lysis Buffer (PLB) (Promega). Luciferase and Renilla activity were measured in a Varioskan Lux Microplate reader (ThermoFisher Scientific) using the Dual-Luciferase Reporter Assay System (Promega). Enhancer activity was calculated as the fold-change of Luciferase normalized to Renilla activity. Each transfection was performed three times, and of each transfection three technical replicates were measured.

For the knockdown (KD) experiments shRNAs for *HOXB2*, *SOX10* or *FOXD2* were co-transfected with the luciferase-constructs. **Table 6.2** lists the shRNA sequences used for the KD. Medium was refreshed 24 hours after transfection, and the cells were harvested 48 h after the transfection. To address the KD efficiency RNA was isolated and cDNA was synthesized as described above and expression levels were tested by Real Time

quantitative PCR (RT-qPCR). RT-qPCR was performed using SybrGreen dye (BioRad) on a CFX96 T1000 thermal cycler. All data shown are averages of three independent biological replicates and three technical replicates, normalized to ß-ACTIN. Primers are listed in Table 3. Luciferase activity after KD was performed as described above.

Data availability

T2C data are available under the GEO accession number GSE147000.

Acknowledgments

We thank Dr. Bruce R Conklin, Dr. Po-Lin So and Dr. Veronica Viray (Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA), and Dr. Femke de Vrij (Department of Psychiatry, Erasmus University Medical Center, Rotterdam, NL) for their valuable help with the WTC line. We would like to thank Dr. Ilyas Singec (NIH, National Center for Advancing Translational Sciences (NCATS), Bethesda MD, USA) for the help in optimizing the iPSCs differentiation protocol, the entire Center for Biomics-Genomics at Erasmus University Medical Center for their technical help, and our local colleagues Dr. Kerstin Wendt and Dr. Niels Galjart, as well as Dr. Akis Papantonis (Institute of Pathology, Georg-August University of Göttingen, DE) for constructive discussions. This work was supported by primary funding to the Department of Cell Biology at Erasmus University Medical Center and by the Theme of Biomedical Sciences Internal Gravitation (BIG) Program, supported by Erasmus University Rotterdam (EUR).

Conflict of Interest statement

The authors declare no competing interests.

Tables

Table 6.1. Antibodies used in this study.

Antibody	Commercial source	Cat. No.	Dilution
OCT4	Abcam	Ab19857	1:250
NCAM1/CD56	R&D Systems	AF2408-SP	1:200
SOX2	Immune systems	GT15098	1:400
NESTIN	Biolegend	839801	1:200
PAX6	Biolegend	901301	1:200
TUJ1	Biolegend	801202	1:500
ZEB2	SantaCruz	sc-48789	1:100
Alexa Fluor Donkey α-Rabbit 488	ThermoFisher Scientific	A32790	1:500
Alexa Fluor Donkey α-Goat 594	ThermoFisher Scientific	A32758	1:500
Alexa Fluor Goat α-Mouse 488	ThermoFisher Scientific	A11001	1:500
Donkey α-Rabbit Cy5	Jackson ImmunoResearch	711-175-152	1:500

Table 6.2. shRNA sequences used in knockdown experiments.

shRNA	Target sequence	Oligonucleotide sequence
shRNA HOXB2 #1	CCGCCAAGAAACCCAGCCAAT	CCGG**CCGCCAAGAAACCCAGCCAAT**CTCGAG**ATTGGCTGGGTTTCTTGGCGG**TTTTT
shRNA HOXB2 #2	CGGCCTTTAGCCGTTCGCTTA	CCGG**CGGCCTTTAGCCGTTCGCTTA**CTCGAG**TAAGCGAACGGCTAAAGGCCG**TTTTT
shRNA HOXB2 #3	CTTGGATGAAAGAGAAGAAAT	CCGG**CTTGGATGAAAGAGAAGAAAT**CTCGAG**ATTTCTTCTCTTTCATCCAAG**TTTTT
shRNA SOX10 #1	CCTCATTCTTTGTCTGAGAAA	CCGG**CCTCATTCTTTGTCTGAGAAA**CTCGAG**TTTCTCAGACAAAGAATGAGG**TTTTT
shRNA SOX10 #2	GCAGCCAGTATATACGACACT	CCGG**GCAGCCAGTATATACGACACT**CTCGAG**AGTGTCGTATATACTGGCTGC**TTTTT
shRNA SOX10 #3	GCTGCTGAACGAAAGTGACAA	CCGG**GCTGCTGAACGAAAGTGACAA**CTCGAG**TTGTCACTTTCGTTCAGCAGC**TTTTT
shRNA FOXD2 #1	CTTCTCTATAGACCACATCAT	CCGG**CTTCTCTATAGACCACATCAT**CTCGAG**ATGATGTGGTCTATAGAGAAG**TTTTT

shRNA FOXD2 #2	GCCTTCCTTCTCTATAGACCA	CCGG**GCCTTCCTTCTCTATAGACCA** CTCGAG**TGGTCTATAGAGAAGGA AGGC**TTTTT
shRNA FOXD2 #3	CGAGGCAGACTTAGCCGAGGA	CCGG**CGAGGCAGACTTAGCCGAG GA**CTCGAG**TCCTCGGCTAAGTCTG CCTCG**TTTTT
scrambled control	CAACAAGATGAAGAGCACCAA	CCGG**CAACAAGATGAAGAGCACC AA**CTCGAG**TTGGTGCTCTTCATCTT GTTG**TTTTT

Table 6.3: Primer sequences used for RT-qPCR analyses.

Primer	Oligonucleotide sequence
ACTIN_Fwd	TCCCTGGAGAAGAGCTACGA
ACTIN_Rev	AGCACTGTGTTGGCGTACAG
HOXB2_Fwd	GAATTTGAGAGGGAGATTGGGT
HOXB2_Rev	GGGAAGGTTTGCTCGAAAGG
SOX10_Fwd	ACAAGAAAGACCACCCGGAC
SOX10_Rev	AAGTGGGCGCTCTTGTAGTG
FOXD2_Fwd	TGCGCCAAAGCCTTCTAC
FOXD2_Rev	TGGCCCATGATGTGGTCTAT

Supplementary files

Supplementary file S6.1. Coordinates of all *Apo*I fragments interactions for the considered T2C region of interest.

Supplementary file S6.2. Coordinates of all proximity interactions as obtained from the 20-kb maps (Fig. 6.3) for the three TADs.

These files are available with the Open Access publication at
https://academic.oup.com/hmg/article/29/15/2535/5867762
in subpage https://academic.oup.com/hmg/article/29/15/2535/5867762#supplementary-data

Literature references

Aksoy I, Utami KH, Winata CL, Hillmer AM, Rouam SL, Briault S, Davila S, Stanton LW, Cacheux V. Personalized genome sequencing coupled with iPSC technology identifies GTDC1 as a gene involved in neurodevelopmental disorders. Hum Mol Genet. 2017 Jan 15;26(2):367-82.

Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen MJAM, Geeven G, van Kranenburg M, Pieterse M, Straver R, Haarhuis JHI, Jalink K, Teunissen H, Renkens IJ, Kloosterman WP, Rowland BD, de Wit E, de Ridder J, de Laat W. Enhancer hubs and loop collisions identified from single-allele topologies. Nat Genet. 2018 Aug;50(8):1151-60.

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166-9.

Baala L, Briault S, Etchevers HC, Laumonnier F, Natiq A, Amiel J, Boddaert N, Picard C, Sbiti A, Asermouh A, Attié-Bitach T, Encha-Razavi F, Munnich A, Sefiani A, Lyonnet S. Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis. Nat Genet. 2007 Apr;39(4):454-6.

Bar Yaacov R, Eshel R, Farhi E, Shemuluvich F, Kaplan T, Birnbaum RY. Functional characterization of the ZEB2 regulatory landscape. Hum Mol Genet. 2019 May 1;28(9):1487-97.

Baxter AL, Vivian JL, Hagelstrom RT, Hossain W, Golden WL, Wassman ER, Vanzo RJ, Butler MG. A Novel Partial Duplication of *ZEB2* and Review of *ZEB2* Involvement in Mowat-Wilson Syndrome. Mol Syndromol. 2017 Jun;8(4):211-18.

Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, Grosveld FG, Marenduzzo D, Papantonis A. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. Mol Syst Biol. 2016 Dec 9;12(12):891.

Chatterjee S, Ahituv N. Gene Regulatory Elements, Major Drivers of Human Disease. Annu Rev Genomics Hum Genet. 2017 Aug 31;18:45-63.

Chng Z, Teo A, Pedersen RA, Vallier L. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. Cell Stem Cell. 2010 Jan 8;6(1):59-70.

Conidi A, Cazzola S, Beets K, Coddens K, Collart C, Cornelis F, Cox L, Joke D, Dobreva MP, Dries R, Esguerra C, Francis A, Ibrahimi A, Kroes R, Lesage F, Maas E, Moya I, Pereira PN, Stappers E, Stryjewska A, van den Berghe V, Vermeire L, Verstappen G, Seuntjens E, Umans L, Zwijsen A, Huylebroeck D. Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFβ/BMP signaling *in vivo*. Cytokine Growth Factor Rev. 2011 Oct-Dec;22(5-6):287-300.

Conidi A, van den Berghe V, Leslie K, Stryjewska A, Xue H, Chen YG, Seuntjens E, Huylebroeck D. Four amino acids within a tandem QxVx repeat in a predicted extended α-helix of the Smad-binding domain of Sip1 are necessary for binding to activated Smad proteins. PLoS One. 2013 Oct 11;8(10):e76733.

Davenne M, Maconochie MK, Neun R, Pattyn A, Chambon P, Krumlauf R, Rijli FM. Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. Neuron. 1999 Apr;22(4):677-91.

de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH. The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell. 2018 Jan 11;172(1-2):289-304.e18.

de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature. 2013 Oct 24;502(7472):499-506.

de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, Heard E, Geijsen N, Poot RA, Chambers I, de Laat W. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature. 2013 Sep 12;501(7466):227-31.

Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002 Feb 15;295(5558):1306-11.

El-Kasti MM, Wells T, Carter DA. A novel long-range enhancer regulates postnatal expression of Zeb2: implications for Mowat-Wilson syndrome phenotypes. Hum Mol Genet. 2012 Dec 15;21(26):5429-42.

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2020 Jan 8;48(D1):D87-D92.

Fufa TD, Baxter LL, Wedel JC, Gildea DE; NISC Comparative Sequencing Program, Loftus SK, Pavan WJ. MEK inhibition remodels the active chromatin landscape and induces SOX10 genomic recruitment in BRAF(V600E) mutant melanoma cells. Epigenetics Chromatin. 2019 Aug 9;12(1):50.

Garavelli L, Donadio A, Zanacca C, Banchini G, Della Giustina E, Bertani G, Albertini G, Del Rossi C, Zweier C, Rauch A, Zollino M, Neri G. Hirschsprung disease, mental retardation, characteristic facial features, and mutation in the gene ZFHX1B (SIP1): confirmation of the Mowat-Wilson syndrome. Am J Med Genet A. 2003 Feb 1;116A(4):385-8.

Garavelli L, Zollino M, Mainardi PC, Gurrieri F, Rivieri F, Soli F, Verri R, Albertini E, Favaron E, Zignani M, Orteschi D, Bianchi P, Faravelli F, Forzano F, Seri M, Wischmeijer A, Turchetti D, Pompilii E, Gnoli M, Cocchi G, Mazzanti L, Bergamaschi R, De Brasi D, Sperandeo MP, Mari F, Uliana V, Mostardini R, Cecconi M, Grasso M, Sassi S, Sebastio G, Renieri A, Silengo M, Bernasconi S, Wakamatsu N, Neri G. Mowat-Wilson syndrome: facial phenotype changing with age: study of 19 Italian patients and review of the literature. Am J Med Genet A. 2009 Mar;149A(3):417-26.

Göndör A, Rougier C, Ohlsson R. High-resolution circular chromosome conformation capture assay. Nat Protoc. 2008;3(2):303-13.

Goossens S, Peirs S, Van Loocke W, Wang J, Takawy M, Matthijssens F, Sonderegger SE, Haigh K, Nguyen T, Vandamme N, Costa M, Carmichael C, Van Nieuwerburgh F, Deforce D, Kleifeld O, Curtis DJ, Berx G, Van Vlierberghe P, Haigh JJ. Oncogenic ZEB2 activation drives sensitivity toward KDM1A inhibition in T-cell acute lymphoblastic leukemia. Blood. 2017 Feb 23;129(8):981-90.

Goossens S, Radaelli E, Blanchet O, Durinck K, Van der Meulen J, Peirs S, Taghon T, Tremblay CS, Costa M, Farhang Ghahremani M, De Medts J, Bartunkova S, Haigh K, Schwab C, Farla N, Pieters T, Matthijssens F, Van Roy N, Best JA, Deswarte K, Bogaert P, Carmichael C, Rickard A, Suryani S, Bracken LS, Alserihi R, Canté-Barrett K, Haenebalcke L, Clappier E, Rondou P, Slowicka K, Huylebroeck D, Goldrath AW, Janzen V, McCormack MP, Lock RB, Curtis DJ, Harrison C, Berx G, Speleman F, Meijerink JP, Soulier J, Van Vlierberghe P, Haigh JJ. ZEB2 drives immature T-cell lymphoblastic leukaemia development via enhanced tumour-initiating potential and IL-7 receptor signaling. Nat Commun. 2015 Jan 7;6:5794.

Gordon CT, Tan TY, Benko S, Fitzpatrick D, Lyonnet S, Farlie PG. Long-range regulation at the SOX9 locus in development and disease. J Med Genet. 2009 Oct;46(10):649-56.

Hegarty SV, Sullivan AM, O'Keeffe GW. Zeb2: A multifunctional regulator of nervous system development. Prog Neurobiol. 2015 Sep;132:81-95.

Helgadottir A, Thorleifsson G, Gretarsdottir S, Stefansson OA, Tragante V, Thorolfsdottir RB, Jonsdottir I, Bjornsson T, Steinthorsdottir V, Verweij N, Nielsen JB, Zhou W, Folkersen L, Martinsson A, Heydarpour M, Prakash S, Oskarsson G, Gudbjartsson T, Geirsson A, Olafsson I, Sigurdsson EL, Almgren P, Melander O, Franco-Cereceda A, Hamsten A, Fritsche L, Lin M, Yang B, Hornsby W, Guo D, Brummett CM, Abecasis G, Mathis M, Milewicz D, Body SC, Eriksson P, Willer CJ, Hveem K, Newton-Cheh C, Smith JG, Danielsen R, Thorgeirsson G, Thorsteinsdottir U, Gudbjartsson DF, Holm H, Stefansson K. Genome-wide analysis yields new loci associating with aortic valve stenosis. Nat Commun. 2018 Mar 7;9(1):987.

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013 Nov 7;155(4):934-47.

Ivanovski I, Djuric O, Caraffi SG, Santodirocco D, Pollazzon M, Rosato S, Cordelli DM, Abdalla E, Accorsi P, Adam MP, Ajmone PF, Badura-Stronka M, Baldo C, Baldi M, Bayat A, Bigoni S, Bonvicini F, Breckpot J, Callewaert B, Cocchi G, Cuturilo G, De Brasi D, Devriendt K, Dinulos MB, Hjortshøj TD, Epifanio R, Faravelli F, Fiumara A, Formisano D, Giordano L, Grasso M, Grønborg S, Iodice A, Iughetti L, Kuburovic V, Kutkowska-Kazmierczak A, Lacombe D, Lo Rizzo C, Luchetti A, Malbora B, Mammi I, Mari F, Montorsi G, Moutton S, Møller RS, Muschke P, Nielsen JEK, Obersztyn E, Pantaleoni C, Pellicciari A, Pisanti MA, Prpic I, Poch-Olive ML, Raviglione F, Renieri A, Ricci E, Rivieri F, Santen GW, Savasta S, Scarano G, Schanze I, Selicorni A, Silengo M, Smigiel R, Spaccini L, Sorge G, Szczaluba K, Tarani L, Tone LG, Toutain A, Trimouille A, Valera ET, Vergano SS, Zanotta N, Zenker M, Conidi A, Zollino M, Rauch A, Zweier C, Garavelli L. Phenotype and genotype of 87 patients with Mowat-Wilson syndrome and recommendations for care. Genet Med. 2018 Sep;20(9):965-75.

Jiang Q, Ho YY, Hao L, Nichols Berrios C, Chakravarti A. Copy number variants in candidate genes are genetic modifiers of Hirschsprung disease. PLoS One. 2011;6(6):e21219.

Katoh M, Katoh M. Integrative genomic analyses of ZEB2: Transcriptional regulation of ZEB2 based on SMADs, ETS1, HIF1alpha, POU/OCT, and NF-kappaB. Int J Oncol. 2009 Jun;34(6):1737-42.

Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357-60.

Knoch TA, Wachsmuth M, Kepper N, Lesnussa M, Abuseiris A, Ali Imam AM, Kolovos P, Zuin J, Kockx CEM, Brouwer RWW, van de Werken HJG, van IJcken WFJ, Wendt KS, Grosveld FG. The detailed 3D multi-loop aggregate/rosette chromatin architecture and functional dynamic organization of the human and mouse genomes. Epigenetics Chromatin. 2016 Dec 24;9:58.

Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, Imam AMA, van de Werken HJG, Wendt KS, Knoch TA, van IJcken WFJ, Grosveld F. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. Nat Protoc. 2018 Mar;13(3):459-77.

Kolovos P, Georgomanolis T, Koeferle A, Larkin JD, Brant L, Nikolicć M, Gusmao EG, Zirkel A, Knoch TA, van Ijcken WF, Cook PR, Costa IG, Grosveld FG, Papantonis A. Binding of nuclear factor κB to noncanonical consensus sites reveals its multimodal role during the early inflammatory response. Genome Res. 2016 Nov;26(11):1478-89.

Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, Wendt KS, van IJcken WF, Grosveld F, Knoch TA. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. Epigenetics Chromatin. 2014 Jun 16;7:10.

Lee AP, Koh EG, Tay A, Brenner S, Venkatesh B. Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters. Proc Natl Acad Sci USA. 2006 May 2;103(18):6994-9.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003 Jul 15;12(14):1725-35.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93.

Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016 Nov 17;167(5):1170-87.

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

Maruhashi M, Van De Putte T, Huylebroeck D, Kondoh H, Higashi Y. Involvement of SIP1 in positioning of somite boundaries in the mouse embryo. Dev Dyn. 2005 Oct;234(2):332-8.

McKinsey GL, Lindtner S, Trzcinski B, Visel A, Pennacchio LA, Huylebroeck D, Higashi Y, Rubenstein JL. Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. Neuron. 2013 Jan 9;77(1):83-98.

Mead TJ, Wang Q, Bhattaram P, Dy P, Afelik S, Jensen J, Lefebvre V. A far-upstream (-70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. Nucleic Acids Res. 2013 Apr;41(8):4459-69.

Miguez A, Ducret S, Di Meglio T, Parras C, Hmidan H, Haton C, Sekizar S, Mannioui A, Vidal M, Kerever A, Nyabi O, Haigh J, Zalc B, Rijli FM, Thomas JL. Opposing roles for Hoxa2 and Hoxb2 in hindbrain oligodendrocyte patterning. J Neurosci. 2012 Nov 28;32(48):17172-85.

Mihalas AB, Hevner RF. Control of Neuronal Development by T-Box Genes in the Brain. Curr Top Dev Biol. 2017;122:279-312.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. A regulatory archipelago controls Hox genes transcription in digits. Cell. 2011 Nov 23;147(5):1132-45.

Murakawa Y, Yoshihara M, Kawaji H, Nishikawa M, Zayed H, Suzuki H, Fantom Consortium, Hayashizaki Y. Enhanced Identification of Transcriptional Enhancers Provides Mechanistic Insights into Diseases. Trends Genet. 2016 Feb;32(2):76-88.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 Apr 11;485(7398):381-5.

Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L. Evolution and functional classification of vertebrate gene deserts. Genome Res. 2005 Jan;15(1):137-45.

Polychronopoulos D, King JWD, Nash AJ, Tan G, Lenhard B. Conserved non-coding elements: developmental gene regulation meets genome organization. Nucleic Acids Res. 2017 Dec 15;45(22):12611-24.

Qiao Y, Shiue CN, Zhu J, Zhuang T, Jonsson P, Wright AP, Zhao C, Dahlman-Wright K. AP-1-mediated chromatin looping regulates ZEB2 transcription: new insights into TNFα-induced epithelial-mesenchymal transition in triple-negative breast cancer. Oncotarget. 2015 Apr 10;6(10):7804-14.

Quintes S, Brinkmann BG, Ebert M, Fröb F, Kungl T, Arlt FA, Tarabykin V, Huylebroeck D, Meijer D, Suter U, Wegner M, Sereda MW, Nave KA. Zeb2 is essential for Schwann cell differentiation, myelination and nerve repair. Nat Neurosci. 2016 Aug;19(8):1050-59.

R Core Team (2018) R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Scott CL, Soen B, Martens L, Skrypek N, Saelens W, Taminau J, Blancke G, Van Isterdael G, Huylebroeck D, Haigh J, Saeys Y, Guilliams M, Lambrecht BN, Berx G. The transcription factor Zeb2 regulates development of conventional and plasmacytoid DCs by repressing Id2. J Exp Med. 2016 May 30;213(6):897-911.

Scott CL, T'Jonck W, Martens L, Todorov H, Sichien D, Soen B, Bonnardel J, De Prijck S, Vandamme N, Cannoodt R, Saelens W, Vanneste B, Toussaint W, De Bleser P, Takahashi N, Vandenabeele P, Henri S, Pridans C, Hume DA, Lambrecht BN, De Baetselier P, Milling SWF, Van Ginderachter JA, Malissen B, Berx G, Beschin A, Saeys Y, Guilliams M. The Transcription Factor ZEB2 Is Required to Maintain the Tissue-Specific Identities of Macrophages. Immunity. 2018 Aug 21;49(2):312-25.e5.

Sexton T, Kurukuti S, Mitchell JA, Umlauf D, Nagano T, Fraser P. Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. Nat Protoc. 2012 Jun 21;7(7):1335-50.

Singec I, Crain AM, Hou J, Tobe BTD, Talantova M, Winquist AA, Doctor KS, Choy J, Huang X, La Monaca E, Horn DM, Wolf DA, Lipton SA, Gutierrez GJ, Brill LM, Snyder EY. Quantitative Analysis of Human Pluripotency and Neural Specification by In-Depth (Phospho)Proteomic Profiling. Stem Cell Reports. 2016 Sep 13;7(3):527-42.

Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protoc. 2013 Mar;8(3):509-24.

Stanchina L, Baral V, Robert F, Pingault V, Lemort N, Pachnis V, Goossens M, Bondurand N. Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development.Dev Biol. 2006 Jul 1;295(1):232-49.

Stanchina L, Van de Putte T, Goossens M, Huylebroeck D, Bondurand N. Genetic interaction between Sox10 and Zfhx1b during enteric nervous system development. Dev Biol. 2010 May 15;341(2):416-28.

Stephens M. False discovery rates: a new deal. Biostatistics. 2017 Apr 1;18(2):275-94.

Stryjewska A, Dries R, Pieters T, Verstappen G, Conidi A, Coddens K, Francis A, Umans L, van IJcken WF, Berx G, van Grunsven LA, Grosveld FG, Goossens S, Haigh JJ, Huylebroeck D. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells. 2017 Mar;35(3):611-25.

Suh H, Consiglio A, Ray J, Sawai T, D'Amour KA, Gage FH. *In vivo* fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. Cell Stem Cell. 2007 Nov;1(5):515-28.

Tatari MN, De Craene B, Soen B, Taminau J, Vermassen P, Goossens S, Haigh K, Cazzola S, Lambert J, Huylebroeck D, Haigh JJ, Berx G. ZEB2-transgene expression in the epidermis compromises the integrity of the epidermal barrier through the repression of different tight junction proteins. Cell Mol Life Sci. 2014 Sep;71(18):3599-609.

Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson Å, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. A subcellular map of the human proteome. Science. 2017 May 26;356(6340):eaal3321.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. Nature. 2012 Sep 6;489(7414):75-82.

Van de Putte T, Francis A, Nelles L, van Grunsven LA, Huylebroeck D. Neural crest-specific removal of Zfhx1b in mouse leads to a wide range of neurocristopathies reminiscent of Mowat-Wilson syndrome. Hum Mol Genet. 2007 Jun 15;16(12):1423-36.

Van de Putte T, Maruhashi M, Francis A, Nelles L, Kondoh H, Huylebroeck D, Higashi Y. Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. Am J Hum Genet. 2003 Feb;72(2):465-70.

van den Berghe V, Stappers E, Vandesande B, Dimidschstein J, Kroes R, Francis A, Conidi A, Lesage F, Dries R, Cazzola S, Berx G, Kessaris N, Vanderhaeghen P, van Ijcken W, Grosveld FG, Goossens S, Haigh JJ, Fishell G, Goffinet A, Aerts S, Huylebroeck D, Seuntjens E. Directed migration of cortical interneurons depends on the cell-autonomous action of Sip1. Neuron. 2013 Jan 9;77(1):70-82.

van Grunsven LA, Taelman V, Michiels C, Verstappen G, Souopgui J, Nichane M, Moens E, Opdecamp K, Vanhomwegen J, Kricha S, Huylebroeck D, Bellefroid EJ. XSip1 neuralizing activity involves the co-repressor CtBP and occurs through BMP dependent and independent mechanisms. Dev Biol. 2007 Jun 1;306(1):34-49.

Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P, Nelles L, Wuytens G, Su MT, Bodmer R, Smith JC, Huylebroeck D. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. J Biol Chem. 1999 Jul 16;274(29):20489-98.

Verstappen G, van Grunsven LA, Michiels C, Van de Putte T, Souopgui J, Van Damme J, Bellefroid E, Vandekerckhove J, Huylebroeck D. Atypical Mowat-Wilson patient confirms the importance of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. Hum Mol Genet. 2008 Apr 15;17(8):1175-83.

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009 Sep10; 461(7261):199-205.

Walton KL, Makanji Y, Harrison CA. New insights into the mechanisms of activin action and inhibition. Mol Cell Endocrinol. 2012 Aug 15;359(1-2):2-12.

Watanabe Y, Stanchina L, Lecerf L, Gacem N, Conidi A, Baral V, Pingault V, Huylebroeck D, Bondurand N. Differentiation of Mouse Enteric Nervous System Progenitor Cells Is Controlled by Endothelin 3 and Requires Regulation of Ednrb by SOX10 and ZEB2. Gastroenterology. 2017 Apr;152(5):1139-50.e4.

Weng Q, Chen Y, Wang H, Xu X, Yang B, He Q, Shou W, Chen Y, Higashi Y, van den Berghe V, Seuntjens E, Kernie SG, Bukshpun P, Sherr EH, Huylebroeck D, Lu QR. Dual-mode modulation of Smad signaling by Smad-interacting protein Sip1 is required for myelination in the central nervous system. Neuron. 2012 Feb 23;73(4):713-28.

Wu LM, Wang J, Conidi A, Zhao C, Wang H, Ford Z, Zhang L, Zweier C, Ayee BG, Maurel P, Zwijsen A, Chan JR, Jankowski MP, Huylebroeck D, Lu QR. Zeb2 recruits HDAC-NuRD to inhibit Notch and controls Schwann cell differentiation and remyelination. Nat Neurosci. 2016 Aug;19(8):1060-72.

Zamboni V, Armentano M, Sarò G, Ciraolo E, Ghigo A, Germena G, Umbach A, Valnegri P, Passafaro M, Carabelli V, Gavello D, Bianchi V, D'Adamo P, de Curtis I, El-Assawi N, Mauro A, Priano L, Ferri N, Hirsch E, Merlo GR. Disruption of ArhGAP15 results in hyperactive Rac1, affects the architecture and function of hippocampal inhibitory neurons and causes cognitive deficits. Sci Rep. 2016 Oct 7;6:34877.

Zweier C, Thiel CT, Dufke A, Crow YJ, Meinecke P, Suri M, Ala-Mello S, Beemer F, Bernasconi S, Bianchi P, Bier A, Devriendt K, Dimitrov B, Firth H, Gallagher RC, Garavelli L, Gillessen-Kaesbach G, Hudgins L, Kääriäinen H, Karstens S, Krantz I, Mannhardt A, Medne L, Mücke J, Kibaek M, Krogh LN, Peippo M, Rittinger O, Schulz S, Schelley SL, Temple IK, Dennis NR, Van der Knaap MS, Wheeler P, Yerushalmi B, Zenker M, Seidel H, Lachmeijer A, Prescott T, Kraus C, Lowry RB, Rauch A. Clinical and mutational spectrum of Mowat-Wilson syndrome. Eur J Med Genet. 2005 Apr-Jun;48(2):97-111.

Chapter 7

Zeb2 DNA-binding sites in neuroprogenitor cells reveal autoregulation and affirm neurodevelopmental defects, including in Mowat-Wilson Syndrome

Judith C. Birkhoff [1], Anne L. Korporaal [1], Rutger W.W. Brouwer [2], **Karol Nowosad** [1,3,4], Claudia Milazzo [1], Lidia Mouritadou [1], Mirjam C.G.N. van den Hout [2], Wilfred F. J. van IJcken [1,2], Danny Huylebroeck [1,5,#], Andrea Conidi [1,#]

[1] Department of Cell Biology, Erasmus University Medical Center, Rotterdam, 3015 CN, the Netherlands
[2] Center for Biomics, Erasmus University Medical Center, Rotterdam, 3015 CN, the Netherlands
[3] Department of Biomedical Sciences, Lab. Molecular Genetics, Medical University of Lublin, Lublin, Poland
[4] The Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland
[5] Department of Development and Regeneration, KU Leuven, Leuven, B-3000, Belgium

[#] *Shared senior authors*

2nd version preprint https://www.biorxiv.org/content/10.1101/2021.07.06.451350v2

7.1. Summary

Perturbation and mechanistic studies have shown that the DNA-binding transcription factor Zeb2 controls cell fate decisions, and differentiation and/or maturation in multiple cell lineages, in embryos and after birth. In cultured embryonic stem cells (ESCs) its strong transcriptional upregulation, resulting in increased protein level, is necessary for the exit from primed pluripotency, and for entering general as well as neural differentiation. We edited mouse ESCs to produce epitope-tagged Zeb2 from one of its two endogenous alleles. Using ChIP-sequencing based on this tag, we mapped 2,432 DNA-binding sites of this Zeb2 in ESC-derived neuroprogenitor cells (NPCs). A hitherto undetected, but major site maps promoter-proximal to *Zeb2* itself, and its homozygous removal demonstrates that *Zeb2* autoregulation is necessary to elicit proper Zeb2-dependent effects in ESC to NPC differentiation. We then cross-referenced all Zeb2 DNA-binding sites with transcriptome data from Zeb2 perturbations in ESCs, ventral forebrain in mouse embryos, and adult neurogenesis from the mouse forebrain V-SVZ, respectively. While the characteristics of these neurodevelopmental systems differ, we still find interesting overlaps. These results contribute to explaining neurodevelopmental disorders caused by *ZEB2* deficiency, including Mowat-Wilson Syndrome.

In subsequent work, we have started to do similar work in ESC-derived mesendodermal cells by treating the mouse ESCs with BMP+Activin. We therefore also include at the end of this chapter as an annex the results on establishing a robust and reproducible protocol for mesendodermal differentiation, applied to not only wild-type, control ESCs, but also the Zeb2-tag ESCs and ESCs from which the *Zeb2* autoregulatory site was deleted.

7.2. Introduction

Zeb2 (also named Sip1/Zfhx1b) and Zeb1 (δEF1/Zfhx1a), the two members of the small family of Zeb transcription factors (TFs) in vertebrates, bind predominantly to two separated CACCT (often also present as CACCTG E2-boxes) and sometimes to CACANNT(G) sequences on DNA via two (between Zeb1 and Zeb2) highly conserved, separated clusters of zinc fingers (Funahashi *et al.*, 1991; Sekido *et al.*, 1994; Remacle *et al.*, 1999; Verschueren *et al.*, 1999). Mutations in *ZEB2* cause Mowat-Wilson Syndrome (MOWS, OMIM#235730; Mowat *et al.*, 2003; Cerruti Mainardi *et al.*, 2004; Ishihara *et al.*, 2004), a rare congenital disease displaying intellectual disability, epilepsy/seizures, typical facial dysmorphism, and often Hirschsprung disease, and multiple other defects in MOWSpatients (Mowat *et al.*, 1998; Cacheux *et al.*, 2001; Wakamatsu *et al.*, 2001). Typically, MOWS patients have a delay in developmental milestones and motoric development, and also have eye and tooth anomalies. Other typical features are specific craniofacial malformation, sensorineural deafness and HSCR, which originate from defects in the ZEB2-positive(+) cells of the embryonic neural crest cell lineage. Mutant *ZEB2* alleles have meanwhile been determined for about 350 patients (Yamada *et al.*, 2001; Wilson *et al.*, 2003; Zweier *et al.*, 2005; Garavelli *et al.*, 2009; Ivanovski *et al.*, 2018). Other reports have described malformations in the central nervous system (CNS) of MOWS patients over broad age range, in which the observed defects locate to the corpus callosum and/or hippocampus, and can be seen by neuroimaging and follow-up of electro-clinical defects, which include focal seizures (Garavelli *et al.*, 2017; Ivanovski *et al.*, 2018; Ricci *et al.*, 2021).

Zeb2's action mechanisms, its partner proteins, and few proven or candidate direct target genes or genes whose normal expression depends on intact Zeb2 levels, have been studied in various cell types derived mainly from *Zeb2* conditional, cell-type specific knockout (KO) mouse models, explaining specific phenotypes caused by Zeb2 perturbation. Zeb2 DNA-binding around candidate direct target genes helped also to explain *Zeb2* loss-of-function phenotypes in ESCs and cells of early and late embryos, and in postnatal and adult mice. These genes are involved in ESC pluripotency (*Nanog, Sox2*), cell differentiation (*Id1*, *Smad7*), and maturation of various cell types, e.g., in embryonic cortical and adult neurogenesis (*Ntf3, Sox6*), as well as epithelial-to-mesenchymal transition (EMT) (*Cdh1*) (Comijn *et al.*, 2001; Seuntjens *et al.*, 2009; Weng *et al.*, 2012; Scott *et al.*, 2016; Wu *et al.*, 2016; Stryjewska *et al.*, 2017; Deryckere *et al.*, 2020; Menuchin-Lasowski *et al.*, 2020). In reverse, subtle mutagenesis of Zeb2 DNA-binding sites in demonstrated target genes has confirmed the importance of Zeb DNA-binding and its activity (e.g., repressive on mesodermal X*Bra* in *Xenopus* early embryos, Lerchner *et al.*, 2000; and repressor on epithelial *Cdh1* in normal and cancer epithelial-to-mesenchymal transition (EMT; Comijn *et al.*, 2001).

Despite its critical function in the precise spatial-temporal regulation of expression of system/process-specific relevant genes during embryogenesis and postnatal development, but recently also adult tissue homeostasis and stem cell-based repair, and acute and chronic disease (van Grunsven *et al.*, 2006; Conidi *et al.*, 2011; for a recent elaborative review, see Birkhoff *et al.*, 2021), ChIP-sequencing (ChIP-seq) data for Zeb2 has been obtained in very few cases only. A major reason is that (at least in our hands) ChIP-seq grade anti-Zeb2 antibodies, stringently specific for Zeb2, are not readily available, and other anti-Zeb antibodies are not exceptionally performing in ChIP-seq, but sometimes do work in ChIP-PCR for individual binding sites. Furthermore, most, if not all, anti-Zeb2 antibodies cross-react with Zeb1, so do not discriminate between both proteins when their presence overlaps or succeeds to one another. However, they compete for the same target genes, which for the individual proteins depends on cell identity/state, extrinsic stimulation or cellular context (e.g., in somitogenesis, Miyoshi *et al.*, 2006; and melanoma, Vandamme *et al.*, 2020). Despite these practical hurdles, data have been published for high-Zeb2 hepatocellular carcinoma and leukemia cell lines, or cultured cells that overproduce tag-Zeb2 from episomal vectors or the safe *Rosa26* locus (Stryjewska *et al.*, 2017; Balcik-Ercin *et al.*, 2018; Yang *et al.*, 2018), but neither of those represents endogenous Zeb2 levels nor expresses *Zeb2* with normal dynamics.

During neural differentiation (ND) of mouse (m) ESCs, Zeb2 mRNA/protein is undetectable in undifferentiated cells, whereas its strong upregulation accompanies efficient conversion of naïve ESCs into epiblast stem cell like cells (EpiLSCs) and is essential for subsequent exit from primed cells and onset of differentiation, including progression to neuroprogenitor cells (NPCs) (Stryjewska *et al.*, 2017). We have edited one *Zeb2* allele of mESCs by inserting a Flag-V5 epitope tag (in brief V5, because anti-V5 antibody will be used for ChIP-seq) just before the stop codon, in-frame with the last exon (ex9 of mouse *Zeb2*, Nelles *et al.*, 2003). Such Zeb2-V5 mESCs were then differentiated to NPCs, and Zeb2 binding sites were determined by V5-tag ChIP-seq. Doing so, we identified 2,432 binding sites, 2,294 of which map to 1,952 protein-encoding genes.

We then cross-referenced the ChIP+ protein-encoding target genes with RNA-seq data of differentially expressed genes (DEGs) in cell-type specific neurodevelopment-relevant Zeb2 perturbations (van den Berghe *et al.*, 2013; Deryckere *et al.*, 2020; Dries *et al.*, 2020). Although we compare non-identical systems, this overall approach still revealed a number of interesting overlaps, as well as Zeb2's role in regulating critical targets in neurodevelopment. Taken together,

and to our knowledge, we are the first to identify endogenous genome-wide binding sites (GWBS) in ESC-derived neural cells for the MOWS TF ZEB2.

7.3. Results

7.3.1. Heterozygous Zeb2-Flag-V5 ESCs differentiate as wild-type cells

Addition of short epitope(s) at the Zeb2 N- or C-terminus, as well as entire activation/repression domains of heterologous TFs (such as VP16 and En2, respectively) at the Zeb2 C-terminus, does not interfere with Zeb2's DNA-binding (as tested in *Xenopus* embryos, Papin *et al.*, 2002; heterologous cells, Verstappen *et al.*, 2008; mouse forebrain, van den Berghe *et al.*, 2013; and mESCs, Stryjewska *et al.*, 2017). Here, we have now used a CRISPR/Cas9 approach (*see* Experimental Procedures) to insert an in-frame Flag-V5-tag encoding sequence in *Zeb2*-ex9 of mESCs (yielding clone 2BE3; **Fig. S7.1**). Allele-specific RT-qPCR using primers that amplify sequences between ex9 and the V5-tag showed mRNA expression from the tagged allele during ND in cell culture at control day (D) 0 and days 4, 6 and 8 (**Fig. 7.1A**). Western blot analysis in nuclear extracts of ND-ESCs at D8, thus in NPCs (Stryjewska *et al.*, 2017), confirmed the presence of Zeb2 of expected relative molecular mass, using either anti-V5 (αV5) or western-blot quality anti-Zeb2 antibodies (**Fig. 7.1B**). Both the Zeb2-V5 and wild-type, control ESCs were then also verified during ND differentiation for temporal expression of *Zeb2*, core pluripotency genes (*Pou5f1*, *Nanog*, both downregulated upon ND, and *Sox2*, also a NPC TF) and an acknowledged NPC marker (*Pax6)* (**Fig. 7.1C,D**), all used as such in our previous work.

Both cell lines displayed comparable *Zeb2* expression dynamics of the respective genes/alleles, indicating that Zeb2-V5 NPCs at D8 of ND can be used for ChIP-seq. Further confirmation came from selective pull-down of Zeb2 on the known target *Cdh1*, using ChIP-qPCR. Zeb2 binds to two (of three) E-boxes in the mouse *Cdh1* promoter (**Fig. 7.1E**), which it represses during EMT (*see* section 7.2; Comijn *et al.*, 2001). A ±25-fold enrichment for Zeb2-V5 was obtained when probing this *Cdh1* region using anti-V5 antibody (αV5) conjugated beads, compared to agarose beads (the negative control) (**Fig. 7.1F**). Hence, the Zeb2-V5 protein binds to a well-characterized, known Zeb2 target site, and the NPCs are suitable for endogenous mapping of the Zeb2 GWBS.

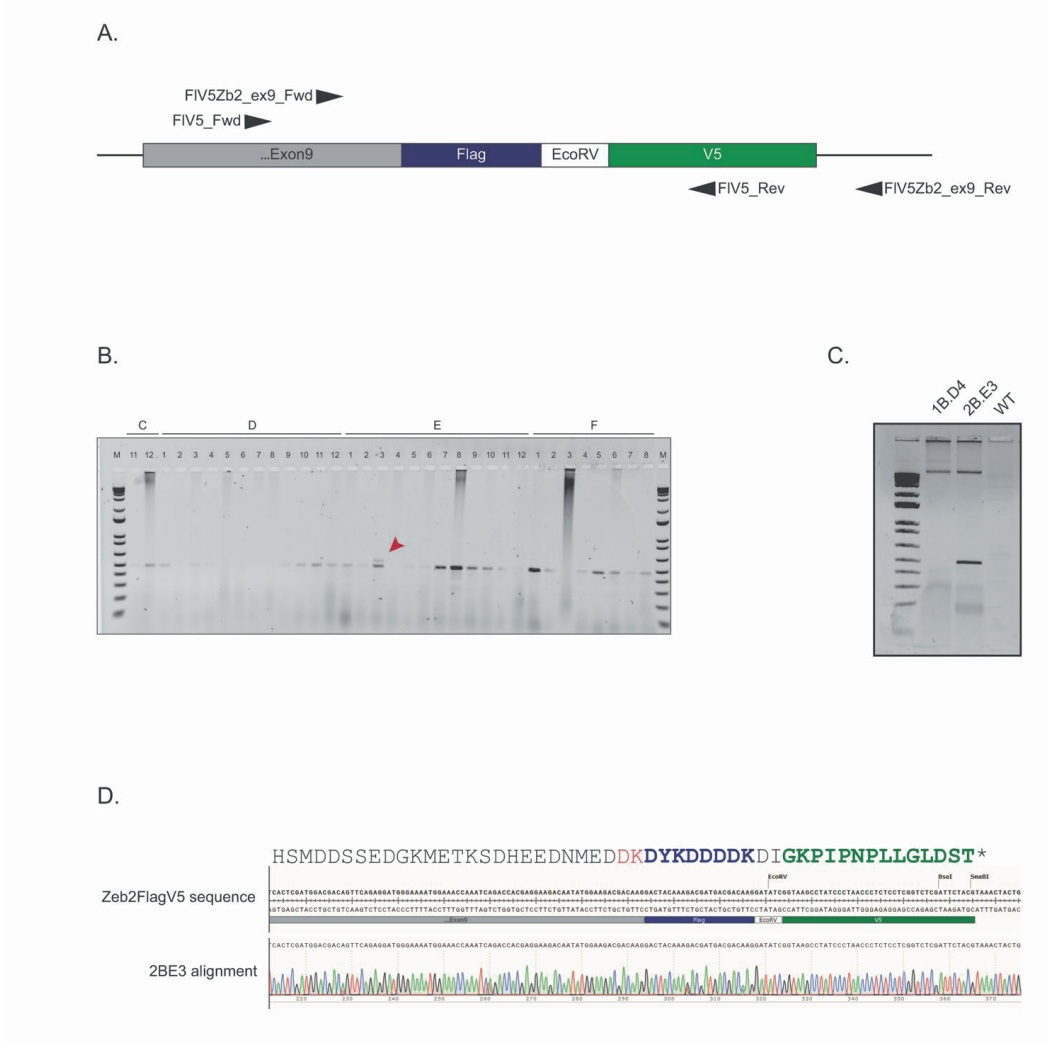


Figure S7.1. Characterization of the Zeb2-V5 mESC lines.

A. Schematic overview of the design strategy, including showing also primers used for genotyping and detection of tagged allele. **B.** Genotyping results (selected part) showing the heterozygous band (red arrow) present in mouse (m) ESC clone 2BE3. Lower band represents the wild-type (WT) allele. **C.** Zeb2-V5 specific PCR showing the presence of the tagged allele only in clone 2BE3, and not in WT genomic DNA material. Clone 1BD4 was also used as negative control. **D.** Sanger DNA-sequencing results of clone 2BE3 showing the alignment with the Zeb2V5 sequence designed *in silico*. The last three amino acids (i.e. GME) of WT Zeb2 have been modified in DK to remove a PAM (protospacer adjacent motif) sequence and mutate the remaining PAM sequence, the target of the gRNA, to avoid multiple cutting by the Cas9. An *Eco*RI site was inserted between the Flag- and the V5-coding sequences to facilitate the screening, and an artificial STOP codon was added downstream of the V5-coding sequence.

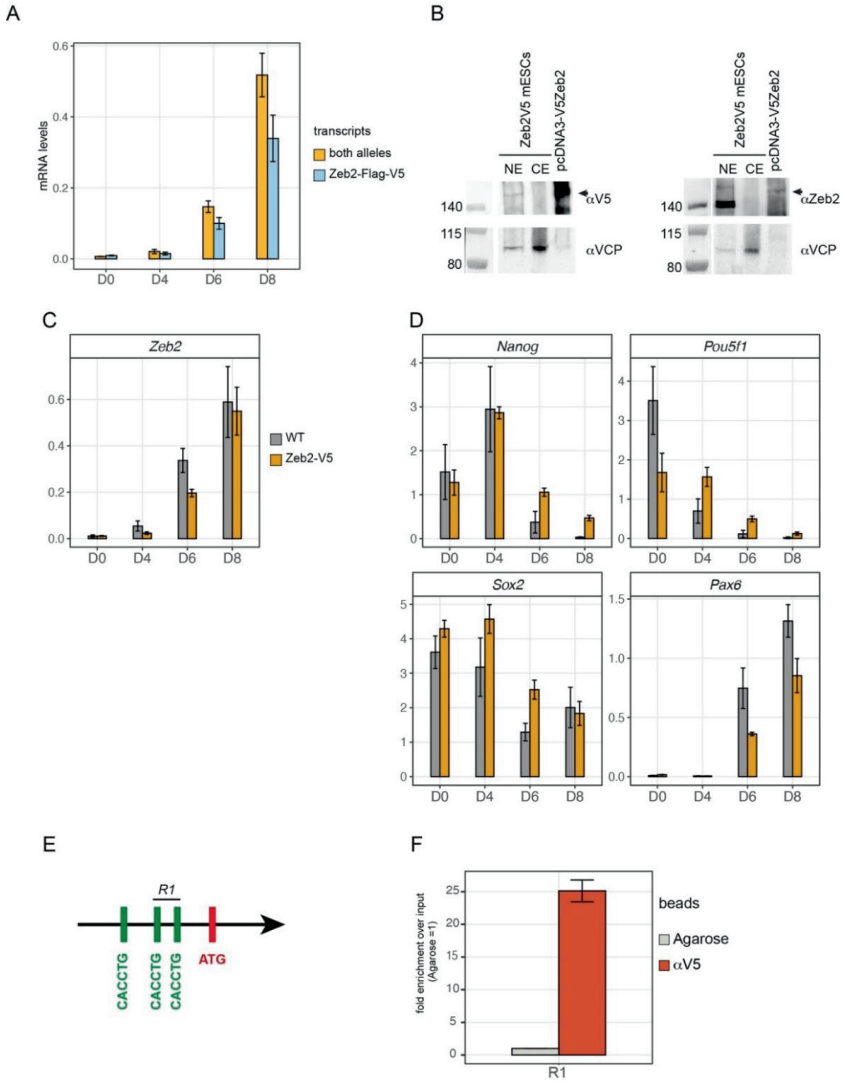


Figure 7.1. Characterization of the heterozygous Zeb2-Flag-V5 mESC line and ChIP-qPCR validation.

A. Allele-specific RT-qPCR using two sets of primers located either in ex7 (and therefore able to detect the whole Zeb2 mRNA produced by both alleles; orange bar) or located in ex9 and V5-tag, respectively (thus recognizing specifically the knocked-in tagged allele; light blue bar). **B.** Western blot analysis showing V5-epitope containing Zeb2 in ESC-derived NPCs (at D8 of neural differentiation, ND) in nuclear extracts (NE), but not in cytoplasmic extracts (CE). Membranes were blotted with anti-V5 antibody (left panel, αV5) or western-quality anti-Zeb2 antibody (right panel, αZeb2; Seuntjens *et al.*, 2009). As control, a fraction of Zeb2-rich extract obtained from HeLa cells transfected with pcDNA3-V5Zeb2 vector was also included. **C.** Zeb2 mRNA levels in wild-type (WT, gray bar) and Zeb2-V5 (orange bar, clone 2BE3, indicated as Zeb2-V5) mESCs during ND. **D.** Pluripotency markers *Nanog*, *Pou5f1* (*Oct4*) and *Sox2* are down regulated in Zeb2-V5 mESCs, similarly to WT. The neuronal marker *Pax6* is also upregulated during differentiation, like in WT mESCs. **E.** Scheme of the mouse *Cdh1* promoter showing the three E-boxes located upstream the ATG start codon (Comijn *et al.*, 2001). Zeb2 binds specifically to only two of these, indicated as R1 (Stryjewska *et al.*, 2017). **F.** ChIP-qPCR showing enrichment for Zeb2-V5 binding to the R1 region of the *Cdh1* promoter. Agarose beads were used as negative control (in grey).

7.3.2. One-third of 2,432 Zeb2 DNA-binding sites map close to the transcription start site of transcriptome-confirmed, system-relevant protein-encoding genes, including *Zeb2* itself

αV5-precipitated samples from upscaled Zeb2-V5 NPCs were used for cross-link ChIP-seq, followed by analysis with Galaxy Software (*see* Experimental Procedures; Afgan *et al.*, 2018). Of the total of 2,432 significant peaks, 2,294 peaks (94% of total) mapped to 1,952 loci that encode protein, while 125 peaks (5% of total) map to micro-RNA (miRNA) genes, and 1% to regions lacking annotation (indicated NA, and using ENSEMBL-GRCm38.99; **Fig. 7.2A**; **File S7.1**). In addition, about 37.5% of all binding sites of Zeb2-V5 are located within -10/+10 kb of annotated transcription start sites (TSS, **Fig. 7.2B**). Gene ontology (GO) pathway enrichment analysis of the aforementioned 1,952 loci revealed binding of Zeb2 to classes of genes annotated to Wnt, integrin, chemokine/cytokine (predominantly as defined in inflammation) and cadherin signaling, respectively, as well as to developmental signaling by EGF, VEGF, TGFβ and FGF family pathways (**Fig. 7.2C**). Among these 1,952 loci, those for genes encoding transcription regulatory proteins, and post-translational modification as well as metabolic enzymes, are well-represented (**Fig. 7.2D**).

In parallel we applied bulk temporal RNA-seq of wild-type mESCs at D0 (undifferentiated), D4 (neural induction), D6 (early NPCs) and D8 (NPCs) and checked the expression dynamics of the 1,952 Zeb2-bound genes from the D8 sample. Among these, 1,244 were subject to significant transcriptional regulation between D4-8 as compared to D0 (**Fig. 7.2E**; log2FoldChange <-0.5 or >0.5 and p-value < 0.05; low-stringency analysis was opted to assess also small differences in mRNA of Zeb2-bound genes). Further, 335 of these genes are commonly expressed between days 4-6-8, but at different levels, including *Zeb2* itself (for the lists of all DEGs, *see* **File S7.2**). **Fig. S7.2A** depicts the D4, D6 and D8 transcriptomes of ND-mESCs, each compared to D0, with indication whether the genes are bound or not by Zeb2, as determined by Zeb2-V5 ChIP-seq. At each of these respective time points, hence at different Zeb2 mRNA level, about 11-14% of the up/downregulated DEGs are bound by Zeb2 (**Fig. S7.2B**).

Among the Zeb2-bound genes that normally become downregulated, *Dnmt3l* and *Esrrb* are present, suggesting that upregulation of *Zeb2* in ND-ESCs (D6 and 8) directly causes downregulation of these two genes accordingly (**Fig. S7.2C**; **File S7.2**). Zeb2 has been suggested to directly repress *Dnmt3l* and *Esrrb*, facilitating the switch from self-renewal of ESCs to their exit from pluripotency and promotion of differentiation. Indeed, expression levels of all *Dnmt3* genes remained higher in *Zeb2*-knockout (KO) ND-ESCs (Stryjewska *et al.*, 2017). However, these *Zeb2*-KO cells convert very inefficiently into EpiLSCs and in any case fail to exit from primed pluripotency.

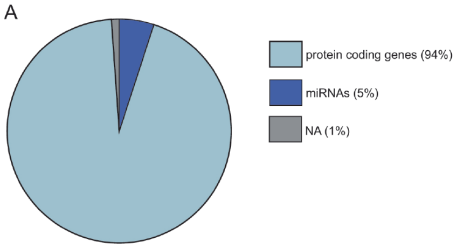
Importantly, among the Zeb2-binding genes whose mRNA levels increased during ND, *Zeb2* itself is also present (yellow dot, **Fig. S7.2C**), indicating autoregulation. In fact, in this ND model the highest recruitment of Zeb2-V5 in ChIP-seq data was mapped upstream the TSS of *Zeb2* (**File S7.1**).

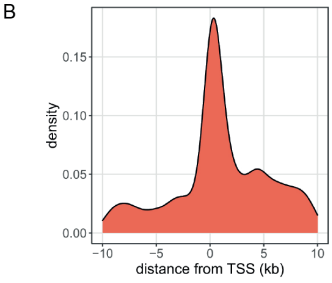
Out of the 1,244 Zeb2-bound genes that significantly changed steady-state mRNA levels in our D4 to D8 transcriptome data sets, 213 are exclusive DEGs in NPCs at D8 (**Fig. 7.2E**; **Fig. S7.2D, File S7.2**). Among these, and as interesting example for many reasons (*see* below; also disucssed in Meert *et al.*, 2022), *Tcf4* is bound by Zeb2 and becomes highly upregulated in NPCs (**Fig. 7.2E**; **Fig. S7.2D, File S7.2**). Tcf4 is a ubiquitous basic helix-loop-helix (bHLH) type TF that

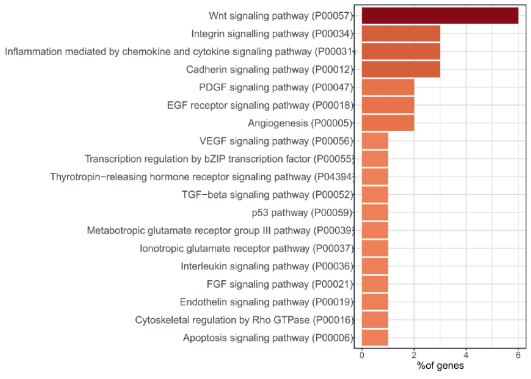
binds to E-boxes, and co-operates in many isoforms (Sepp *et al.*, 2012; Teixeira *et al.*, 2021) with cell-type specific bHLH TF monomers in heterodimers, which are active during CNS development (Corneliussen *et al.*, 1991; Forrest *et al.*, 2014; for a review, see Wang and Baker, 2015). In oligodendrocyte precursors (OPCs), Tcf4 is essential for their subsequent differentiation. Tcf4 dimerizes here with the lineage-specific bHLH-TF Olig2, further promoting differentiation and maturation (Fu *et al.*, 2009), while Zeb2 together with upstream Olig1/2 are essential for myelinogenesis in the embryonic CNS (Weng *et al.*, 2012). Here, Zeb2 generates anti-BMP(-Smad)/anti-Wnt(-β-catenin) activities, which are crucial for progression of embryonic CNS myelinogenesis emanating from differentiation of OPCs. The regulatory action of Zeb2 on the *Tcf4* target gene, as found in mouse cells by our ChIP-seq, may underpin phenotypic similarities between MOWS and Pitt-Hopkins syndrome patients (PTHS, OMIM #610954) (for a detailed discussion, see Meert *et al.*, 2022), the latter caused by mutations in *TCF4* (Brockschmidt *et al.*, 2007), making us speculate that *TCF4* may be deregulated in MOWS neural cells.

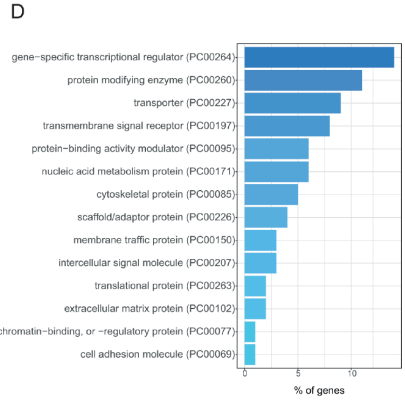
Figure 7.2. (*Next page*). Zeb2-V5 protein is recruited at the TSS of transcriptional regulator encoding genes, predominantly those classified in Wnt signaling.

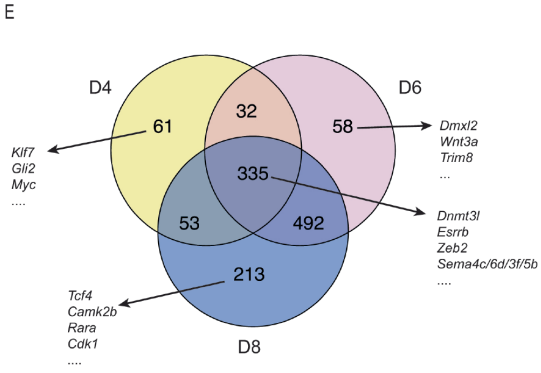
A. 2,432 peaks were selected from our ChIP-seq data set (*see* Experimental procedures). Of these, 94% are associated with protein coding loci, 5% with miRNAs and the remaining 1% map to regions without functional annotation (NA). **B.** Frequency plot showing the binding of Zeb2-V5 at and around (-10 to +10 kb) the TSS. **C.** The 2,294 peaks map to 1,952 protein encoding genes, many of which operate in Wnt signaling (**File S4.1**) or are (as shown in panel **D**) transcriptional regulators. **E.** Of the 1,952 protein-encoding genes, 1,244 are differentially expressed during ND, when compared to the undifferentiated state (D0). Of these 1,244 genes, 335 are differentially expressed at all three time points of ND. A few examples are listed of DEGs uniquely expressed at one time point, as well as these that are shared among three time points; a full list is provided in **File S4.2**. **F.** Overlap of the Zeb2-bound regions with H3K27ac, H3K4me1 and H3K4me3 histone marks in the -10/+10 kb from the TSS of up or downregulated genes at D8 of mESCs differentiation. **G.** Overlap of the Zeb2-bound regions outside the -10/+10kb region from the TSS with histone marks.



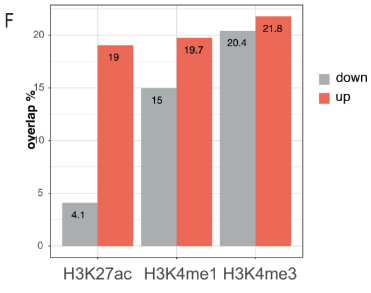


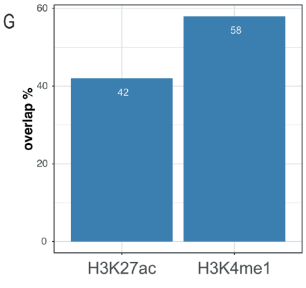






	up	down
D4	297	184
D6	523	394
D8	612	481





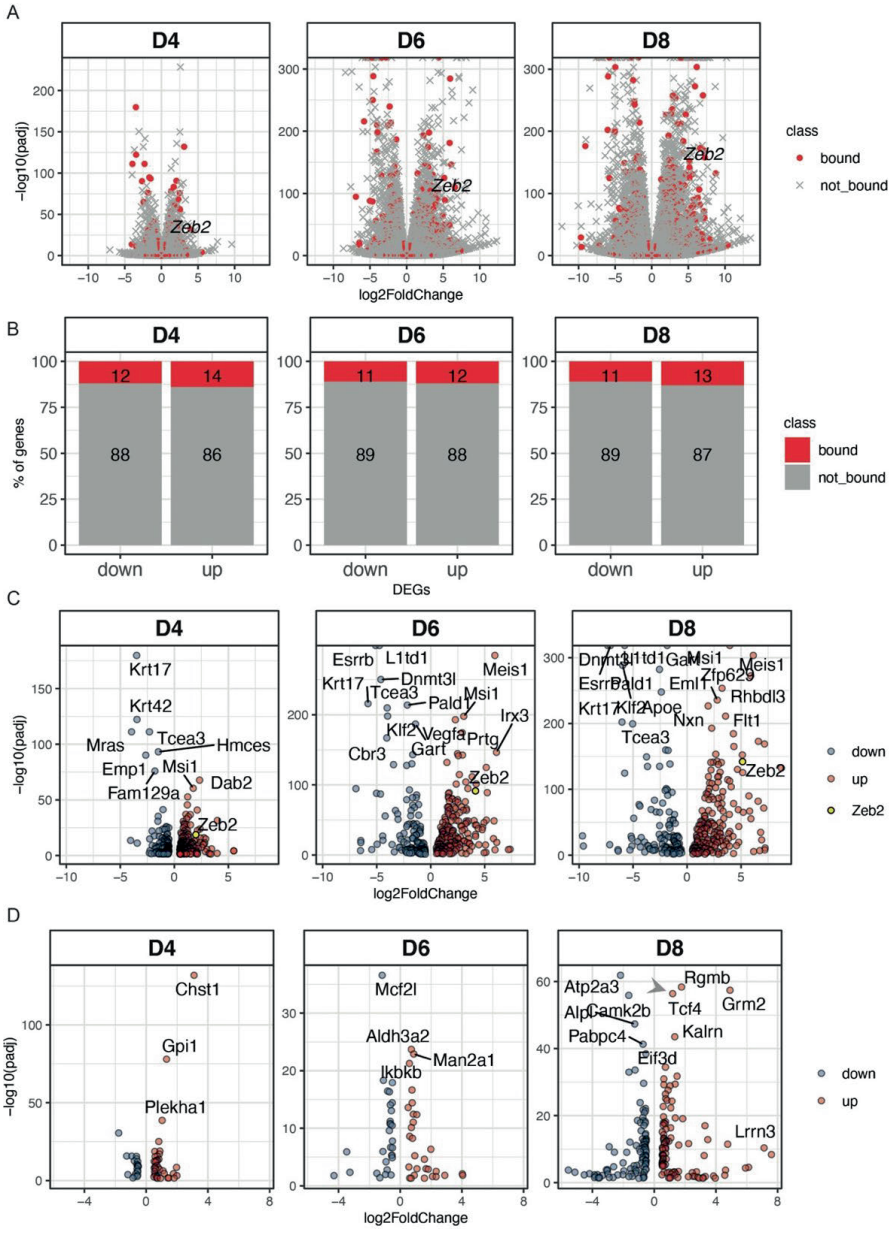


Figure S7.2. Cross-reference of Zeb2-V5-bound protein-coding genes and transcriptome of differentiating mESCs.

A. Volcano plots showing the distribution of Zeb2 bound genes in the whole transcriptomes of D4, D6 and D8 neural differentiated mESCs. Red dots depict the genes bound by Zeb2 and grey crosses those not bound. **B.** About 11-14% of the up or downregulated genes are bound by Zeb2 (red bars) at D4-D8 **C.** Of the 1,952 genes bound by Zeb2 and found being differentially expressed during mESC differentiation, 335 are in common among the three considered time points. *Zeb2* itself is among these common genes and its expression increases during differentiation (yellow dot). The volcano plots show the distribution of the common genes during differentiation. **D.** Timepoint specific DEGs depicted as volcano plots. The grey arrow in D8 panel indicates *Tcf4*.

7.3.3. Zeb2-peaks overlap with active enhancers and promoters within -10/+10 kb from the TSS

To assess whether Zeb2-peaks are present in the regulatory regions of up or downregulated genes from our transcriptome data, we cross-referenced the coordinates of the Zeb2 broad peaks within -10/+10kb from the TSS, with the mouse ChIP-seq datasets available in ENCODE for central nervous system parts (cortical plate, olfactory bulb, forebrain, midbrain, hindbrain and cerebellum, neural tube). We found that of the different histones ChIP-seq datasets available in ENCODE H3K27ac, H3K4me1 and H3K4me3 marks overlapped with our ChIP-seq data (**Fig. 7.2F**). In particular the H3K27ac signature strongly overlapped with the Zeb2-peaks in genes upregulated at D8 (19% in upregulated genes vs. 4% in downregulated genes). For H3K4me1 and H4K4me3 marks, no big difference in overlap between up and down-regulated genes was observed. While the H3K27ac mark is associated with active enhancers, H3K4me1 is associated with primed enhancers, and H3K4me3 is considered a "promoter" marker (Zhang *et al.*, 2020). Taken together these data suggest an activating role for Zeb2 at this stage of differentiation. Outside the -10/+10kb considered range, about 48% of the identified peaks overlap for about 42% with H3K27ac and 58% with H3K4me1 histone marks (**Fig. 7.2G**).

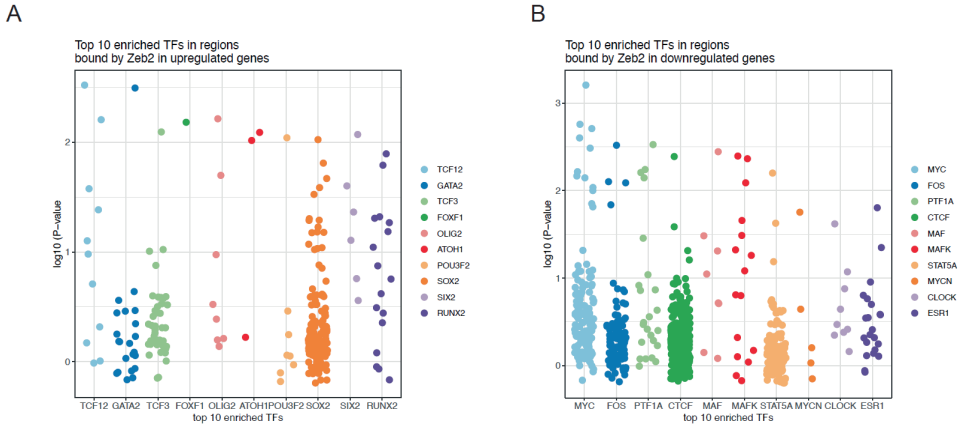


Figure S7.3. Top-10 TF motifs enrichment at peak present in the -10/+10 kb segment from the TSS of up or downregulated genes.

Scatter plots represent the top-10 TF motifs found in the peaks present in the -10/+10 kb from the TSS range in genes which are up- or down-regulated during mESC neural differentiation (panel **A** and **B**, respectively). Each dot represents a peak.

We then performed motif enrichment analysis using UniBind (https://unibind.uio.no) for candidate TFs that would bind the Zeb2-bound peaks and in proximity of up or downregulated genes at D8 of mESCs differentiation (**Fig. S7.3A,B**). In those peaks close to the TSS of upregulated genes, motifs for Sox2, Gata2 and Tcf3 were very abundant. These TFs are known to function during NPC or neural differentiation in general. Sox2 is also an acknowledged marker for neurogenesis (Suh *et al.*, 2007) and also acts during later stages of neurodifferentiation. Zeb2 exhibits anti-Sox2 activities in CNS myelinogenesis and in (re-)myelination by adult Schwann cells in the PNS, and is needed for normal progression of differentiation and maturation in this glial cell lineage (Conidi *et al.*, 2011; Weng *et al.*, 2012; Quintes *et al.*, 2016; Wu *et al.*, 2016).

Tcf3 (also named E2A) plays a role in stem cell self-renewal (Dunn *et al.*, 2014), but is also important during neural fate commitment and possibly represses Nodal signaling during ND (Rao *et al.*, 2020). Gata2 is associated with negatively regulating proliferation in NPCs and, by doing so, directs these cells further into differentiation (El Wakil *et al.*, 2006). However, how Zeb2 acts upon, or together with, these TFs during NPC differentiation is not fully worked-out yet.

In the peaks close to the TSS of downregulated genes there is a prevalence for CTCF, Fos, Myc and Stat5a binding. CTCF is likely of specific interest for its actions that link 3D genome architecture with gene expression regulation. During NPC differentiation however, it was observed that 40% of the NPC-specific DNA-loops were not CTCF-dependent, whereas in other cell-state specific loops, this was only 10%, indicating a less important role for CTCF in the regulation of NPC differentiation compared to differentiation into other cell lineages (Beagan *et al.*, 2017). Also here, more studies are required to get more insights into the co-operativity or counteracting actions of Zeb2 with candidate TFs in the regulation of the candidate Zeb2 targets.

7.3.4. Meta-analysis of RNA-seq data from selected neural-system *Zeb2* perturbations and the NPC ChIP-seq data reveal interesting overlaps

We performed a meta-analysis of three published transcriptome data sets from controls and *Zeb2*-KOs: sorted E14.5 mouse ventral forebrain interneurons (Nkx2.1-Cre driven *Zeb2*-KO; van den Berghe *et al.*, 2013) and sorted (at P2) progenitors of the V-SVZ, an adult neurogenic niche (Gsh2-Cre, *see* van den Berghe *et al.*, 2013). In addition, we used high-throughput RT-qPCR data generated on a Fluidigm platform and obtained after esiRNA-based knockdown (KD) of Zeb2, as part of a systems-biology study in ND-mESCs (the Zeb2 KD data subset was kindly provided by R. Dries; Dries *et al.*, 2020). From these respective datasets, the DEGs upon the Zeb2 perturbations (p.value <0.05; log2FoldChange<-1 and >1) were filtered. This identified (i) genes that depend on normal Zeb2 levels for their downregulation/repression (if directly by Zeb2, then as repressor) or (ii) other genes that depend on Zeb2 for their upregulation/activation (if directly by Zeb2, as activator) (for Zeb2 as dual TF, *see* Conidi *et al.*, 2011; Hegarty *et al.*, 2015; Birkhoff *et al.*, 2021). In parallel, our 2,294 Zeb2-V5 sites in the 1,952 protein-encoding genes were filtered from the complete ChIP-seq dataset, and then used as reference for the RNA datasets (**Fig. 7.3A**).

This cross-referencing enabled the identification of 108 protein-encoding genes among the three transcriptomic datasets and the ChIP-seq dataset (**Fig. 7.3A**). **Fig. S7.4** shows a heatmap of the changes in mRNA levels of these 108 genes during ND of wild-type ESCs, as well as their correlation with the analyzed datasets. Noteworthy, only *Cxcr4* was common to all RNA data sets. This is definitely due to the fact that two RNA-seq sets are generated in different brain/neuron cell-type *in vivo* mouse models, while the other steady-state RNA level data documented the effects of Zeb2-KD on mRNA levels of (only 96 in total) TGFβ/BMP-system components (Dries *et al.*, 2020), so the timing does not completely overlap with our ChIP-seq dataset obtained here. However, Cxcr4 and its ligand Cxcl12/Sdf-1 are crucial for migration of interneurons from the ventral forebrain to the neocortex (Stumm *et al.*, 2003; Nash and Meucci, 2014), processes co-controlled by Zeb2 (van den Berghe *et al.*, 2013). Furthermore, the identified 108 genes are involved in regulation of stem cell pluripotency, signaling by TGFβ, FoxO and Hippo, and also operate in axon guidance. Taken together, these data actually confirm the pivotal role of Zeb2 in these processes further.

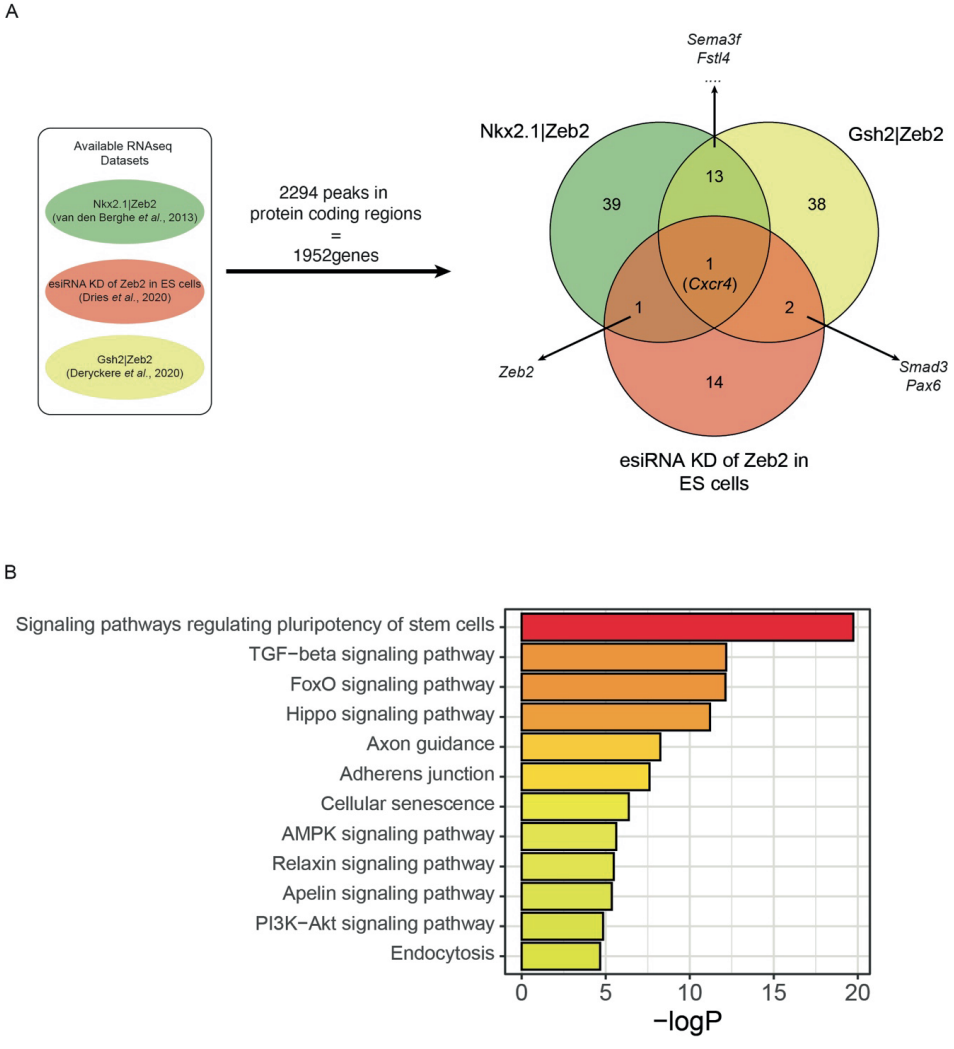


Figure 7.3. Schematic representation of the meta-analysis of Zeb2-bound genes versus RNA-seq datasets.

A. 108 genes bound by Zeb2 are also differentially expressed in the three datasets from other studies in mouse models and ESCs (*see* main text for details). *Cxcr4* is the only DEG bound by Zeb2 and common among the three datasets (for discussion, *see* main text). **B.** These 108 genes mainly map to signaling pathways regulating stem cell pluripotency, and effects of TGFβ family, FoxO, and Hippo signaling/activity.

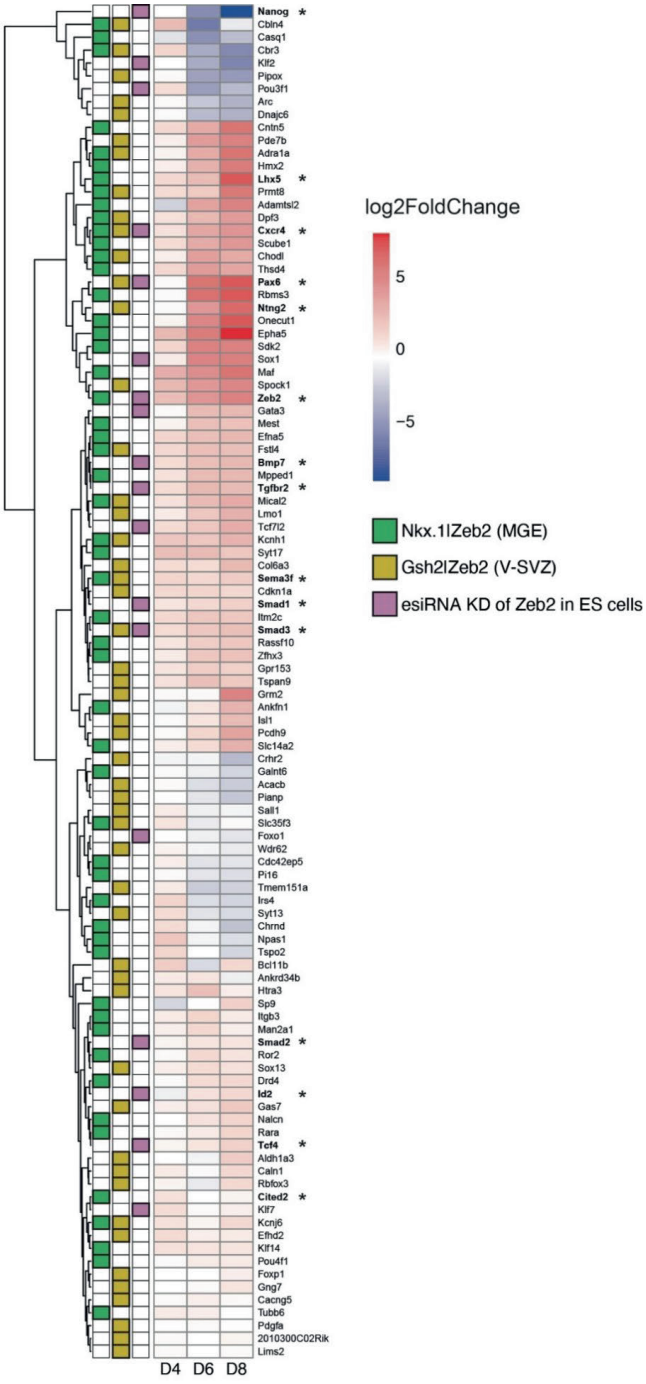


Figure S7.4. Expression of genes resulting from the meta-analysis of Zeb2-bound genes versus three independent RNA-seq datasets.

Heatmap visualizes the log2FoldChange of the resulting genes during mESC differentiation. RNA-seq datasets where the genes have been found to be differentially expressed are annotated as green (Nkx2.1-Cre|Zeb2), purple (Zeb2 KD in mESCs) or yellow (Gsh2-Cre|Zeb2) squares.

7.3.5. Zeb2 directly controls TGFβ/BMP-system component and neuronal differentiation/migration genes

We then validated 14 out of the 108 cross-referenced target genes, selected based on either being already known as a target of Zeb2 (*Nanog*) or as TGFβ/BMP-system component (*Bmp7*, *Tgfbr2*, *Smad1*, *Smad2*, *Smad3*, *Id2*, *Cited2*), or having a crucial role in neurogenesis up till neuronal maturation (*Sema3f, Cxcr4*, *Lhx5*, *Ntng2*, *Pax6*, *Tcf4*; their mRNA levels in ND-ESCs are highlighted in the heatmap in **Fig. S7.4,** genes have been marked with asterix).

Because *Zeb2*-KO ESCs do neither exit from primed pluripotency nor enter differentiation (Stryjewska *et al.*, 2017), we chose to validate our findings using shRNA-mediated Zeb2-KD at ND-D8 (shZeb2; *see* Experimental procdures), and analyzed these aforementioned 14 genes two days later (at D10 in NPCs) (**Fig. 7.4A**). At the latter read-out time point, >50% reduction of *Zeb2* mRNA was obtained (**Fig. 7.4B**). Zeb2 KD resulted in reduced mRNA levels of *Cxcr4*, *Ntng2* and *Pax6* (**Fig. 7.4B**), genes that each are involved in neuron specification and migration (Georgala *et al.*, 2011; Dias *et al.*, 2019). Zeb2 KD also caused downregulation of *Lhx5,* involved in differentiation of interneurons, including cytoskeletal rearrangements during dendritogenesis (Lui *et al.*, 2017), and of *Tcf4*, which acts in neurogenesis (Corneliussen *et al.*, 1991; Forrest *et al.*, 2014; Meert *et al.*, 2022). *Sema3f* is a cue for axon outgrowth and migration guidance of neurons, and was slightly upregulated (**Fig. 7.4B**). These results confirm the regulation by Zeb2 of its direct targets in later phases of neuronal differentiation and/or migration.

The expression of *Nanog*, the promoter of which binds Zeb2 as a repressor (Stryjewska *et al.*, 2017), was increased in Zeb2-KD cells (**Fig. 7.4B**). The KD of Zeb2 caused increased mRNA of *Bmp7*, *Tgfbr2, Smad1*, *Smad2, Smad3* and *Cited2* (**Fig. 7.4B**), fitting with the normal levels of Zeb2 that mount anti-TGFβ/BMP family effects (Conidi *et al.*, 2011). In contrast, *Id2* is downregulated in shZeb2-treated ESCs (**Fig. 7.4B**). *Id2* is normally activated by BMP-Smads and together with other Id proteins (Id1, Id3 and Id4) inhibits cell differentiation, e.g., Zeb2 represses *Id2* in immune cells to promote differentiation (Scott *et al.*, 2016). However, *Id2* as well as other *Id* genes (Tzeng and de Vellis, 1998; Blomfield *et al.*, 2019; Havrda *et al.*, 2008) is, like *Zeb2* (Seuntjens *et al.*, 2009; van den Berghe *et al.*, 2013), expressed in the developing forebrain.

Taken together, these data suggest an active and direct role for Zeb2 in repressing genes regulating stem cell pluripotency as well as a number of TGFβ/BMP-system components (*Bmp7*, *Smad1/2/3*), but also in activating genes in neurogenesis (*Cxcr4, Ntng2, Lhx5*).

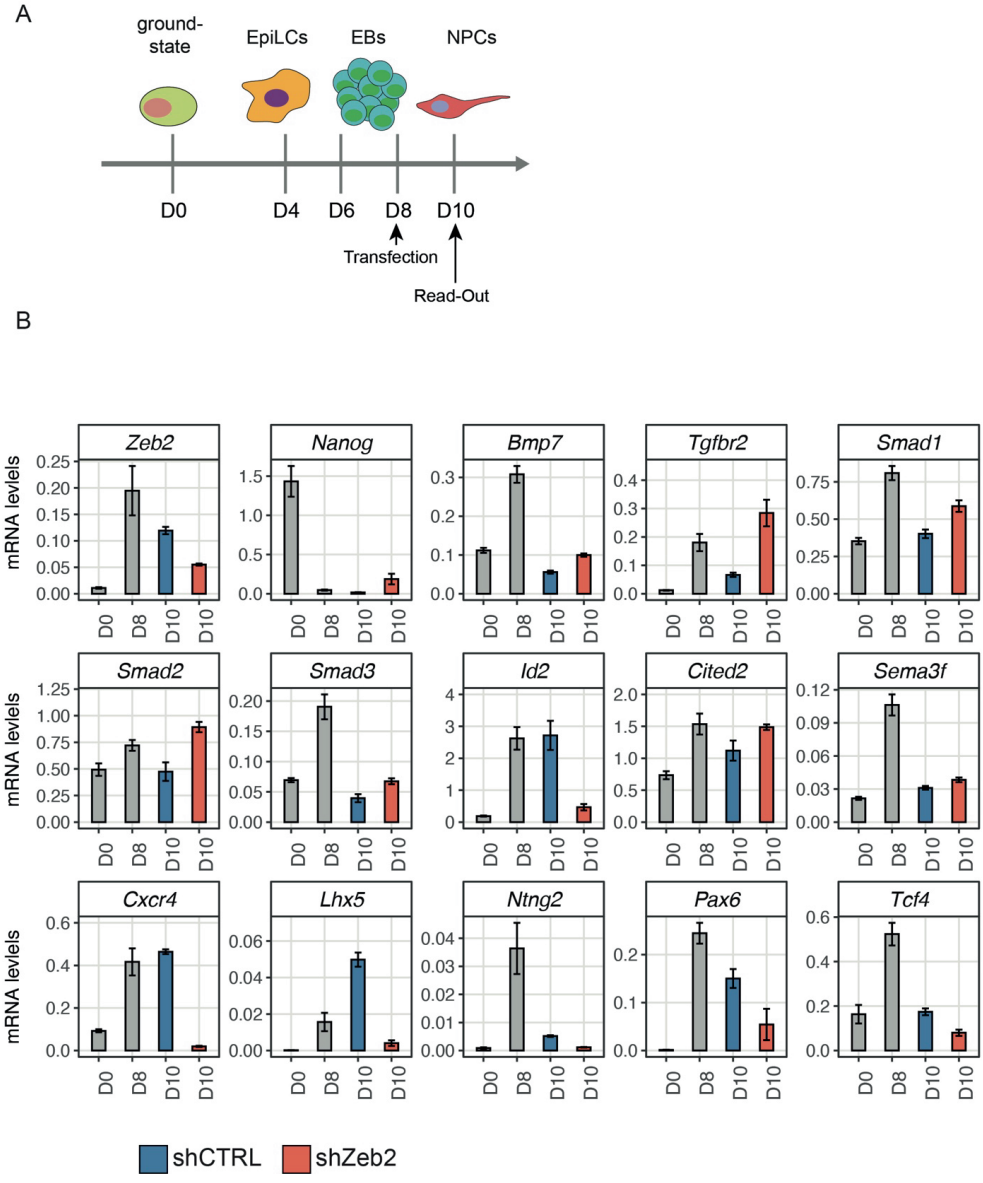


Figure 7.4. shRNA-mediated KD of Zeb2 discriminates between primary and secondary target genes.

A. Schematic overview of the shRNA transfection targeting Zeb2 (shZeb2) and read-out of the effect. Cellular aggregates at D8 of ND are dissociated and transfected with shZeb2 or against a scrambled, control sequence (shCTRL). Read-out is done two days after the start of shRNA addition. The list of shRNAs is given in **Table S4.2**. **B.** *Zeb2* levels after KD were reduced to 40-50% of their normal level (shZeb2, orange bars) compared to shCTRL (blue bars). *Bmp7*, *Cited2*, *Nanog*, *Sema3f*, *Smad1*, *Smad2*, *Smad3*, and *Tgfbr2* were upregulated following Zeb2 KD, whereas genes encoding for neuronal specification and migration (*Cxcr4*, *Lhx5*, *Ntng2*, *Pax6* and *Tcf4*) were downregulated.

7.3.6. Zeb2 potentiates its own gene expression, which is crucial for proper control of some of its direct target genes

Strikingly, in our ChIP-seq dataset, the peak with the highest enrichment (~200-fold) mapped 232 bp upstream of the Zeb2 TSS (**Fig. 7.5A**; **File S7.1**). For further functional studies of this site, we deleted the encompassing region (chr2:45109746-45110421), using CRISPR/Cas9 in wild-type mESCs, thereby obtaining *Zeb2*$^{\Delta P/\Delta P}$ ESCs (*see* Experimental procedures; **Fig S7.5**). **Fig. 7.5B** shows that the Zeb2 mRNA levels in the aforementioned homozygous *ΔP* clone stayed very low during ND, already from D8 onwards, as compared to control cells, classifying this *Zeb2*$^{\Delta P/\Delta P}$ clone as an alternative Zeb2-KD cell line. We used these *Zeb2*$^{\Delta P/\Delta P}$ mESCs to read-out the same genes that depend on intact Zeb2 levels and are Zeb2 ChIP+ (*see* **Fig. 7.4B**; in the *Annex* to this chapter we also do similar experiments in ESC to mesendodermal differentiation of this mutant cell line).

Levels of *Zeb2* mRNA stayed abnormally very low at D10 in *Zeb2*$^{\Delta P/\Delta P}$ ND-mESCs, whereas *Nanog* was still expressed and remained higher than in control cells, in line with Zeb2 being a direct repressor of *Nanog* in wild-type cells (**Fig. 7.5B**). Hence, Zeb2 levels, including those achieved by autoregulation, are critical, but to a different degree, for sets of genes.

The latter genes include neuronal-relevant genes such as *Cxcr4*, *Lhx5*, *Ntng2*, *Pax6*, *Tcf4* (**Fig. 7.5B**). Among the TGFβ/BMP-system components (*see* **Fig. 7.4**) we observed a slight reduction of *Bmp7*, *Smad1* and *Smad3*, whereas *Tgfbr2*, *Smad2*, *Cited2* and *Sema3f* were not affected in *Zeb2*$^{\Delta P/\Delta P}$ ND-mESCs. We speculate that Zeb2, and now including its newly identified autoregulation, plays an active role in regulating a number of genes involved in neuron determination (this is also the case in mesendodermal differentiation, *see Annex* to this chapter) and, together with other TFs, TGFβ/BMP-system component genes. Thus, the precise amounts of Zeb2, and in a critical stage also its autoregulation, are crucial in discriminating genes where Zeb2 plays the aforementioned primary, active role (as for *Cxcr4*, *Lhx5*, *Ntng*2 etc). For these genes, ~50% reduction (as observed in the Zeb2-KD) or the novel *ΔP/ΔP* "peak" mutation is sufficient to strongly deregulate them, but other genes' expression is either not or just slightly affected upon ND applied here (*Sema3f*, *Smad2*, *Cited2* vs. *Bmp7*, *Tgfbr2*, *Smad1*, *Smad3*).

Because Zeb2 also binds to phospho(p)-Smads (Verschueren *et al.*, 1999; Conidi *et al.*, 2011; Weng *et al.*, 2012; Deryckere *et al.*, 2020), we also scanned the Zeb2 ChIP+ direct target genes, and for which we saw strong deregulation upon Zeb2 KD and/or in *Zeb2*$^{\Delta P/\Delta P}$ cells during ND (wherein notably Smad activation is not stimulated; Smads will be activated in mesendodermal differentiation protocols, *see Annex* to this chapter). In particular, we scanned for the presence of (i) the Zeb half-sites CACCT(G) (pragmatically neglecting the variable spacing between them; Remacle *et al.*, 1999) that combine with (ii) candidate p-Smad binding and responsive genes (using GTC($^{T}/_{G}$)CT($^{T}/_{G}$)($^{A}/_{C}$)GCC for p-Smad1/Smad5, GTCTAGAC for p-Smad2/3) and (iii) the co-Smad Smad4 (C($^{C}/_{T}$)AGAC) (Hill, 2016), using the Jaspar database. **Fig. S7.6** shows the distribution of such identified Zeb and Smad-binding motifs (threshold score >85%) in those genes strongly affected by Zeb2 KD and/or in *Zeb2*$^{\Delta P/\Delta P}$ cells. Interestingly, in the regions where Zeb2 binds close to the TSS (*Zeb2*, *Ntng2*, *Lhx5, Nanog*), the p-Smad and Smad4 binding elements are sometimes present in very close proximity of the ChIP+ Zeb2-bound E-box, indicating a possible cross-talk between receptor-activated Smads and Zeb2 in regulating target genes indicating a possible cross-talk between receptor-activated Smads and Zeb2 in regulating target genes.

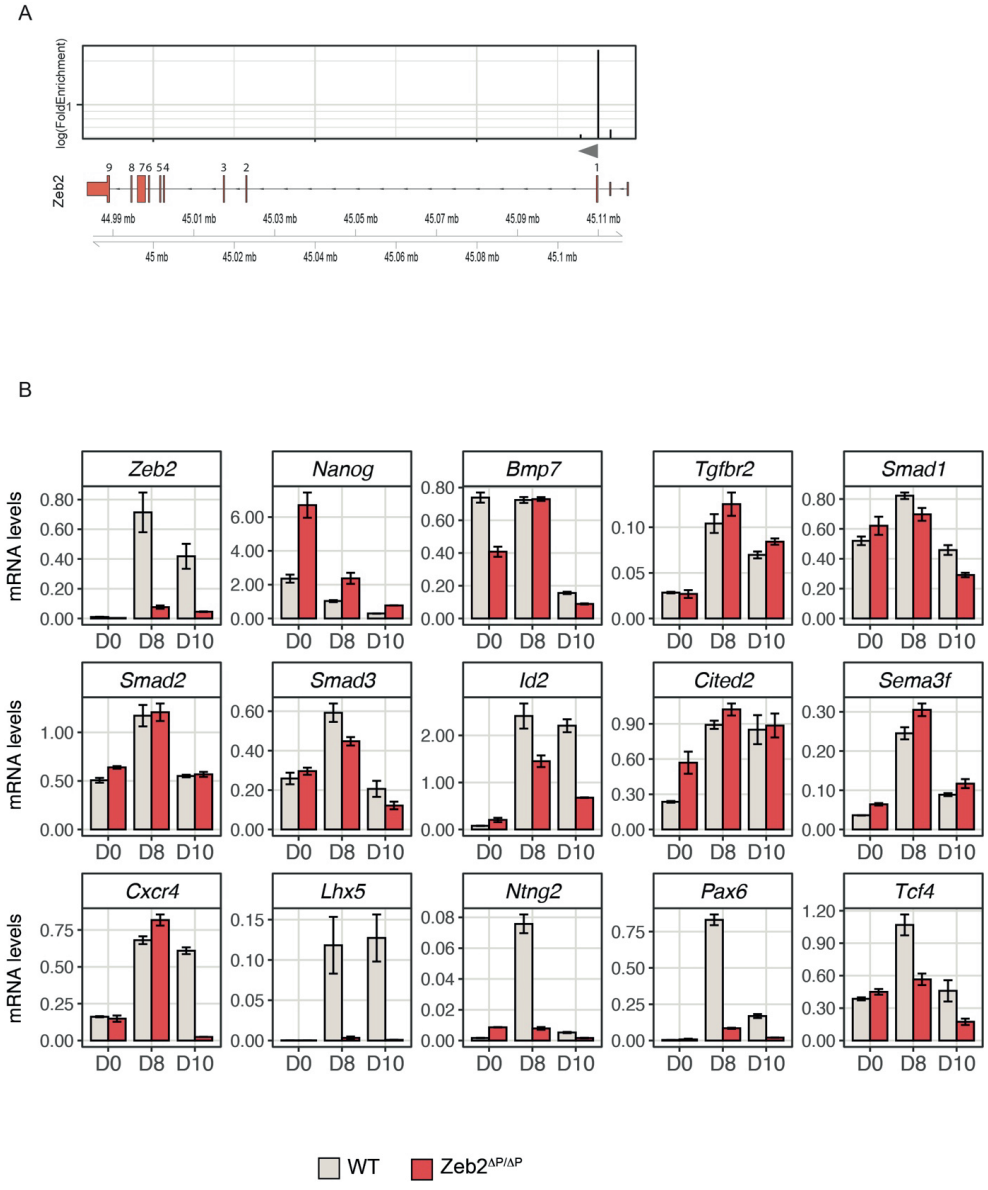


Figure 7.5. Deletion of the Zeb2-binding, candidate autoregulatory site (chr2:45109746-45110421) impairs Zeb2 mRNA levels and neuronal markers.

A. Schematic overview of the log(FoldEnrichment) of the peaks identified by ChIP-seq located in the mouse *Zeb2* locus, and localization on top of *Zeb2* intron/exon structure. The highest peak is located 232 bp upstream of the first translated exon. Grey arrow indicates the TSS of *Zeb2*. **B.** Zeb2 mRNA levels are strongly reduced in the $Zeb2^{\Delta P/\Delta P}$ clone. Expression levels (mRNA) of the target genes validated with shRNA (**Fig. 4.4**). Most, but not all of the genes found to be affected following Zeb2 KD are also deregulated in $Zeb2^{\Delta P/\Delta P}$ mESCs, in particular *Id2* and the neuronal markers *Cxcr4*, *Lhx5*, *Ntng2*, *Pax6* and *Tcf4*.

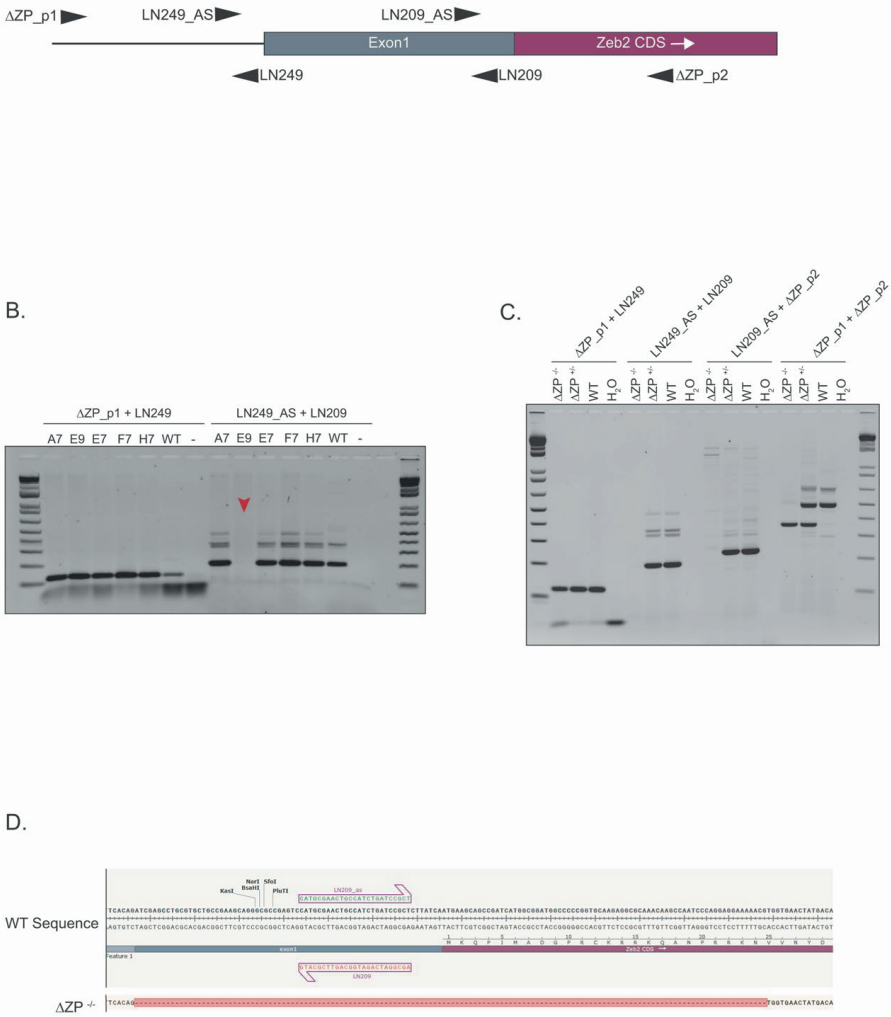


Figure S7.5. Genotyping of *Zeb2*$^{\Delta P/\Delta P}$ mESCs.

A. Primers (listed in **Table 4.1**) used for genotyping. **B.** PCR confirming the proper deletion of the selected region in clone E9. **C.** Further PCR showing the difference in molecular weight for regions amplified with different set of primers and discriminating between WT, heterozygous (*Zeb2*$^{\Delta P/+}$) and homozygous (*Zeb2*$^{\Delta P/\Delta P}$) deletion clones. **D.** Sanger sequencing showing the correct deletion of the selected region in *Zeb2*$^{\Delta P/\Delta P}$ clone.

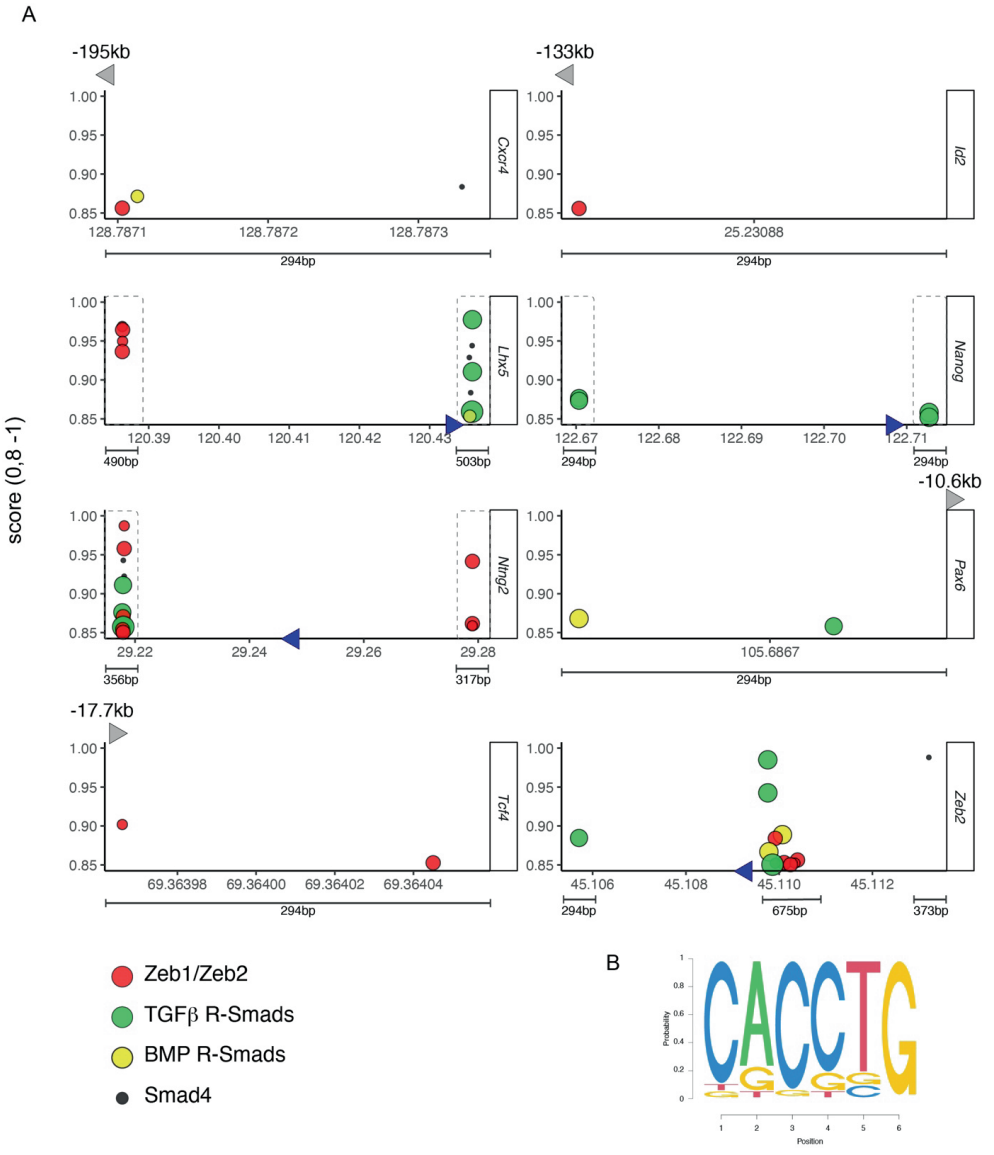


Figure S7.6. Distribution of Zeb binding sites (performed as single E-boxes; for details, *see* main text) **and TGFβ/BMP activated (phospho-)Smads and Smad4 binding elements within the identified Zeb2-bound regions for primary target genes.**

A. Grey triangles represent distance from the TSS, while blue triangles depict the TSS of the individual genes. Direction of the arrow shows also whether the genes is transcribed on the + or – strand. **B.** Motif analysis identifies the binding motif for Zeb2 as CACCTG.

7.4. Discussion

To our knowledge, we mapped for the first time the endogenous GWBS for Zeb2 by ChIP-seq, notably in ESC-derived NPCs gene-edited for this purpose. We have previously used ESCs established from *Zeb2*$^{\Delta ex7/\Delta ex7}$-KO pre-implantation embryos (Higashi *et al.*, 2002) and, for rescue purposes, such KO cells in which $Flag_3$-Strep-tagged Zeb2 (at its N-terminus) was produced, as active Zeb2, from a Cre-controllable *Rosa26* locus (Stryjewska *et al.*, 2017). Conceptually, with regard to Zeb2 levels, the latter cells are different from the mESCs that were newly established here. In the original *Rosa26*-Zeb2 cells *Zeb2* is not subjected to its normal temporal regulation during cell differentiation, although its levels in ND are never leading to unwanted vast/massive overproduction of this Zeb2, neither in cultured cells (Stryjewska *et al.*, 2017) nor in mouse models (van den Berghe *et al.*, 2013).

However, precise dosage of Zeb2 mRNA and protein is a critical factor *in vivo* (for a recent discussion, see Birkhoff *et al.*, 2021). This is concluded from transgenic Zeb2 cDNA-based rescues in *Zeb2*-KO ESCs and similar genetic rescues in *Zeb2*-mutant cells in mice, which can via heterozygous/homozygous combinations create a large panel of allele-number dependent Zeb2 levels (in interneurons, van den Berghe *et al.*, 2013; in NK cells, van Helden *et al.*, 2015; in ESCs, Stryjewska *et al.*, 2017). Another illustration of fine-tuned control of Zeb2 levels are miRs targeting Zeb2, and lncRNAs that regulate these miRs (Guan *et al.*, 2020; Jiang *et al.*, 2020; Yao *et al.*, 2020; Cheng *et al.*, 2021), with Zeb2 on its turn also controlling some of its own miR-encoding genes or clusters (Brabletz and Brabletz, 2010; Exposito-Villen *et al.*, 2018; Gregory *et al.*, 2008). In our ChIP-seq we find 125 peaks (~5% of the total) that correspond to the TSSs of 98 miR-genes (**File S7.1**). Among these miR-genes, Zeb2 binds to loci encoding miR-144, miR-148a, miR-9 and miR-153, known to target Zeb2 in the context of e.g., tumor progression (Guan *et al.*, 2015; Pan *et al.*, 2016; Nourmohammadi *et al.*, 2019; Wahab *et al.*, 2020). We recently added the identification, in human iPSCs subjected to ND, of *Zeb2* distant (~600 kb upstream) enhancers, which act through DNA-looping to the *ZEB2* promoter-proximal region (Birkhoff *et al.*, 2020). In addition, we have documented dynamic expression patterns of *Zeb2* in early embryos (Lerchner *et al.*, 2000; van Grunsven *et al.*, 2000; Van de Putte *et al.*, 2003; Miyoshi *et al.*, 2006; Takagi *et al.*, 2015). We have also shown that cDNA-based expression of various tag-Zeb2 proteins is compatible with functional embryology-type and action mechanism studies (Papin *et al.*, 2002; Verstappen *et al.*, 2008; Stryjewska *et al.*, 2017). Importantly, our Zeb2-V5 allele thus enables normal production of tag-Zeb2 from its endogenous locus.

Only two studies present ZEB2 ChIP-seq data in human cells, i.e. SNU398 hepatocellular carcinoma and K562 erythroleukemia cells, respectively (Balcik-Ercin *et al.*, 2018; Yang *et al.*, 2018). In K562 cells, ZEB2 binds to the promoters of *NR4A2*, *NEUROG2* and *PITX3*, expressed in midbrain dopaminergic neurons, wherein Zeb2 negatively regulates axon growth and target innervation (Hegarty *et al.*, 2017). In SNU398, ZEB2 represses *GALNT3*, which is normally expressed in epithelial cells. This repression co-incides with acquisition of a mesenchymal phenotype, linking ZEB2 here again to an EMT-like process. These two valuable studies also present limitations. The use of cancer cell lines of genomic instable nature may create possible bias in ChIP-seq, and in any case they overproduce ZEB2. Our ChIP-seq identifies >2,400 peaks for Zeb2-V5 in mESCs at D8 of ND. This is in cells not stimulated with TGFβ family ligands, with 37.5% of Zeb2 sites mapping close to the TSS (shen defined as -10/+10kb). Most of these genes function in growth factor/cytokine signaling and/or encode transcriptional regulators, the latter suggesting that Zeb2 orchestrates other co-operating TFs driving the transcriptomic

signature of NPCs. The regulation of Wnt signaling by Zeb2 is in line with observations that inhibition of the Wnt-βcat pathway suppresses ND *in vitro* and *in vivo*, and that Wnt (and Zeb2)-controlled *Tcf4* expression promotes neurogenesis and is required for normal brain development (Hirabayashi *et al.*, 2004; Slawny and O'Shea, 2011; Li *et al.*, 2018; Li *et al.*, 2019; Mesman *et al.*, 2020; for a discussion of the consequences for phenotypic convergence between MOWS and PTHS, *see* Meert *et al.*, 2022).

The 2,294 Zeb2-peaks map to 1,952 protein-coding genes, of which 1,244 are DEGs in ND-mESCs. Strikingly, the strongest enrichment of Zeb2 occurs on the *Zeb2* promoter itself, leading to the identification of a novel self-regulatory mechanism where Zeb2 binds upstream its TSS to maintain its levels sufficiently high, also and at least during ND. While this autoregulation needs further investigation in Zeb2-dependent differentiation and/or maturation of other cell types (e.g. in cKO mouse models or in ND-iPSCs derived from appropriate MOWS patient cells), we propose that lower Zeb2 levels might compromise this autoregulatory loop. Deletion of the autoregulatory site from both *Zeb2* alleles (in the *ΔP/ΔP* cells) results in a significant decrease of Zeb2 mRNA levels, but *Zeb2* is still partially expressed, and these cells can still exit from pluripotency and differentiate (contained in part in **Fig. 7.5**; *data not shown*). A number of genes, which are mainly linked to neuron maturation, are significantly affected in $Zeb2^{\Delta P/\Delta P}$ ESCs, whereas TGFβ/BMP-system component genes are not deregulated. Zeb2 dosage might thus underlie this difference in regulating its direct, ChIP+ genes in our ESCs. Zeb2 might be key to maintaining expression of neuronal genes, while for TGFβ/BMP-system genes Zeb2 may co-operate with other TFs (including p-Smads) or DNA-modifying enzymes to regulate the expression of target genes.

Zeb2 binds to receptor-activated p-Smads, and several studies indicate its negative regulation of BMP-Smad activation of specific target genes, although Zeb2 also has Smad-independent functions (Conidi *et al.*, 2011; Deryckere *et al.*, 2020). BMP-pSmads bind to GGCGCC with high affinity (BabuRajendran *et al.*, 2010). Morikawa *et al.* (2011) have confirmed these results using ChIP-seq, and identified also a lower-affinity (so, higher BMP-doses required) BMP-Smad element (GGAGCC). For achieving full responsiveness it was proposed that the GG($^{A}/_{C}$)GCC element needs to be coupled with a Smad4 site, ideally located 5 bp away (Morikawa *et al.*, 2011), although Smad4-(in)dependent and robust BMP vs. Nodal-activated Smad signaling was recently nicely documented for its contribution to zebrafish morphogenesis (Guglielmi *et al.*, 2021). We find that in primary targets affected by varying levels of Zeb2, E-boxes are located close to Smad-binding motifs, with BMP-Smads not being activated in ND. However, whether Zeb2 and Smads are co-present in target regions will require further experiments, such as ChiP-on-ChIP assays, and (non-neural) differentiation protocols (involving stimulation of the cells by addition of BMP and/or Nodal/Activin; see *Annex* to this chapter). However, these may even be further complicated because of post-translational modification status of Zeb2, nuclear p-Smads and/or Smad4 (Liu *et al.*, 2016; Xu *et al.*, 2016; Kim *et al.*, 2020; Lin *et al.*, 2020).

Striking are also the 1,093 Zeb2-binding DEGs at D8. When we performed a gene-to-disease association using the human orthologues of these D8-DEGs, we found a clear association with several disorders (**Fig. S7.7**), including neurodevelopmental, mental and eye defects, which also occur in MOWS. Altogether, our data may provide novel insights into molecular aspects of MOWS, focusing on the *Zeb2* autoregulation mechanism, and Zeb2-driven regulation of genes linked to other congenital disorders.

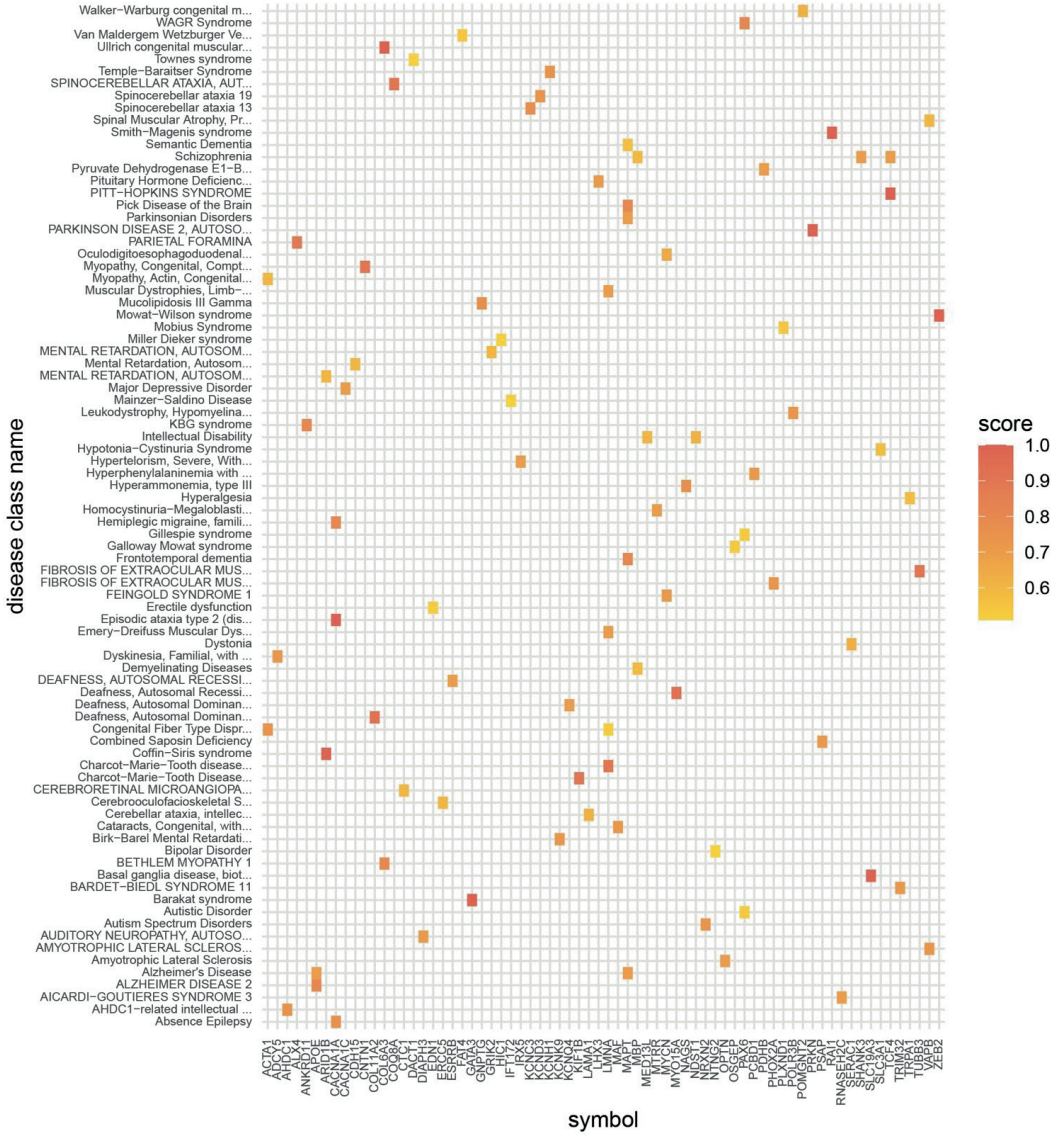


Figure S7.7. Gene-to-Disease association of the D8 DEGs bound by Zeb2.

The gene names of the DEGs expressed at D8 of neural differentiation were initially subjected to a mouse to human gene conversion. After that, gene to disease association analysis was performed showing that a number of those genes, could be associated with neurodevelopmental disorders, mental disorders, eye defects, seizures and speech impairment.

Several *Zeb2*-cKO mouse models have been generated, and for many bulk RNA-seq data are available. Here, we selected two such RNA-seq data sets (van den Berghe *et al.*, 2013; McKinsey *et al.*, 2013; Deryckere *et al.*, 2020). In addition, similar data were obtained for cultured mESCs, either *Zeb2*-KO cells (Stryjewska *et al.*, 2017) or cells submitted to ND wherein e.g., Zeb2 KD was performed (Dries *et al.*, 2020). Unfortunately, we could not include *Zeb2*-KO mESCs in these comparsons, for they convert far less efficiently to EpiLSCs, and fail in any case to differentiate beyond (Stryjewska *et al.*, 2017). The meta-analysis of these three different data sets, overlaid with the 1,952 Zeb2-V5+ loci/genes, show therefore a limited number of common targets, *Cxcr4* being the only common one in all data sets. This specifically narrows the Zeb2-bound gene collection to 108 in total. However, and interestingly, the latter enrich for GO terms such as pluripotency of stem cells, signaling by TGFβ and Wnt, cell fate commitment and neuron differentiation, all processes where Zeb2 plays a crucial role.

Out of these 108 genes, we selected 14 genes covering TGFβ/BMP, pluripotency, neuron migration, and neuron differentiation/maturation, and checked their levels 2 days after Zeb2 KD at ND-D8. Most of these 14 genes relevant to NPC status show to be critically depend on intact levels of Zeb2. They may help to explain why the defects caused by MOWS are observed later after birth and why (the few) missense mutations in MOWS (besides the far more frequent large deletions etc) present with milder syndromic manifestation. Both the cross-reference of Zeb2-ChIP+ genes with the transcriptome of ND-ESCs, and the meta-analysis, identify a number of common genes, such as *Bmp7*, *Tgfbr2*, *Tcf4*, *Smad1/2/3*, and *Sema3f* (**File S7.3**; **Fig. S7.4**), making Zeb2 a likely direct regulator of these. It is also intriguing that Zeb2 is recruited to and controlling *Tcf4* at D8, and that *Tcf4* is deregulated upon Zeb2 KD (using esiRNA, Dries *et al.*, 2020; and shRNA here). Mutations in *TCF4* cause Pitt-Hopkins syndrome (PTHS, OMIM#610954), a rare neurodevelopmental disorder with some defects overlapping with MOWS. The binding of Zeb2 to *Tcf4* opens new attractive roads to further investigate the crosstalk between these two TFs and their role in regulating crucial aspects of neurodevelopment (for an elaborate discussion, see Meert *et al.*, 2022).

7.5. Experimental procedures

ESC culture conditions and differentiation

CGR8 (strain 129) wild-type and Zeb2-Flag-V5+ mESCs were cultured and differentiated towards the neural lineage (Bibel *et al.*, 2007, with few modifications). Briefly, mESCs were cultured on 0.1% Gelatin-coated plates in ESC-medium: DMEM supplemented with 15% heat-inactivated (HI) FBS, 2mM L-Glutamine, 1x Non-Essential Amino Acids (NEAA), 143 µM β-Mercapto-Ethanol (β-EtSH) (all ThermoFisher Scientific, TFS) and LIF (10^3 U/ml).

For ND, $4x10^6$ cells were plated on non-adherent 10-cm dishes (Greiner) and allowed to form cellular aggregates (CAs) in 10 ml CA-medium (DMEM, 10% HI-FBS, 2 mM L-Glutamine, 1x NEAA, and 143 µM β-EtSH). From D4 of ND, cells were grown in CA-medium supplemented with 5 µM Retinoic Acid (RA). During the aggregation stages of ND, medium was changed every other day by carefully collecting the aggregates with a 10-ml pipet and transferring them to a 15-ml conical tube. The CAs were allowed to sink to the bottom of the tube where, after the previous medium was carefully discarded, the CAs were then resuspended in fresh medium and transferred back to the dishes. At D8 of ND, the aggregates were harvested and dissociated by resuspension in 1 ml Accutase (TFS) and pipetting them up-and-down using a 1-ml pipet, after shaking them in a 37°C water bath for 5 min. The Accutase was deactivated by adding 9 ml of fresh N2-medium (DMEM with 2 mM L-Glutamine, 50 µg BSA/ml and 1x N2-supplement) to the dissociated cells and pelleting the cells gently for 5 min at 200 g. The cells were resuspended in fresh N2. To ensure single-cell suspension, the cells were filtered by passing them through a 40-mm nylon cell strainer (Corning). 2.5×10^5 cells/cm^2 were plated on Poly-DL-Ornithine hydrobromide (Sigma) / Laminin (Sigma) coated plates. Cells were harvested at D8 or D10 of ND.

Western blots

To check Zeb2-V5 protein, the 2BE3-clone ESCs were subjected to ND till D8. Cytoplasmic and nuclear fractions were extracted using NePer-kit® (TFS). Protein concentrations were measured using the Bradfort BCA (TFS) and equal quantities of protein lysates were loaded on 6% SDS-PAGs and thereafter cut based on protein relative molecular mass. Gels were then transferred onto nitrocellulose membranes (Amersham Bioscience), which were incubated overnight with anti-Zeb2 (Seuntjens *et al.*, 2009) and anti-V5 (Life Technologies) antibody, followed by incubation at RT with HorseRadish Peroxidase (HRP) conjugated secondary anti-rabbit and anti-mouse antibodies, respectively (Jackson ImmunoResearch). Protein bands corresponding to Zeb2 or Zeb2-V5 were visualized on an AI-600 digital imager (Amersham|). As loading control, we used Valosin-containing Protein (VCP) and anti-VCP antibody (Santa Cruz sc-57492, mouse).

RNA extraction and RT-qPCR analysis

Total RNA was extracted from ESCs using TRI Reagent (Sigma), and used for cDNA synthesis with RevertAid RT Kit (TFS) with oligodT-primers. RT-qPCR was performed using SybrGreen dye (BioRad) on a CFX96 T1000 thermal cycler (BioRad). All data shown are averages of 3 independent biological replicates and 3 technical replicates, normalized to β-Actin mRNA levels. Primers are listed in **Table S7.1**. Analysis and data visualization was performed in R environment for statistical computing version 3.5.3, implemented with the *tidyverse* version 1.3 package (https://github.com/tidyverse).

Tag-Zeb2 mouse ESCs

gRNAs (**Table 7.1**) targeting *Zeb2*-ex9, and tracrRNA (Integrated DNA Technologies, IDT), were diluted to 125 ng/µl in duplex buffer (IDT). gRNAs were annealed to tracrRNA at 1:1 ratio at 95°C for 5 min and cooling the samples to room temperature (RT, 24°C). 250 ng of these annealed gRNAs were transfected in 350,000 mESCs together with 2 µg pX459-Cas9-puro vector and 1 µg ssDNA oligo of the Donor Template containing the FlagV5-tag sequence (**Table 7.1**). Transfection was done in a gelatin-coated 6-well plate using DNA:Lipofectamine2000 (ratio of 1:2). Six hours after transfection the medium was refreshed, at 24 hours the cells were Puromycin-selected (2 µg/ml). After 2 days, the remaining cells were transferred to gelatin-coated 10-cm dishes and given fresh ESC-medium (*see* below). Per dish 1,000; 1,500; or 2,000 cells were plated and allowed to form colonies. Medium was changed every other da7. Colonies were picked, expanded and genotyped by PCR (both outer and inner primer sets were used (**Table S7.1**; **Fig. S7.1**). All candidate clones were validated by Sanger-sequencing; correct clones were expanded and validated by Western blot.

CRISPR/Cas9-mediated deletion of Zeb2 binding site located at chr2:45109746-45110421

Oligonucleotides for gRNAs (**Table 7.1**) with target outside of this chr2-region were cloned into *Bbs*I-digested pX330-hspCas9-T2A-eGFP plasmid. All plasmids were sequenced. 4 µg of gRNA-plasmids (1 µg each) were transfected in 350,000 mESCs and selected (*see* above). After 24 hours these cells were sorted as GFP+ cells (Becton-Dickinson LSR Fortessa). Per well of a 6-well plate 1,000; 1,500; or 2,000 GFP+ cells were plated and colonies allowed to form, picked (*see* above) and genotyped by PCR using primers flanking this deletion, and within and outside of it. Clones showing a possible heterozygous or homozygous deletion, as concluded from the PCR analysis, were subjected to ND. At D8, they were harvested, RNA was isolated and cDNA synthesized (*see* below), and amplified (for the primers, see **Table S7.1**). All candidate clones were validated by Sanger-sequencing.

Chromatin immunoprecipitation (ChIP)

For ChIP $2x10^8$ cells were harvested at ND-D8 in 10 ml of PBS, and cross-linked at 1% formaldehyde (Sigma Aldrich) for 15 min, rotating at RT. Quenching followed with 125 mM glycine during 5 min, again rotating at RT. Cross-linked cells were washed twice with ice-cold PBS (5 min, 1,500 rpm (240 rcf), 4˚C), the pelleted cells snap-frozen, and stored at -80˚C. For sonication, the cell pellets were thawed on ice, resuspended in 1 ml sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease and phosphatase inhibitors (PPI, Roche) and incubated on ice for 10 min. DNA was sheared by sonicating the cells using a probe sonicator (32 cycles, 30 sec-on amplitude 9, and 30 sec-off). These samples were centrifuged at 13,200 rpm (17,000 rcf) for 10 min at 4˚C. Chromatin pellets were snap-frozen and stored at -80˚C.

To check sonication efficiency, 50 µl of sample was de-crosslinked overnight by adding 5 mM NaCl at 65˚C, shaking (950 rpm, Eppendorf ThermoMixer C). The next morning 5, 10 and 20 µl of sample were loaded on a 2% agarose gel, revealing ideally a DNA-smear around 300 bp. A 50-µl sample was used as control input. For immunoprecipitation, chromatin of 10^7 cells was diluted in ChIP-dilution buffer (17 mM Tris-HCl pH8, 170 mM NaCl, 1.2 mM EDTA, 0.01% SDS, 1.1% Triton X-100, with 1xPPI) to a final volume of 1 ml. Samples were pre-cleared by adding pre-washed Protein A/G agarose beads (Santa Cruz) and further incubation for 1 hour, rotating at 4˚C. Then, samples were centrifuged for 1 min (1,000 rpm; 106 rcf) at 4˚C, and the pre-cleared chromatin (supernatant) was transferred to a new low-binding 1.5-ml tube and incubated with 50 µl of pelleted V5-Agarose beads (Sigma Aldrich), rotating at 4˚C overnight. Before addition of V5-agarose beads, they were washed 5 times (5 min each) in PBS by rotating them. As negative control, half of the sample was incubated with Protein A/G beads (Santa Cruz, sc-2003).

The following day the beads were pelleted (1,000 rpm; 1 min) and washed as follows: once with lower-salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), transferred to non-stick low-binding 1.5-ml tubes and then washed once with high-salt buffer (i.e. lower-salt buffer, but now 500 mM NaCl), once washed with LiCl buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate (DOC), and twice washed with 10 mM Tris-HCl pH 8.0, 1 mM EDTA (each incubation for 5 min, rotating at 4˚C, followed by gently spinning down.

The protein-chromatin was then eluted from the beads by adding 250 µl of elution buffer (1% SDS + 100 mM $NaHCO_3$), rotating for 1 hour at RT twice, and combining the eluates from both steps. To the input sample, 450 µl of elution buffer was also added, and all samples were de-crosslinked by addition of 5 mM NaCl at 65˚C overnight, shaking 950 rpm. The day-after, 2µl Proteinase-K (10 mg/ml), 20 mM (final concentration) Tris-HCl pH6.5, 5mM (final concentration) EDTA pH8.0 and 10 mg/ml RNase-A (Sigma) were added to each sample and incubated for 1 hour at 45°C while shaking (700 rpm). DNA was extracted from the samples using the PCI method and diluted in water. Five independent ChIPs were performed, for a total of 10^8 mESCs used per condition, and pulled-down chromatin was pooled. ChIP efficiency was assessed by qPCR using primers amplifying *Cdh1* promoter sequences bound by Zeb2 (Stryjewska *et al.*, 2017). All primers used are listed in **Table S7.1**.

ChIP-seq

DNA libraries from input (i.e. control) and V5 ChIPs were prepared using ThruPLEX DNA protocol (TakaraBio) specific for low amounts of DNA, and sequenced on Illumina HiSeq-2500, and single reads of 50 bp were generated. Adapter sequences were trimmed from the 3'-end of the reads, after which the reads were aligned to the mm10/GRCm38 genome using HISAT2 (Kim *et al.*, 2015). From the alignments, secondary or supplementary, low quality and fragmented alignments (fragments > 150 bp) were filtered awa7. Peaks were called with MACS (Zhang *et al.*, 2008), and coverage was determined. 42 and 25 million reads were generated for input and V5 ChIP, respectively.

ChIP-seq data analysis

Peak calling was performed with MACS2 (Galaxy version 2.1.1.20160309.6) (Zhang *et al.*, 2008; Feng *et al.*, 2012), with default parameters (narrow peak calling, *Mm*1.87e9, FDR < 0.05) using the input sample as background. The No model parameter was used, and the extension size was set on 210 bp based on the predicted fragment lengths from the alignments (MACS2 predict-tool, Galaxy version 2.1.1.20160309.1)] (Zhang *et al.*, 2008; Feng *et al.*, 2012). The distance of the aligned reads from the TSS of the gene was analyzed using ComputeMatrix (Galaxy version 3.3.2.0.0) and PlotHeatmap Galaxy version 3.3.2.0.1; the used matrix is based on the log2ratio of the aligned ChIP peaks over the input, calculated using BamCompare (Galaxy version 3.3.2.0.0) (Ramirez *et al.*, 2016).

Transcription Factors motifs enrichment analysis

To identify the TFBS in Zeb2-binding regions associated with DEGs, we first extracted unique Zeb2-peaks located 10kb +/- from TSS. Next, we analyzed the TFBS enrichment using UniBind enrichment tool with motifs from the UniBind database (using reference genome GRCm38/mm10) (Puig *et al.*, 2021). As a background for the analysis all Zeb2-peaks were used. The p-value from Fisher's exact test after multitest adjustments was used to identify significantly enriched TFBS. Further, the max rank index calculated based on the odds ratio, p-value from Fisher's exact test and the number of overlapping regions, was applied to rank the top enriched motifs.

RNA-seq

The quality of total RNA (of biologically independent triplicates) of wild-type mESCs at D0, and at ND-D4, D6 and D8, was checked on Agilent Technologies-2100 Bioanalyzer, using a RNA nano- assa7. All samples had RIN value of 9.8 or higher. Triplicate RNA-seq libraries were prepared (Illumina TruSeq stranded mRNA protocol; www.illumina.com). Briefly, 200 ng of total RNA was purified using polyT-oligo-attached magnetic beads for ending with polyA-RNA. The polyA-tailed RNA was fragmented, and cDNA synthesized (SuperScript II, random primers, in the presence of Actinomycin D). cDNA fragments were end-repaired, purified (AMPureXP beads), A-tailed using Klenow exo-enzyme and dATP. Paired-end adapters with dual index (Illumina) were ligated to the A-tailed cDNA fragments and purified (AMPureXP beads).

The resulting adapter-modified cDNAs were enriched by PCR (Phusion polymerase) as follows: 30 sec at 98°C, 15 cycles of (10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C), 5 min at 72°C. PCR products were purified (AMPureXP beads) and eluted in 30 µl resuspension buffer. One µl was loaded on an Agilent 2100 Bioanalyzer using a DNA-1000 assay to determine concentration and for quality check. Cluster generation was performed according to the Illumina TruSeq SR Rapid Cluster kit v2 Reagents Preparation Guide (www.illumina.com). After hybridization of the sequencing primer, sequencing-by-synthesis was performed using a HiSeq-2500 with a single-read 50-cycle protocol followed by dual index sequencing. Illumina adapter sequences have been trimmed off the reads, which were subsequently mapped against the GRCm38 mouse reference (using HiSat2 version 2.1.0) (Kim *et al.*, 2015). Gene expression values were called (using HTSeq-count version 0.9.1) (Anders *et al.*, 2015) and Ensembl release 84 gene and transcript annotation. Sample QC and DEG analysis have been performed in the R environment for statistical computing (version 3.5.3, using DESeq2 version 1.22.1 and Tidyverse version 1.2.1 (https://github.com/tidyverse, R Core Team, 2018; https://www.r-project.org/; Love *et al.*, 2014).

Meta-analysis

RNA-seq datasets (as DEG tables, from van den Berghe *et al.* (2013) and Deryckere *et al.* (2020) were downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/, GSE35616 and GSE103003, respectively). Cross-referencing and visualization was performed in R, using Tidyverse, VennDiagram and pheatmap packages.

Pathways Enrichment, Gene Ontology, Function analysis and Gene to Disease Association

These were performed with StringDB package for R (Szklarczyk *et al.*, 2019), while for Gene-to-Disease association Disgenet2R for R was used (Pinero *et al.*, 2020).

Zeb2 shRNA-mediated knock-down

Zeb2-KD in was done by transfecting Zeb2-shRNAs into mESCs at ND-D8. For this, the CAs were dissociated as described above, and single-cell suspensions were transfected using Amaxa Nucleofector II (using kit V, program A-33). In total 4 µg of shRNA was used for transfection of $4.5x10^6$ cells. After transfection, ells were plated in 5 ml N2-medium on a poly-ornithine/laminin-coated 6-cm cell culture dish. Two hours post-

transfection, the medium was refreshed, and 24 hours after transfection was changed to N2-medium+Puromycin (*see* above) for 48 hours. The cells were then harvested, and KD efficiencies examined by RT-qPCR. As a control, scrambled shRNA was used.

Acknowledgments

This work was supported mainly by Erasmus University Medical Center via departmental funds, extra funds coordinated with its Executive Board, and BIG project funding, a collaborative extra support between Erasmus University Rotterdam and its University Medical Center for promoting fundamental research at the Theme of Biomedical Sciences, and hence the Department of Cell Biology. We thank all co-workers of the Center for Biomics-Genomics at Erasmus University Medical Center for their expert technical assistance, and colleagues Frank Grosveld and Raymond Poot for constructive discussions.

Tables

Table 7.1: **gRNAs and donor template used for CRISPR/Cas9-mediated *Zeb2* editing.**

Name	Sequence	CRISPR/Cas9
FlagV5 Donor Template	aaaatggaaaccaaatcagaccacgaagaagacaatatgga agatggcatcgaaGACTACAAAGACGATGACGACA AG***gatatc***GGTAAGCCTATCCCTAACCCTCTCCTC GGTCTCGATTCTACG**TAA**actactgcattttaagcttcc tatttttttttttccagtagtattgtt	In-frame knock-in of Flag-V5 tag
gRNA_ex9_1	GGAAACCAAATCAGACCACGAGG	
gRNA_ΔZP1	CCCGCGCGCGTTTCAATGGGCGC	Zeb2 peak deletion
gRNA_ΔZP2	CCCTCGCGAGTGCAACACACCAA	
gRNA_ΔZP3	GGGCTCGGAGCGCTGCCGATCGG	
gRNA_ΔZP4	CCGCTGGACCGGGGGGGAGTTGA	

Donor template: lowercase: homology arms located in exon 9 and 3' UTR of *Zeb2*, underlined lowercase: mutated PAM sequences, uppercase: *FLAG* coding sequence, lowercase italics bold: *EcorI* restriction site, underlined uppercase: *V5* coding sequence, bold uppercase: STOP codon.

Supplementary tables

Table S7.1: List of primers used in the study.

Primer name	Sense/Antisense	Sequence (5' -> 3')	Application
FlV5mZeb2Ex9_Fwd	Sense	GGCTTACCTGCAGAGCATCA	genotyping
FlV5mZeb2Ex9_Rev	Antisense	CTCCATCTAACTCTGTCTTGGC	genotyping
FlV5_Fwd	Sense	CTACTCGCAGCACATGAATC	genotyping
FlV5_Rev	Antisense	GAGAGGGTTAGGGATAGGC	genotyping
ΔZP_P1_Fwd	Sense	GTCAGTCCGTCCCCAGGTTT	genotyping
ΔZP_P2_Rev	Antisense	GGCATGCTAGCTGGGCTGGT	genotyping
LN249_Fwd	Sense	GGAGCAAACTGAACAAAACCTCGCC	genotyping
LN249_Rev	Antisense	GGCGAGGTTTTGTTCAGTTTGCTCC	genotyping
LN209_Fwd	Sense	AGCGGATCAGATGGCAGTTCGCATG	genotyping
LN209_Rev	Antisense	CATGCGAACTGCCATCTGATCCGCT	genotyping
Zeb2_Fwd	Sense	CAATGCAGCACTTAGGTGTA	qPCR
Zeb2_Rev	Antisense	TTGCCTAGAAACCGTATTGT	qPCR
Zeb2V5_Fwd	Sense	GAAACGATACGGGATGAGGA	qPCR
Zeb2V5_Rev	Antisense	AGGAGAGGGTTAGGGATAGG	qPCR
Nanog_Fwd	Sense	TCT TCC TGG TCC CCA CAG TTT	qPCR
Nanog_Rev	Antisense	GCA AGA ATA GTT CTC GGG ATG AA	qPCR
Pou5f1_Fwd	Sense	AGA GGA TCA CCT TGG GGT ACA	qPCR
Pou5f1_Rev	Antisense	CGA AGC GAC AGA TGG TGG TC	qPCR
Sox2_Fwd	Sense	GCGGAGTGGAAACTTTTGTCC	qPCR
Sox2_Rev	Antisense	CGGGAAGCGTGTACTTATCCTT	qPCR
Pax6_Fwd	Sense	ACATCTTTTACCCAAGAGCA	qPCR
Pax6_Rev	Antisense	GGCAAACACATCTGGATAAT	qPCR
Acrv1b_Fwd	Sense	CTGCCTACAGACCAACTACACC	qPCR
Acrv1b_Rev	Antisense	CCACGCCATCCAGGTTAAAGA	qPCR
Lhx5_Fwd	Sense	AGAACCGAAGGTCCAAAGAA	qPCR
Lhx5_Rev	Antisense	TCACTTTGGTAGTCTCCGTA	qPCR
Ntng2_Fwd	Sense	CAAGGACTCTACGCTTTTCG	qPCR
Ntng2_Rev	Antisense	AGCACTCGCAGTCTTGAAAT	qPCR
Sema3f_Fwd	Sense	CTACACAGCATCCTCCAAGA	qPCR
Sema3f_Rev	Antisense	ACGGCATTCTTGTTTGCATT	qPCR
Smad1_Fwd	Sense	TACTATGAGCTCAACAACCG	qPCR
Smad1_Rev	Antisense	GAAGCGGTTCTTATTGTTGG	qPCR
Smad3_Fwd	Sense	CACGCAGAACGTGAACACC	qPCR
Smad3_Rev	Antisense	GGCAGTAGATAACGTGAGGGA	qPCR
Sox13_Fwd	Sense	CTTACAGGAGGTTGTGCCA	qPCR

Sox13_Rev	Antisense	TCCTTAGCTTCCACATTGCT	qPCR
Stat3_Fwd	Sense	CAATACCATTGACCTGCCGAT	qPCR
Stat3_Rev	Antisense	GAGCGACTCAAACTGCCCT	qPCR
Tcf4_Fwd	Sense	TTGAAGATGTTTTCGCCTCC	qPCR
Tcf4_Rev	Antisense	CCTGCTAGTCATGTGGTCAT	qPCR
Tgfbr2_Fwd	Sense	GAAGGAAAAGAAAAGGGCGG	qPCR
Tgfbr2_Rev	Antisense	TGCTGGTGGTGTATTCTTCC	qPCR
Amylase_Fwd	Sense	GGCTGAGTGTTCTGGGAT	ChIP-qPCR
Amylase_Rev	Antisense	CACGGTGCTCTGGTAGAT	ChIP-qPCR
Cdh1_R1_Fwd	Sense	GCTAGGCTAGGATTCGAACGAC	ChIP-qPCR
Cdh1_R1_Rev	Antisense	TGCAGGGCCCTCAACTT	ChIP-qPCR

Table S7.2: shRNAs used.

Name	Sequence
shZeb2_1	CCGG**CCGAATGAGAAACAATATCAA**CTCGAGTTGATATTGTTTCTCATTCGGTTTTTG
shZeb2_2	CCGG**CCTCAGGAATTTGTGAAGGAA**CTCGAGTTCCTTCACAAATTCCTGAGGTTTTTG
shZeb2_3	CCGG**CCAGTGTCAGATTTGTAAGAA**CTCGAGTTCTTACAAATCTGACACTGGTTTTTG
shZeb2_4	CCGG**CCCATTTAGTGCCAAGCCTTT**CTCGAGAAAGGCTTGGCACTAAATGGGTTTTTG
shCTRL	CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT

Supplementary files

Supplementary File S7.1: Narrow peaks obtained from ChIP-seq.

Supplementary File S7.2: Transcriptomic data of Zeb2-bound genes in differentiating mESCs.

Supplementary File S7.3: Target genes bound by Zeb2 and deregulated in different already published RNA-seq datasets.

These files are available in the uploaded preprint at
https://www.biorxiv.org/content/10.1101/2021.07.06.451350v2

or can be downloaded as pdf file there at
https://www.biorxiv.org/content/10.1101/2021.07.06.451350v2.supplementary-material
as the .zip file [supplements/451350_file07.zip]

Literature references

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018 Jul 2;46(W1): W537-W44.

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015 Jan 15;31(2): 166-9.

BabuRajendran N, Palasingam P, Narasimhan K, Sun W, Prabhakar S, Jauch R, Kolatkar PR. Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. Nucleic Acids Res. 2010 Jun;38(10): 3477-88.

Balcik-Ercin P, Cetin M, Yalim-Camci I, Odabas G, Tokay N, Sayan AE, Yagci T. Genome-wide analysis of endogenously expressed ZEB2 binding sites reveals inverse correlations between ZEB2 and GalNAc-transferase GALNT3 in human tumors. Cell Oncol (Dordr). 2018 Aug;41(4): 379-93.

Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski CV, Gillis DR, Phillips-Cremins JE. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. 2017 Jul;27(7):1139-52.

Bibel M, Richter J, Lacroix E, Barde YA. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc. 2007;2(5): 1034-43.

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, Nowosad K, van den Hout M, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. Hum Mol Genet. 2020 Aug 29;29(15): 2535-50.

Birkhoff JC, Huylebroeck D, Conidi A. ZEB2, the Mowat-Wilson Syndrome Transcription Factor: Confirmations, Novel Functions, and Continuing Surprises. Genes (Basel). 2021 Jul 3;12(7):1037.

Blomfield IM, Rocamonde B, Masdeu MDM, Mulugeta E, Vaga S, van den Berg DL, Huillard E, Guillemot F, Urban N. Id4 promotes the elimination of the pro-activation factor Ascl1 to maintain quiescence of adult hippocampal stem cells. Elife. 2019 Sep 25;8: e48561.

Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? EMBO Rep. 2010 Sep;11(9): 670-7.

Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, Frenck W, Radlwimmer B, Lichter P, Engels H, Driever W, Kubisch C, Weber RG. Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. Hum Mol Genet. 2007 Jun 15;16(12):1488-94.

Cacheux V, Dastot-Le Moal F, Kääriäinen H, Bondurand N, Rintala R, Boissier B, Wilson M, Mowat D, Goossens M. Loss-of-function mutations in SIP1 Smad interacting protein 1 result in a syndromic Hirschsprung disease. Hum Mol Genet. 2001 Jul 1;10(14):1503-10.

Cerruti Mainardi P, Pastore G, Zweier C, Rauch A. Mowat-Wilson syndrome and mutation in the zinc finger homeo box 1B gene: a well defined clinical entit7. J Med Genet. 2004 Feb;41(2):e16.

Cheng H, Zhao H, Xiao X, Huang Q, Zeng W, Tian B, Ma T, Lu D, Jin Y, Li 7. Long Non-coding RNA MALAT1 Upregulates ZEB2 Expression to Promote Malignant Progression of Glioma by Attenuating miR-124. Mol Neurobiol. 2021 Mar;58(3): 1006-16.

Comijn J, Berx G, Vermassen P, Verschueren K, van Grunsven L, Bruyneel E, Mareel M, Huylebroeck D, van Roy F. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell. 2001 Jun;7(6): 1267-78.

Conidi A, Cazzola S, Beets K, Coddens K, Collart C, Cornelis F, Cox L, Joke D, Dobreva MP, Dries R, Esguerra C, Francis A, Ibrahimi A, Kroes R, Lesage F, Maas E, Moya I, Pereira PN, Stappers E, Stryjewska A, van den Berghe V, Vermeire L, Verstappen G, Seuntjens E, Umans L, Zwijsen A, Huylebroeck D. Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFbeta/BMP signaling *in vivo*. Cytokine Growth Factor Rev. 2011 Oct-Dec;22(5-6): 287-300.

Corneliussen B, Thornell A, Hallberg B, Grundstrom T. Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. J Virol. 1991 Nov;65(11): 6084-93.

Deryckere A, Stappers E, Dries R, Peyre E, van den Berghe V, Conidi A, Zampeta FI, Francis A, Bresseleers M, Stryjewska A, Vanlaer R, Maas E, Smal IV, van IJcken WFJ, Grosveld FG, Nguyen L, Huylebroeck D, Seuntjens E. Multifaceted actions of Zeb2 in postnatal neurogenesis from the ventricular-subventricular zone to the olfactory bulb. Development. 2020 May 26;147(10): dev184861.

Dias CM, Punetha J, Zheng C, Mazaheri N, Rad A, Efthymiou S, Petersen A, Dehghani M, Pehlivan D, Partlow JN, Posey JE, Salpietro V, Gezdirici A, Malamiri RA, Al Menabawy NM, Selim LA, Vahidi Mehrjardi MY, Banu S, Polla DL, Yang E, Rezazadeh Varaghchi J, Mitani T, van Beusekom E, Najafi M, Sedaghat A, Keller-Ramey J, Durham L, Coban-Akdemir Z, Karaca E, Orlova V, Schaeken LLM, Sherafat A, Jhangiani SN, Stanley V, Shariati G, Galehdari H, Gleeson JG, Walsh CA, Lupski JR, Seiradake E, Houlden H, van Bokhoven H, Maroofian R. Homozygous Missense Variants in NTNG2, Encoding a Presynaptic Netrin-G2 Adhesion Protein, Lead to a Distinct Neurodevelopmental Disorder. Am J Hum Genet. 2019 Nov 7;105(5): 1048-56.

Dries R, Stryjewska A, Coddens K, Okawa S, Notelaers T, Birkhoff J, Dekker M, Verfaillie CM, Del Sol A, Mulugeta E, Conidi A, Grosveld FG, Huylebroeck D. Integrative and perturbation-based analysis of the transcriptional dynamics of TGFbeta/BMP system components in transition from embryonic stem cells to neural progenitors. Stem Cells. 2020 Feb;38(2): 202-17.

Dunn SJ, Martello G, Yordanov B, Emmott S, Smith AG. Defining an essential transcription factor program for naïve pluripotenc7. Science. 2014 344(6188):1156-60.

El Wakil A, Francius C, Wolff A, Pleau-Varet J, Nardelli J. The GATA2 transcription factor negatively regulates the proliferation of neuronal progenitors. Development. 2006 Jun;133(11):2155-65.

Exposito-Villen A, A EA, Franco D. Functional Role of Non-Coding RNAs during Epithelial-To-Mesenchymal Transition. Noncoding RNA. 2018 May 28;4(2): 14.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep;7(9): 1728-40.

Forrest MP, Hill MJ, Quantock AJ, Martin-Rendon E, Blake DJ. The emerging roles of TCF4 in disease and development. Trends Mol Med. 2014 Jun;20(6): 322-31.

Fu H, Cai J, Clevers H, Fast E, Gray S, Greenberg R, Jain MK, Ma Q, Qiu M, Rowitch DH, Taylor CM, Stiles CD. A genome-wide screen for spatially restricted expression patterns identifies transcription factors that regulate glial development. J Neurosci. 2009 Sep 9;29(36): 11399-408.

Funahashi J, Kamachi Y, Goto K, Kondoh H. Identification of nuclear factor delta EF1 and its binding site essential for lens-specific activity of the delta 1-crystallin enhancer. Nucleic Acids Res. 1991 Jul 11;19(3): 3543-7.

Garavelli L, Donadio A, Zanacca C, Banchini G, Della Giustina E, Bertani G, Albertini G, Del Rossi C, Zweier C, Rauch A, Zollino M, Neri G. Hirschsprung disease, mental retardation, characteristic facial features, and mutation in the gene ZFHX1B (SIP1): confirmation of the Mowat-Wilson syndrome. Am J Med Genet A. 2003 Feb 1;116A(4): 385-8.

Garavelli L, Ivanovski I, Caraffi SG, Santodirocco D, Pollazzon M, Cordelli DM, Abdalla E, Accorsi P, Adam MP, Baldo C, Bayat A, Belligni E, Bonvicini F, Breckpot J, Callewaert B, Cocchi G, Cuturilo G, Devriendt K, Dinulos MB, Djuric O, Epifanio R, Faravelli F, Formisano D, Giordano L, Grasso M, Grønborg S, Iodice A, Iughetti L, Lacombe D, Maggi M, Malbora B, Mammi I, Moutton S, Møller R, Muschke P, Napoli M, Pantaleoni C, Pascarella R, Pellicciari A, Poch-Olive ML, Raviglione F, Rivieri F, Russo C, Savasta S, Scarano G, Selicorni A, Silengo M, Sorge G, Tarani L, Tone LG, Toutain A, Trimouille A, Valera ET, Vergano SS, Zanotta N, Zollino M, Dobyns WB, Paciorkowski AR. Neuroimaging findings in Mowat-Wilson syndrome: a study of 54 patients. Genet Med. 2017 Jun;19(6):691-700.

Georgala PA, Carr CB, Price DJ. The role of Pax6 in forebrain development. Dev Neurobiol. 2011 Aug;71(8): 690-709.

Gregory PA, Bracken CP, Bert AG, Goodall GJ. MicroRNAs as regulators of epithelial-mesenchymal transition. Cell Cycle. 2008 Oct;7(20): 3112-8.

Guan H, Liang W, Xie Z, Li H, Liu J, Liu L, Xiu L, Li 7. Down-regulation of miR-144 promotes thyroid cancer cell invasion by targeting ZEB1 and ZEB2. Endocrine 2015 48: 566-74.

Guan J, Liu P, Wang A, Wang B. Long noncoding RNA ZEB2AS1 affects cell proliferation and apoptosis via the miR1225p/PLK1 axis in acute myeloid leukemia. Int J Mol Med. 2020 Oct;46(4): 1490-1500.

Guglielmi L, Heliot C, Kumar S, Alexandrov Y, Gori I, Papaleonidopoulou F, Barrington C, East P, Economou AD, French PMW, McGinty J, Hill CS. Smad4 controls signaling robustness and morphogenesis by differentially contributing to the Nodal and BMP pathways. Nat Commun. 2021 Nov 4;12(1):6374.

Havrda MC, Harris BT, Mantani A, Ward NM, Paolella BR, Cuzon VC, Yeh HH, Israel MA. Id2 is required for specification of dopaminergic neurons during adult olfactory neurogenesis. J Neurosci. 2008 Dec 24;28(52): 14074-86.

Hegarty SV, Sullivan AM, O'Keeffe GW. Zeb2: A multifunctional regulator of nervous system development. Prog Neurobiol. 2015 Sep;132: 81-95.

Hegarty SV, Wyatt SL, Howard L, Stappers E, Huylebroeck D, Sullivan AM, O'Keeffe GW. Zeb2 is a negative regulator of midbrain dopaminergic axon growth and target innervation. Sci Rep. 2017 Aug 17;7(1): 8568.

Higashi Y, Maruhashi M, Nelles L, Van de Putte T, Verschueren K, Miyoshi T, Yoshimoto A, Kondoh H, Huylebroeck D. Generation of the floxed allele of the SIP1 (Smad-interacting protein 1) gene for Cre-mediated conditional knockout in the mouse. Genesis. 2002 Feb;32(2): 82-4.

Hill CS. Transcriptional Control by the SMADs. Cold Spring Harb Perspect Biol. 2016Oct 3;8(10): a022079.

Hirabayashi Y, Itoh Y, Tabata H, Nakajima K, Akiyama T, Masuyama N, Gotoh 7. The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. Development. 2004 Jun;131(12): 2791-01.

Ishihara N, Yamada K, Yamada Y, Miura K, Kato J, Kuwabara N, Hara Y, Kobayashi Y, Hoshino K, Nomura Y, Mimaki M, Ohya K, Matsushima M, Nitta H, Tanaka K, Segawa M, Ohki T, Ezoe T, Kumagai T, Onuma A, Kuroda T, Yoneda M, Yamanaka T, Saeki M, Segawa M, Saji T, Nagaya M, Wakamatsu N. Clinical and molecular analysis of Mowat-Wilson syndrome associated with ZFHX1B mutations and deletions at 2q22-q24.1. J Med Genet. 2004 May;41(5):387-93.

Ivanovski I, Djuric O, Caraffi SG, Santodirocco D, Pollazzon M, Rosato S, Cordelli DM, Abdalla E, Accorsi P, Adam MP, Ajmone PF, Badura-Stronka M, Baldo C, Baldi M, Bayat A, Bigoni S, Bonvicini F, Breckpot J, Callewaert B, Cocchi G, Cuturilo G, De Brasi D, Devriendt K, Dinulos MB, Hjortshøj TD, Epifanio R, Faravelli F, Fiumara A, Formisano D, Giordano L, Grasso M, Grønborg S, Iodice A, Iughetti L, Kuburovic V, Kutkowska-Kazmierczak A, Lacombe D, Lo Rizzo C, Luchetti A, Malbora B, Mammi I, Mari F, Montorsi G, Moutton S, Møller RS, Muschke P, Nielsen JEK, Obersztyn E, Pantaleoni C, Pellicciari A, Pisanti MA, Prpic I, Poch-Olive ML, Raviglione F, Renieri A, Ricci E, Rivieri F, Santen GW, Savasta S, Scarano G, Schanze I, Selicorni A, Silengo M, Smigiel R, Spaccini L, Sorge G, Szczaluba K, Tarani L, Tone LG, Toutain A, Trimouille A, Valera ET, Vergano SS, Zanotta N, Zenker M, Conidi A, Zollino M, Rauch A, Zweier C, Garavelli L. Phenotype and genotype of 87 patients with Mowat-Wilson syndrome and recommendations for care. Genet Med. 2018 Sep;20(9): 965-75.

Jiang Y, Wu K, Cao W, Xu Q, Wang X, Qin X, Wang X, Li Y, Zhang J, Chen W. Long noncoding RNA KTN1-AS1 promotes head and neck squamous cell carcinoma cell epithelial-mesenchymal transition by targeting miR-153-3p. Epigenomics. 2020 Mar;12(6): 487-505.

Kim D, Langmead B, Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4): 357-60.

Kim YJ, Kang MJ, Kim E, Kweon TH, Park YS, Ji S, Yang WH, Yi EC, Cho JW. O-GlcNAc stabilizes SMAD4 by inhibiting GSK-3beta-mediated proteasomal degradation. Sci Rep. 2020 Nov 16;10(1): 19908.

Lerchner W, Latinkic BV, Remacle JE, Huylebroeck D, Smith JC. Region-specific activation of the Xenopus brachyury promoter involves active repression in ectoderm and endoderm: a study using transgenic frog embryos. Development. 2000 Jun;127(12): 2729-39.

Li H, Zhu Y, Morozov YM, Chen X, Page SC, Rannals MD, Maher BJ, Rakic P. Disruption of TCF4 regulatory networks leads to abnormal cortical development and mental disabilities. Mol Psychiatr7. 2019 Aug;24(8): 1235-46.

Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, Sunkin SM, Li Z, Shin Y, Zhu Y, Sousa AMM, Werling DM, Kitchen RR, Kang HJ, Pletikos M, Choi J, Muchnik S, Xu X, Wang D, Lorente-Galdos B, Liu S, Giusti-Rodríguez P, Won H, de Leeuw CA, Pardiñas AF; BrainSpan Consortium; PsychENCODE Consortium; PsychENCODE Developmental Subgroup, Hu M, Jin F, Li Y, Owen MJ, O'Donovan MC, Walters JTR, Posthuma D, Reimers MA, Levitt P, Weinberger DR, Hyde TM, Kleinman JE, Geschwind DH, Hawrylycz MJ, State MW, Sanders SJ, Sullivan PF, Gerstein MB, Lein ES, Knowles JA, Sestan N. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science. 2018 Dec 14;362(6420): eaat7615.

Lin X, Wang Y, Jiang Y, Xu M, Pang Q, Sun J, Yu Y, Shen Z, Lei R, Xu J. Sumoylation enhances the activity of the TGF-beta/SMAD and HIF-1 signaling pathways in keloids. Life Sci. 2020 Aug 15;255: 117859.

Liu S, Long J, Yuan B, Zheng M, Xiao M, Xu J, Lin X, Feng XH. SUMO Modification Reverses Inhibitory Effects of Smad Nuclear Interacting Protein-1 in TGF-beta Responses. J Biol Chem. 2016 Nov 18;291(47): 24418-30.

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

Lui NC, Tam WY, Gao C, Huang JD, Wang CC, Jiang L, Yung WH, Kwan KM. Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. Nat Commun. 2017 May 18;8: 15079.

McKinsey GL, Lindtner S, Trzcinski B, Visel A, Pennacchio LA, Huylebroeck D, Higashi Y, Rubenstein JL. Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. Neuron. 2013 Jan 9;77(1): 83-98.

Meert L, Birkhoff JC, Conidi A, Poot RA, Huylebroeck D. Different E-box binding transcription factors, similar neuro-developmental defects: ZEB2 (Mowat-Wilson syndrome) and TCF4 (Pitt-Hopkins syndrome). Rare Dis Orphan Drugs J. 2022;1:8.

Menuchin-Lasowski Y, Dagan B, Conidi A, Cohen-Gulkar M, David A, Ehrlich M, Giladi PO, Clark BS, Blackshaw S, Shapira K, Huylebroeck D, Henis YI, Ashery-Padan R. Zeb2 regulates the balance between retinal interneurons and Muller glia by inhibition of BMP-Smad signaling. Dev Biol. 2020 Dec 1;468(1-2): 80-92.

Mesman S, Bakker R, Smidt MP. Tcf4 is required for correct brain development during embryogenesis. Mol Cell Neurosci. 2020 Jul;106: 103502.

Miyoshi T, Maruhashi M, Van De Putte T, Kondoh H, Huylebroeck D, Higashi 7. Complementary expression pattern of Zfhx1 genes Sip1 and deltaEF1 in the mouse embryo and their genetic interaction revealed by compound mutants. Dev Dyn. 2006 Jul;235(7): 1941-52.

Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH, Aburatani H, Miyazono K. ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. Nucleic Acids Res. 2011 Nov 1;39(20): 8712-27.

Mowat DR, Croaker GD, Cass DT, Kerr BA, Chaitow J, Adès LC, Chia NL, Wilson MJ. Hirschsprung disease, microcephaly, mental retardation, and characteristic facial features: delineation of a new syndrome and identification of a locus at chromosome 2q22-q23. J Med Genet. 1998 Aug;35(8):617-23.

Mowat DR, Wilson MJ, Goossens M. Mowat-Wilson syndrome. J Med Genet. 2003 May;40(5):305-10.

Nash B, Meucci O. Functions of the chemokine receptor CXCR4 in the central nervous system and its regulation by mu-opioid receptors. Int Rev Neurobiol. 2014;118: 105-28.

Nelles L, Van de Putte T, van Grunsven L, Huylebroeck D, Verschueren K. Organization of the mouse Zfhx1b gene encoding the two-handed zinc finger repressor Smad-interacting protein-1. Genomics. 2003 Oct;82(4): 460-9.

Nourmohammadi B, Tafsiri E, Rahimi A, Nourmohammadi Z, Daneshvar Kakhaki A, Cho W, Karimipoor M. Expression of miR-9 and miR-200c, ZEB1, ZEB2 and E-cadherin in Non-Small Cell Lung Cancers in Iran. Asian Pac J Cancer Prev. 2019 Jun 1;20(6): 1633-9.

Pan Y, Zhang J, Fu H, Shen L. miR-144 functions as a tumor suppressor in breast cancer through inhibiting ZEB1/2-mediated epithelial mesenchymal transition process. Onco Targets Ther. 2016 Oct 11;9: 6247-55.

Papin C, van Grunsven LA, Verschueren K, Huylebroeck D, Smith JC. Dynamic regulation of Brachyury expression in the amphibian embryo by XSIP1. Mech Dev. 2002 Feb;111(1-2): 37-46.

Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI.The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020 Jan 8;48(D1): D845-55.

Puig RR, Boddie P, Khan A, Castro-Mondragon JA, Mathelier A. UniBind: maps of high-confidence direct TF-DNA interactions across nine species. BMC Genomics. 2021 Jun 26;22(1):482.

Quintes S, Brinkmann BG, Ebert M, Fröb F, Kungl T, Arlt FA, Tarabykin V, Huylebroeck D, Meijer D, Suter U, Wegner M, Sereda MW, Nave KA. Zeb2 is essential for Schwann cell differentiation, myelination and nerve repair. Nat Neurosci. 2016 Aug;19(8):1050-9.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1): W160-5.

Rao C, Malaguti M, Mason JO, Lowell S. The transcription factor E2A drives neural differentiation in pluripotent cells. Development. 2020 Jun 22;147(12):dev184093.

Remacle JE, Kraft H, Lerchner W, Wuytens G, Collart C, Verschueren K, Smith JC, Huylebroeck D. New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. EMBO J. 1999 Sep 15;18(18): 5073-84.

Ricci E, Fetta A, Garavelli L, Caraffi S, Ivanovski I, Bonanni P, Accorsi P, Giordano L, Pantaleoni C, Romeo A, Arena A, Bonetti S, Boni A, Chiarello D, Di Pisa V, Epifanio R, Faravelli F, Finardi E, Fiumara A, Grioni D, Mammi I, Negrin S, Osanni E, Raviglione F, Rivieri F, Rizzi R, Savasta S, Tarani L, Zanotta N; Mowat Wilson Epilepsy Study Group, Dormi A, Vignoli A, Canevini M, Cordelli DM. Further delineation and long-term evolution of electroclinical phenotype in Mowat Wilson Syndrome. A longitudinal study in 40 individuals. Epilepsy Behav. 2021 Oct 4;124:108315.

Scott CL, Soen B, Martens L, Skrypek N, Saelens W, Taminau J, Blancke G, Van Isterdael G, Huylebroeck D, Haigh J, Saeys Y, Guilliams M, Lambrecht BN, Berx G. The transcription factor Zeb2 regulates development of conventional and plasmacytoid DCs by repressing Id2. J Exp Med. 2016 May 30;213(6): 897-911.

Sekido R, Murai K, Funahashi J, Kamachi Y, Fujisawa-Sehara A, Nabeshima Y, Kondoh H. The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. Mol Cell Biol. 1994 Sep;14(9): 5692-700.

Sepp M, Pruunsild P, Timmusk T. Pitt-Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. Hum Mol Genet. 2012 Jul 1;21(13):2873-88.

Seuntjens E, Nityanandam A, Miquelajauregui A, Debruyn J, Stryjewska A, Goebbels S, Nave KA, Huylebroeck D, Tarabykin V. Sip1 regulates sequential fate decisions by feedback signaling from postmitotic neurons to progenitors. Nat Neurosci. 2009 Nov;12(11): 1373-80.

Slawny NA, O'Shea KS. Dynamic changes in Wnt signaling are required for neuronal differentiation of mouse embryonic stem cells. Mol Cell Neurosci. 2011 Nov;48(3): 205-16.

Stryjewska A, Dries R, Pieters T, Verstappen G, Conidi A, Coddens K, Francis A, Umans L, van IWF, Berx G, van Grunsven LA, Grosveld FG, Goossens S, Haigh JJ, Huylebroeck D. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells. 2017 Mar;35(3): 611-25.

Stumm RK, Zhou C, Ara T, Lazarini F, Dubois-Dalcq M, Nagasawa T, Hollt V, Schulz S. CXCR4 regulates interneuron migration in the developing neocortex. J Neurosci. 2003 Jun 15;23(12): 5123-30.

Suh H, Consiglio A, Ray J, Sawai T, D'Amour KA, Gage FH. *In vivo* fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. Cell Stem Cell. 2007 Nov;1(5):515-28.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1): D607-13.

Takagi T, Nishizaki Y, Matsui F, Wakamatsu N, Higashi 7. De novo inbred heterozygous Zeb2/Sip1 mutant mice uniquely generated by germ-line conditional knockout exhibit craniofacial, callosal and behavioral defects associated with Mowat-Wilson syndrome. Hum Mol Genet. 2015 Nov 15;24(22): 6390-402.

Teixeira JR, Szeto RA, Carvalho VMA, Muotri AR, Papes F. Transcription factor 4 and its association with psychiatric disorders. Transl Psychiatr7. 2021 Jan 5;11(1):19.

Tzeng SF, de Vellis J. Id1, Id2, and Id3 gene expression in neural cells during development. Glia. 1998 Dec;24(4): 372-81.

Van de Putte T, Maruhashi M, Francis A, Nelles L, Kondoh H, Huylebroeck D, Higashi 7. Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. Am J Hum Genet. 2003 Feb;72(2): 465-70.

van den Berghe V, Stappers E, Vandesande B, Dimidschstein J, Kroes R, Francis A, Conidi A, Lesage F, Dries R, Cazzola S, Berx G, Kessaris N, Vanderhaeghen P, van IJcken W, Grosveld FG, Goossens S, Haigh JJ, Fishell G, Goffinet A, Aerts S, Huylebroeck D, Seuntjens E. Directed migration of cortical interneurons depends on the cell-autonomous action of Sip1. Neuron. 2013 Jan 9;77(1): 70-82.

van Grunsven LA, Papin C, Avalosse B, Opdecamp K, Huylebroeck D, Smith JC, Bellefroid EJ. XSIP1, a Xenopus zinc finger/homeodomain encoding gene highly expressed during early neural development. Mech Dev. 2000 Jun;94(1-2): 189-93.

van Grunsven LA, Taelman V, Michiels C, Opdecamp K, Huylebroeck D, Bellefroid EJ. deltaEF1 and SIP1 are differentially expressed and have overlapping activities during Xenopus embryogenesis. Dev Dyn. 2006 Jun;235(6): 1491-500.

van Helden MJ, Goossens S, Daussy C, Mathieu AL, Faure F, Marcais A, Vandamme N, Farla N, Mayol K, Viel S, Degouve S, Debien E, Seuntjens E, Conidi A, Chaix J, Mangeot P, de Bernard S, Buffat L, Haigh JJ, Huylebroeck D, Lambrecht BN, Berx G, Walzer T. Terminal NK cell maturation is controlled by concerted actions of T-bet and Zeb2 and is essential for melanoma rejection. J Exp Med. 2015 Nov 16;212(12): 2015-25.

Vandamme N, Denecker G, Bruneel K, Blancke G, Akay O, Taminau J, De Coninck J, De Smedt E, Skrypek N, Van Loocke W, Wouters J, Nittner D, Köhler C, Darling DS, Cheng PF, Raaijmakers MIG, Levesque MP, Mallya UG, Rafferty M, Balint B, Gallagher WM, Brochez L, Huylebroeck D, Haigh JJ, Andries V, Rambow F, Van Vlierberghe P, Goossens S, van den Oord JJ, Marine JC, Berx G. The EMT Transcription Factor ZEB2 Promotes Proliferation of Primary and Metastatic Melanoma While Suppressing an Invasive, Mesenchymal-Like Phenotype. Cancer Res. 2020 Jul 15;80(14): 2983-95.

Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P, Nelles L, Wuytens G, Su MT, Bodmer R, Smith JC, Huylebroeck D. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. J Biol Chem. 1999 Jul 16;274(29): 20489-98.

Verstappen G, van Grunsven LA, Michiels C, Van de Putte T, Souopgui J, Van Damme J, Bellefroid E, Vandekerckhove J, Huylebroeck D. Atypical Mowat-Wilson patient confirms the importance of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. Hum Mol Genet. 2008 Apr 15;17(8): 1175-83.

Wahab NA, Othman Z, Nasri NWM, Mokhtar MH, Ibrahim SF, Hamid AA, Raja Ali RA, Mokhtar NM. Inhibition of miR-141 and miR-200a Increase DLC-1 and ZEB2 Expression, Enhance Migration and Invasion in Metastatic Serous Ovarian Cancer. Int J Environ Res Public Health. 2020 Apr 17;17(8): 2766.

Wakamatsu N, Yamada Y, Yamada K, Ono T, Nomura N, Taniguchi H, Kitoh H, Mutoh N, Yamanaka T, Mushiake K, Kato K, Sonta S, Nagaya M. Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. Nat Genet. 2001 Apr;27(4):369-70.

Wang LH, Baker NE. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. Dev Cell. 2015 Nov 9;35(3):269-80.

Weng Q, Chen Y, Wang H, Xu X, Yang B, He Q, Shou W, Chen Y, Higashi Y, van den Berghe V, Seuntjens E, Kernie SG, Bukshpun P, Sherr EH, Huylebroeck D, Lu QR. Dual-mode modulation of Smad signaling by Smad-interacting protein Sip1 is required for myelination in the central nervous system. Neuron. 2012 Feb 23;73(4): 713-28.

Wu LM, Wang J, Conidi A, Zhao C, Wang H, Ford Z, Zhang L, Zweier C, Ayee BG, Maurel P, Zwijsen A, Chan JR, Jankowski MP, Huylebroeck D, Lu QR. Zeb2 recruits HDAC-NuRD to inhibit Notch and controls Schwann cell differentiation and remyelination. Nat Neurosci. 2016 Aug;19(8): 1060-72.

Xu P, Lin X, Feng XH. Posttranslational Regulation of Smads. Cold Spring Harb Perspect Biol. 2016 Dec 1;8(12): a022087.

Yamada K, Yamada Y, Nomura N, Miura K, Wakako R, Hayakawa C, Matsumoto A, Kumagai T, Yoshimura I, Miyazaki S, Kato K, Sonta S, Ono H, Yamanaka T, Nagaya M, Wakamatsu N. Nonsense and frameshift mutations in ZFHX1B, encoding Smad-interacting protein 1, cause a complex developmental disorder with a great variety of clinical features. Am J Hum Genet. 2001 Dec;69(6):1178-85.

Yang S, Toledo EM, Rosmaninho P, Peng C, Uhlen P, Castro DS, Arenas E. A Zeb2-miR-200c loop controls midbrain dopaminergic neuron neurogenesis and migration. Commun Biol. 2018 Jun 25;1: 75.

Yao H, Hou G, Wang QY, Xu WB, Zhao HQ, Xu YC. LncRNA SPRY4IT1 promotes progression of osteosarcoma by regulating ZEB1 and ZEB2 expression through sponging of miR101 activit7. Int J Oncol. 2020 Jan;56(1): 85-100.

Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol. 2020 Feb 21;21(1):45.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9): R137.

Zweier C, Horn D, Kraus C, Rauch A. Atypical ZFHX1B mutation associated with a mild Mowat-Wilson syndrome phenotype. Am J Med Genet A. 2006 Apr 15;140(8):869-72.

Zweier C, Thiel CT, Dufke A, Crow YJ, Meinecke P, Suri M, Ala-Mello S, Beemer F, Bernasconi S, Bianchi P, ier A, Devriendt K, Dimitrov B, Firth H, Gallagher RC, Garavelli L, Gillessen-Kaesbach G, Hudgins L, Kääriäinen H, Karstens S, Krantz I, Mannhardt A, Medne L, Mücke J, Kibaek M, Krogh LN, Peippo M, Rittinger O, Schulz S, Schelley SL, Temple IK, Dennis NR, Van der Knaap MS, Wheeler P, Yerushalmi B, Zenker M, Seidel H, Lachmeijer A, Prescott T, Kraus C, Lowry RB, Rauch A. Clinical and mutational spectrum of Mowat-Wilson syndrome. Eur J Med Genet. 2005 Apr-Jun;48(2): 97-111.

Annex to Chapter 7:

Zeb2 in mesodermal and endodermal differentiation

Annex 7.1. Background and rationale

This PhD thesis also contains teamwork aiming at studying the regulation of the *Zeb2* locus (**Chapter 6**) and the mapping of Zeb2 DNA-binding sites (**Chapter 7**) in cultured mouse ESC-derived NPCs. Neural differentiation of pluripotent stem cells is relevant to studies of actions of ZEB2 e.g., in human brain early development (Chng *et al.*, 2010; Hegarty *et al.*, 2015; Birkhoff *et al.*, 2020; for a review, *see* Birkhoff *et al.*, 2021). The ESCs used in these studies included an established ESC line in which one of the two *Zeb2* alleles is epitope-tagged to produce Zeb2-V5 protein (clone 2BE3, *see* **Chapter 7**, section 7.3.1), together with two other lines (one heterozygous, one homozygous mutant, indicated as *ΔP* lines) from which the *Zeb2* promoter-proximal Zeb2 binding-site (located ~232 bp upstream of the TSS), which normally acts autoregulatory and positively in neural differentiation of wild-type ESCs, has been deleted (*see* **Chapter 7**, section 7.3.6). In addition, using homozygous *Zeb2*-KO mouse ESCs, we have previously shown that intact *Zeb2* is needed for their neural differentiation, but also general differentiation in embryoid bodies, thus including mesodermal and endodermal cells (Stryjewska *et al.*, 2017).

Efficient neural differentiation of ESCs is achieved by preventing BMP activities and/or signaling, whereas co-stimulation of wild-type mouse ESCs with BMPs and Nodal ligands (the Nodal-related Activin is used, also here, for such experiments) induce embryonic mesoderm, including primitive streak in it, allowing to obtain ESC-derived meso/endodermal (ME) cells in cell culture rapidly (around 3-4 days), and fairly efficiently (>80% and often >90% of the cells being ME cells).

In a late phase of this PhD research, we therefore still started to (i) optimize such ME differentiation of wild-type ESCs, and (ii) check ME cell formation by the *Zeb2-V5* and *ΔP* ESC lines. This would open roads to (i) repeat Zeb2-V5 ChIP-seq, but now in ME cells, and (ii) document whether the *ΔP* sequence is needed for efficient BMP+Activin stimulated ME differentiation and, if so, whether – e.g., as determined by RNA-seq – in one or both types of *ΔP* cells the normal expression of direct Zeb2 target genes or sets thereof would deviate from control, wild-type cells, like our team analyzed for neural differentiation (*see* **Chapter 7**).

Altogether, this would then document in these ME cells the Zeb2 DNA-binding sites in the presence of activated Smads, with the extra possibility to relate these sites to cognate Smad-responsive elements and/or mapped (by ChIP-seq, by others in the field) Smad-binding genes. In longer term, this would open other perspectives to genome-edit ESCs. For example, an interesting project would be to delete or more subtly mutate (i.e. 4 amino acids in) the short Smad-binding domain (SBD; Verschueren *et al.*, 1999; Conidi *et al.*, 2013) of Zeb2, leaving the rest of the protein intact, and repeat the same experiments. Such project would be the decisive step towards the unprecedented identification of intact-domain (in this case the SBD) dependent genes in the multi-domain TF Zeb2, and at the same time document Zeb2 as a direct Smad-interacting TF on the transcriptome, in this case of TGFβ family ligand-stimulated ESCs, and differentiating cells, altogether (Verschueren *et al.*, 1999; *see* also Birkhoff *et al.*, 2021).

Annex 7.2. Results and brief discussion

Annex 7.2.1. Differentiation of CGR8 ESCs (129 mouse strain) into ME cells

Various published protocols were tested and one was selected for further use (Kishimoto *et al.*, 2020), based on efficiency of differentiation and highest reproducibility (Korporaal, *data not shown*). The protocol of Kishimoto *et al.* (2020) was originally used to study bidirectional signaling (involving Wnt and BMP) between endoderm and mesoderm, which confers tracheal identity during development and could, in addition to *in vivo* studies in mice, also be studied by Kishimoto and co-workers in co-cultures of mouse as well as human ESC-derived mesodermal and endodermal cells.

For mouse ESCs their protocol starts by driving these cells into EpiLSC formation by treating undifferentiated ESCs with Activin-A and FGF2 between days 0-2, followed by continued treatment of the cells with to these two factors between days (D) 2-4, but now also adding BMP4 and XAV939. XAV939 inhibits Wnt/β-catenin-mediated transcription through tankyrase1/ 2 inhibition, thereby regulating Axin levels, and does not affect other pathways, including the TGFβ/BMP pathway. This induces mid-primitive streak like (mid-PS) cells by the end of D4, and these can then be further differentiated, using further adapted (always BMP-containing) medium. Kishimoto *et al.* (2020) obtained this way lateral plate mesoderm (LPM) like cells (at day 5) and eventually tracheal mesoderm (from day 5 onwards, up till day 12, where chondrocytes and smooth muscle cells can also be detected). We exploited the first steps in the Kishimoto *et al.* (2020) protocol, and optimized it for wild-type CGR8 ESCs and our purposes, i.e. with earlier read-outs, shortly after *Zeb2* is induced and strongly upregulated (Stryjewska *et al.*, 2017). The testing included (i) at least 2 independent clones or subclones (for 2BE3 cells), and (ii) different concentrations of Activin-A (from 0 to 50 ng/ml) and/or BMP4 (from 0 to 100 ng/ml) (*data not shown*). This yielded an efficient and nicely reproducible, as well as experimenter-independent protocol for obtaining ME cells (**Fig. A7.1**; *see* also Experimental Procedures).

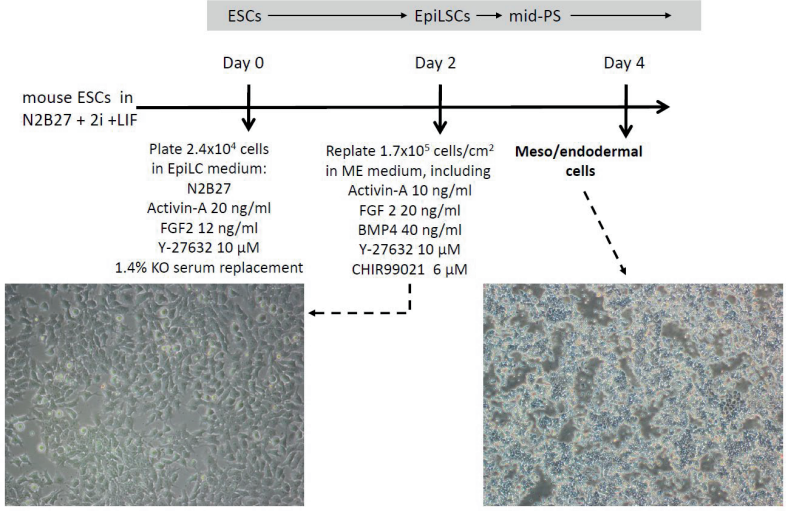


Figure A7.1. Overview of ESC→EpiLSC→meso/endodermal cell differentiation in cell culture.
For more details on the used media and specific additives, *see* below, Experimental procedures.

Differentiation of the wild-type CGR8 ESCs at D0, D3 and D4 was, in addition to brightfield imaging, for each experiment and tested ESC clone documented by RT-qPCR, mainly for acknowledged pluripotency, differentiation and germ layer marker genes, e.g., *Klf4*, *Otx2* and *Fgf5*, the *Mixl* gene *Mixl1* as well as *T*, *Sox17*, and *Zeb2* itself (a typical example is given in **Fig. A7.2**). We also added indirect immunofluorescence staining for marker proteins, e.g., Sox2, Eomes, T, Cdx2, Sox17 and Zeb2 (*data not shown*), identical to previous analyses (**Chapters 6** and **7**). For initial experiments, also the viability of the cells was assessed at D3 and D4. This was done by specific staining followed by FACS of the viable (always >81, and routinely 92%), Annexin-V+ (apoptotic; 1.1-1.6%), PI+ (necrotic, 0.2-1.6%) and double+ (late apoptotic; 2-6.9%) cells, respectively (*data not shown*; *see* Experimental procedures). Taken together, RT-qPCR confirmed differentiation to ME cells, while staining (*data not shown*) of the technical and biological replicates confirmed this too and furthermore showed low presence (routinely <5%) of neural cells.

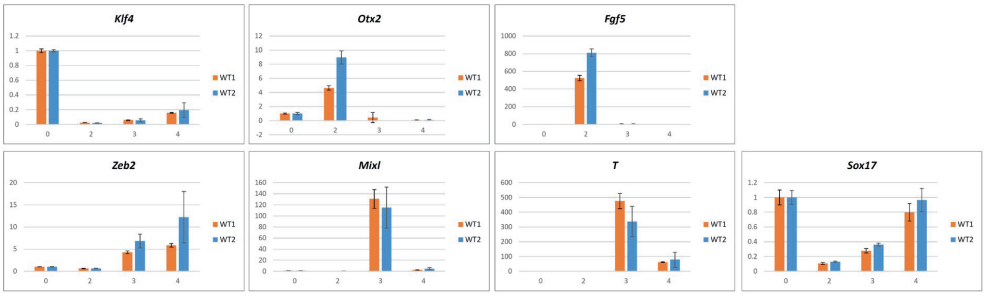


Figure A7.2. RT-qPCR analysis of wild-type (WT) CGR8 ESCs submitted to ME differentiation.

Two independent clones (WT1 and WT2) were submitted to the ME differentiation (*see* **Fig. A7.1**). In the typical experiment shown, 3 technical replicates for each WT cell line and each time point (D0, D2, D3 and D4, respectively) were used, and the averages calculated and compared to D0 levels after taking the latter as =1. Such entire experiments were repeated at least twice (biological replicates). For further details, *see* Experimental procedures.

7.2.2. The autoregulatory site in the *Zeb2* promoter-proximal region determines Zeb2 levels in ME differentiation of ESCs

During ME differentiation of mouse ESCs, and like in embryoid bodies wherein general differentiation was obtained, and in neural differentiation of ESCs (Stryjewska *et al.*, 2017; Birkhoff *et al.*, 2020; *see* also **Chapters 6** and **7**), *Zeb2* was found rapidly and strongly upregulated. This enabled us to test if its autoregulatory site was also critical for ME differentiation, like in neural differentiation. For this, we compared ME differentiation of wild-type and *ΔP* mutant cells (*see* **Chapter 7**), i.e. the E9 (homozygous for *ΔP*) and G2 line (heterozygous for *ΔP*), respectively. The respective steady-state *Zeb2* mRNA levels following the induction of differentiation indicated (**Fig. A7.3**) that the levels at D3 and D4 are critically dependent on the presence of the autoregulatory site, which is striking for line E9 ($Zeb2^{\Delta P/\Delta P}$). In the G2 ($Zeb2^{+/\Delta P}$) cells, Zeb2 levels were reduced, reaching intermediate levels between WT and E9 cells, which is indicative of the in-*cis* positive regulatory nature of the Zeb2 binding site just upstream of the *Zeb2* TSS. In both types of *ΔP* cells, in particular the expression of mesodermal (*Mixl1, T*) marker genes remained dramatically low from D3 onwards, certainly in the E9 clone, while this reduction is also

seen – but was less dramatic – for the endodermal (*Sox17*) marker gene, which in any case fails to be upregulated from D4 in both types of *ΔP* clone.

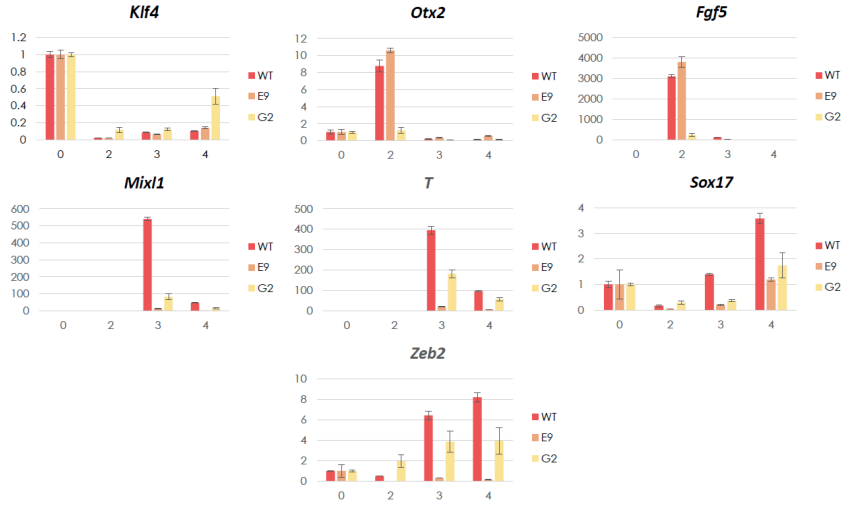


Figure A7.3. RT-qPCR analysis of wild-type (WT) and *ΔP* mutant ESCs submitted to ME differentiation.

WT, E9 (*Zeb2*$^{ΔP/ΔP}$) and G2 (*Zeb2*$^{+/ΔP}$) cells were submitted to ME differentiation (*see* **Fig. A7.1**) and the expression of a set of relevant marker genes verified. In all experiments, including the one shown, 3 technical replicates for each cell line and each time point (D0, D2, D3 and D4, respectively) were used, the averages calculated, and compared to D0 levels after taking the latter as =1. For further details, *see* Experimental procedures.

Interestingly, but this requests further investigation, *Zeb2* transcripts in the G2 (*Zeb2*$^{+/ΔP}$) cells were already detectable at D2, earlier than in wild-type cells, albeit at low level. This may suggest that the region encompassing the autoregulatory site, prior to *Zeb2* upregulation and the positive autoregulation, is also and preceding the upregulation a target for repressive regulation of *Zeb2*.

This observation of earlier detection of *Zeb2* also coincides with dramatically low levels of *Fgf5* at that same time point and this, remarkably, exclusively in this G2 *ΔP*-line, although these cells still formed some and in any case more ME cells than the E9 *ΔP*-line. The same was true for *Otx2* in the G2 line, whereas Klf4 seems normally downregulated in all cell lines. It cannot be excluded that we might have missed *Fgf5* and *Otx2* induction, certainly in case of faster induction followed then by a very fast downregulation. In addition, but the effect is moderate, the E9 *Zeb2*$^{ΔP/ΔP}$ cells tend even to express higher steady-state levels of *Otx2* and *Fgf5* than wild-type cells.

FGF2 is added to our cultures and together with Activin-A maintains primed pluripotency. At least in early anterior neurectoderm *in vivo*, Fgf2 promotes via upregulation of *Zeb2* (neural) stem cell formation (Dang and Tropepe, 2010), but is this is likely not the mechanism that operates here in ME formation in FGF2-supported cell culture. Otx2 is known to be required for mounting appropriate responses to FGF2, transition from naïve to primed pluripotency and thus progression to a stable EpiLSC state (Acampora *et al.*, 2013), direct repression of *Nanog* and *Oct4* (Acampora *et al.*, 2016; di Giovannantonio *et al.*, 2021) and creating functional antagonism with

Nanog, which co-leads to heterogeneity in ESCs (Acampora *et al.*, 2017), and is also required for ESC differentiation. Hence, while the *ΔP/ΔP* mutant cells clearly indicate the relevance of the intact autoregulatory site for reaching proper Zeb2 levels for Zeb2-dependent neural (*see* **Chapter 7**) and ME differentiation of ESCs, the situation of the *Zeb2*$^{+/\Delta P}$ cells presents as more complex. Furthermore, different compensatory responses to inappropriate Zeb2 mRNA/protein levels might be mounted in these cells. These preliminary results in the *Zeb2*$^{\Delta P/\Delta P}$ in the first place, but also the *Zeb2*$^{+/\Delta P}$ cells, call for future RNA-seq studies in these cells, perhaps at more time points intervals, and possibly involving a D1.5 time point as well, in ME differentiation (i.e. in the presence of BMP/Activin-A). It will be interesting to compare this transcriptomic profiling at different time points of differentiation, even in human *ΔP* ESCs and control iPSCs with wild-type hESCs and iPSCs from MOWS patients, including next-generation mutant iPSCs in e.g., the gene desert located enhancers of *ZEB2* (Birkhoff *et al.*, 2020; *see* **Chapter 6**).

7.2.3. Zeb2-V5 tag ESCs differentiate into ME cells, making Zeb2 ChIP-seq analysis possible in the presence of Activin-A/BMP

One of the aims of the team's research is to map genome-wide binding sites for Zeb2 in TGFβ family ligand-stimulated cells. **Fig. A7.4** shows that Zeb2-V5 tag CGR8 ESCs (clone 2BE3, *see* **Chapter 7**) are capable of ME cell formation. Furthermore, these Zeb2-tag cells reach also in this ME differentiation protocol Zeb2 protein levels (indirect immunofluorescence staining and western blot analysis; *data not shown*) that will enable ChIP-seq approaches in upscaled cultures.

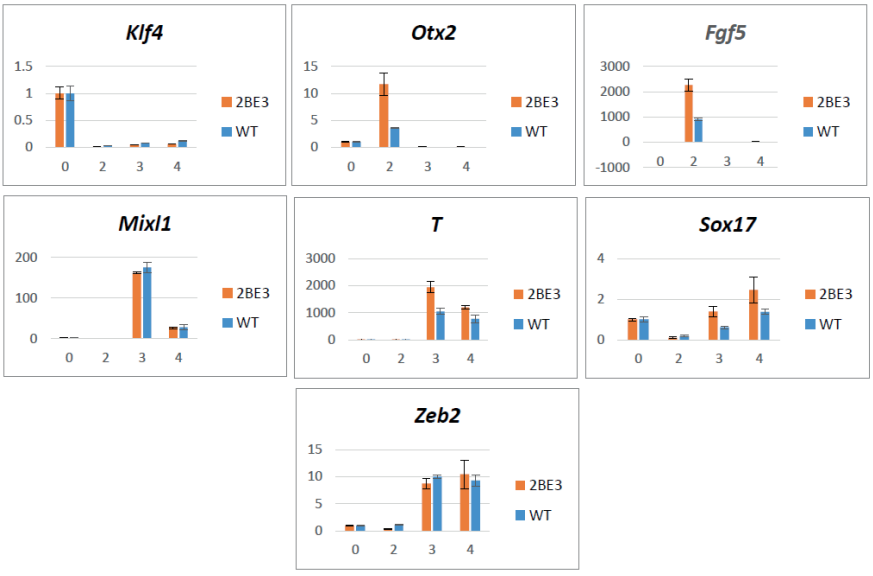


Figure A7.4. RT-qPCR analysis of wild-type (WT) and Zeb2-V5 tag ESCs submitted to ME differentiation.

WT and 2BE3 cells (*see* main part of **Chapter 7** and its Experimental procedures) were submitted to ME differentiation (*see* **Fig. A7.1**) and again the expression of the routinely used set of marker genes verified. In all experiments, including the one shown, 3 technical replicates for each cell line and each time point (D0, D2, D3 and D4, respectively) were used, the averages calculated, and compared to D0 levels after taking the latter as =1. For further details, *see* Experimental procedures.

Annex 7.3. Experimental procedures

ESC cultures

CGR8 (strain 129) mouse ESCs, i.e. the wild-type, Zeb2-V5 and *Zeb2-ΔP* mutant cells are described elsewhere (**Chapters 6** and **7**). They were maintained on a gelatin-coated (0.1%) dish in ESC-medium according to Bibel *et al.* (2007). The media consisted of DMEM supplemented with 15% heat-inactivated FBS, 2 mM L-Glutamine, 1x Non-Essential Amino Acids (MEM-NEAA), 143 µM β-mercapto-ethanol (all ThermoFisher Scientific) and LIF (10^3 U/ml). The cells were passaged every other day with using trypsin-EDTA.

ESC differentiation to ME cells

ESCs were differentiated towards ME cells according to Kishimoto *et al.* (2020) with modifications (*see* also **Fig. A7.1**). In brief, 3 days prior to the start of the differentiation the aforementioned ESC-medium was changed to N2B27+2i+LIF medium, which is a 1:1 DMEM/F12 and Neuralbasal medium, 0.5x B-27 supplement, 0.5x N-2 supplement, 0.5x Non-Essential Amino Acids (NEAA), 0.5x sodium pyruvate, 50 µM β-mercapto-ethanol, 3 µM CHIR99021, 1.2 µM PD325901 (all ThermoFisher Scientific) and LIF (10^3 U/ml).

On D0, $2.4x10^4$ cells/cm^2 were plated on a Geltrex (Gibco)-coated plate in EpiLC medium consisting of N2B27 supplemented with 20 ng recombinant Activin-A/ml (R&D systems), 12 ng GFG2/ml (R&D systems), 10 µM Y-27632 (STEMCELL) and 1% KnockOut (KO) serum replacement (ThermoFisher Sientific). After 48 h, the cells were replated at the density of 1.7 $x10^5$ cells/cm^2 in ME medium consisting of DMEM, 1x MEM-NEAA, 1x sodium pyruvate, 2mM L-Glutamine, 2% B27-supplement, 10 ng recombinant Activin A/ml, 20ng FGF2/ml, 40 ng BMP4/ml (R&D systems), 6 µM CHIR99021 and 10 µM Y-27632, and cultured for another 2 days.

RNA extraction and RT-qPCR analysis

Total RNA was isolated from the CGR8 ESCs at different time points, using TRI reagent (Sigma). Thereafter cDNA was synthesized using RevertAid First Strand cDNA synthesis kit (ThermoFisher Scientific), and RT-qPCR was performed on a CFX96 T1000 thermal cycler (Bio-Rad) using SybrGreen dye (Bio-Rad) (*see* also **Chapters 6** and **7**). Data consists of averages of biological replicates and technical replicates (all at least n=3), normalized to *β-actin*. The used primers are:

Primer	Sequence 5' - 3'
β-actin_Fwd	TCTCCTTCTGCATCCTGTCAGCAA
β-actin_Rev	TCTTGGGTATGGAATCCTGTGGCA
Klf4_Fwd	CTGGCGAGTCTGACATGG
Klf4_Rev	CTCACGCCAACGGTTAGTC
Otx2_Fwd	GCGCTTTCTCAGCAAATCT
Otx2_Rev	TCCAAATAGCCAGCTATCAAAG
Fgf5_Fwd	CGAAGCCAGTGTGTTAAGTAT
Fgf5_Rev	TTGTTGCTGAAAACTCCTCG
mMixl1_Fwd	ATGTACCCAGACATCCACTT
mMixl1_Rev	TGAAATGACTTCCCACTCTG
mT_Fwd	CTCGGATTCACATCGTGAGAG
mT_Rev	AAGGCTTTAGCAAATGGGTTGTA
mSox17_Fwd	CGAGCCAAAGCGGAGTCTC
mSox17_Rev	TGCCAAGGTCAACGCCTTC
Sip1_Fwd	CAATGCAGCACTTAGGTGTA
Sip1_Rev	TTGCCTAGAAACCGTATTGT
mEomes Fwd	AACCTTCCAAGACTCAGACC
mEomes Rev	TCTGATGGGATGAATCGTAGT
mGsc Fwd	AGACAGTCGATGCTACTTGC
mGsc Rev	AGTCCTGGGCCTGTACATTA

Pax6_Fwd	ACATCTTTTACCCAAGAGCA
Pax6_Rev	GGCAAACACATCTGGATAAT
mGata6_Fwd	GGGGTAGGGGCATCAG
mGata6_Rev	CCGTCTTGACCTGAATACTT
mFoxa2_Fwd	AACATGAACTCGATGAGCC
mFoxa2_Rev	ATGTACGAGTAGGGAGGTTT

Indirect immunofluorescence

ESCs were differentiated on coverslips and fixed for 15 min with 4% PFA on D0, D3 and D4 (*see* also **Chapter 6** and **7**). Then, the cells were permeabilized with flash addition of cold methanol (kept at -20°C) and extra 10 min incubation at -20°C, and then switched for 1 h at room temperature (24°C) in blocking buffer (3% bovine serum albumin, BSA (Roche) + 0.3% TritonX-100). Thereafter the cells were incubated overnight at 4°C with the primary antibodies in antibody dilution buffer (3% BSA + 0.3% TritonX-100).

The next day the cells were washed 3 times with PBS. Then the cells were incubated with the corresponding secondary antibodies in antibody dilution buffer for at least 1.5 h, in the dark. Next the cells were washed 3 times with PBS. Then the slides were mounted with Mowiol (Sigma-Aldrich) containing DAPI (1:1000, Sigma-Aldrich) and dried overnight. Images were made using the Leica SP5 confocal microscope.

Antibody	Species	Brand	Cat.	Used at dilution
Sox2	Goat	Immune systems	GT15098	1:400
Cdx2	Mouse	Biogenex	MU392A-5UC	1:100
Eomes	Rabbit	Abcam	ab23345	1:100
Sox17	Goat	R & D systems	AF1924	1:500
Zeb2	Rabbit	Santa-Cruz	Sc-48789	1:100
Goat 594	Donkey	Jackson	705-585-147	1:500
Mouse 488	Donkey	Invitrogen	A-11029	1:500
Rabbit Cy5	Donkey	Jackson	711-175-152	1:500

Apoptotic assay

In order to quantify the level of apoptotis and necrotis during differentiation, the cells were stained with Annexin-V and propidium iodide (PI) according to the manufacturer's protocol (ab214484, abcam). In short, at D0, D3 and D4 10^6 cells were collected and stained with 5 µl of Annexin-V and/or PI in 100 µl 1xbinding buffer for 15 min at room temperature in the dark. Thereafter 400 µl of 1x binding buffer was added. The samples were analyzed using the BD LSRFortessa flow cytometer (BD LSRFortessa).

Protein isolation, BCA and western blot

To check the levels of Zeb2 during ME differentiation of Zeb2-Flag-V5 mESCs we extracted (mostly at D3) nuclear and cytoplasmic proteins, using the NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (Thermo-Fisher Scientific) according to the manufacturer's protocol. The protein concentration was measured by BCA assay (ThermoFisher Scientific). For SDS-PAGE 70 µg of nuclear and cytoplasmic protein were loaded on a 6% Tris-Glycine SDS-PAGE gel. As a positive control, nuclear and cytoplasmic protein from HEK293 cells overproducing Zeb2V5 from a transfected expression vector (*see* **Chapter 7**) was used. The proteins were then transferred from the gel onto nitrocellulose membranes (Amersham Bioscience). These membranes were first blocked in 5% BSA in TBS-T blocking buffer and then incubated overnight with the primary antibody. The next day the blots were washed and incubated with the corresponding HRP-conjugated secondary antibody. The following antibodies were used:

Antibody	Species	Brand	Cat.	Used at dilution
Zeb2	Rabbit	Seuntjens *et al.*, 2009		
V5	Mouse	Life Technologies	R960-25	
Flag	Mouse	Sigma	F1804	
Rabbit-HRP	Goat	Jackson ImmunoResearch	111-035-045	1:10,000
Mouse-HRP	Goat	Jackson ImmunoResearch	115-035-003	1:10,000

Literature references for this annex

Acampora D, Di Giovannantonio LG, Garofalo A, Nigro V, Omodei D, Lombardi A, Zhang J, Chambers I, Simeone A. Functional Antagonism between OTX2 and NANOG Specifies a Spectrum of Heterogeneous Identities in Embryonic Stem Cells. Stem Cell Reports. 2017 Nov 14;9(5):1642-59.

Acampora D, Di Giovannantonio LG, Simeone A. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. Development. 2013 Jan 1;140(1):43-55.

Acampora D, Omodei D, Petrosino G, Garofalo A, Savarese M, Nigro V, Di Giovannantonio LG, Mercadante V, Simeone A. Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. Cell Rep. 2016 Jun 21;15(12):2651-64.

Bibel M, Richter J, Lacroix E, Barde YA. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc. 2007;2(5):1034-43.

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, Nowosad K, van den Hout MCGN, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. Hum Mol Genet. 2020 Aug 29;29(15):2535-50.

Birkhoff JC, Huylebroeck D, Conidi A. ZEB2, the Mowat-Wilson Syndrome Transcription Factor: Confirmations, Novel Functions, and Continuing Surprises. Genes (Basel). 2021 Jul 3;12(7):1037.

Chng Z, Teo A, Pedersen RA, Vallier L. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. Cell Stem Cell. 2010 Jan 8;6(1):59-70.

Conidi A, van den Berghe V, Leslie K, Stryjewska A, Xue H, Chen YG, Seuntjens E, Huylebroeck D. Four amino acids within a tandem QxVx repeat in a predicted extended α-helix of the Smad-binding domain of Sip1 are necessary for binding to activated Smad proteins. PLoS One. 2013 Oct 11;8(10):e76733.

Dang LT, Tropepe V. FGF dependent regulation of Zfhx1b gene expression promotes the formation of definitive neural stem cells in the mouse anterior neurectoderm. Neural Dev. 2010 May 6;5:13.

Di Giovannantonio LG, Acampora D, Omodei D, Nigro V, Barba P, Barbieri E, Chambers I, Simeone A. Direct repression of Nanog and Oct4 by OTX2 modulates the contribution of epiblast-derived cells to germline and somatic lineage. Development. 2021 May 15;148(10):dev199166.

Hegarty SV, Sullivan AM, O'Keeffe GW. Zeb2: A multifunctional regulator of nervous system development. Prog Neurobiol. 2015 Sep;132:81-95.

Kishimoto K, Furukawa KT, Luz-Madrigal A, Yamaoka A, Matsuoka C, Habu M, Alev C, Zorn AM, Morimoto M. Bidirectional Wnt signaling between endoderm and mesoderm confers tracheal identity in mouse and human cells. Nat Commun. 2020 Aug 27;11(1):4159.

Seuntjens E, Nityanandam A, Miquelajauregui A, Debruyn J, Stryjewska A, Goebbels S, Nave KA, Huylebroeck D, Tarabykin V. Sip1 regulates sequential fate decisions by feedback signaling from postmitotic neurons to progenitors. Nat Neurosci. 2009 Nov;12(11): 1373-80.

Stryjewska A, Dries R, Pieters T, Verstappen G, Conidi A, Coddens K, Francis A, Umans L, van IJcken WF, Berx G, van Grunsven LA, Grosveld FG, Goossens S, Haigh JJ, Huylebroeck D. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells. 2017 Mar;35(3):611-25.

Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P, Nelles L, Wuytens G, Su MT, Bodmer R, Smith JC, Huylebroeck D. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. J Biol Chem. 1999 Jul 16;274(29):20489-98.

Chapter 8

General discussion

8.1. Background

Studies of gene transcription and its dynamic changes in vertebrate embryos or embryonic cells, resulting of exposure of these cells to polypeptide growth/differentiation factors, have always been a research focus of (both) Tylzanowski teams at MUL (from Lublin, also from the Leuven base) and (both) Huylebroeck teams at Erasmus MC (from Rotterdam, also from the Leuven base). The global aim of the research of both these entire teams is to understand the molecular mechanisms that control cell fate changes, with focus on how signaling by ligands, their receptors and downstream effectors, including TFs, achieve this in TGFβ/BMP and Wnt family signaling and their regulatory networks.

To better characterize aforementioned mechanisms both teams used loss-of-function approaches in mouse and zebrafish embryos, targeting genes encoding individual components of the TGFβ/BMP (Conidi *et al.,* 2011; Stryjewska *et al.,* 2017; Peeters *et al.,* 2018; Dries *et al.,* 2020; Birkhoff *et al.*, 2021) or Wnt systems (Lana-Elola *et al.*, 2011; Mommaerts *et al.*, 2014). Such experimental work then precedes additional studies, often using cultured cells, aiming at elucidating the mechanism underlying the primary phenotype(s), and also further investigating the role of such components of TGFβ/BMP or Wnt signaling (Peeters *et al.,* 2018; Birkhoff *et al.* 2021). For instance, pluripotent stem cell cultures (e.g., human and mouse ESCs, human iPSCs) submitted to differentiation protocols nowadays provide excellent alternative approaches to *in vivo* studies (Hong and Do, 2019).

Additionally to investigation of the functional role of TGFβ/BMP or Wnt components via loss-of-function approaches, future studies would ideally include characterization of the mechanisms orchestrating the spatio-temporal expression of such components, as for instance, the investigation of the dynamic changes of 3D chromatin architecture, and also the identification and functional characterization of promoter-proximal and distal enhancer(s). Moreover, in the case of TFs, identification of its/their binding sites, which overlap with promoter/enhancer(s), would provide better understanding of the regulation of its target gene(s).

This PhD research, by combining experimental work in embryos (here with focus on synovial joint formation in limb development) and cultured cells (mainly on ESC-derived NPCs, and starting to use BMP+Activin-stimulated mesendodermal differentiation) investigate the role of 3D chromatin structure and *cis*-REs (enhancers) in genomic regions encoding modulators of Wnt and/or BMP signaling (*Dact2*, *Smoc2*), and also the DNA-binding TF *Zeb2*, which is known to act in part through binding to activated Smads in TGFβ/BMP family signaling (Birkhoff *et al.*, 2021) and of course *trans*-REs (for it is a DNA-binding TF). Additionally, this PhD research focuses on the genome-wide investigation of the regulation of gene expression via characterization of the candidate enhancers during synovial joint formation and identification of *Zeb2* binding sites during neuronal differentiation of ESCs. Taking into consideration the use of different models to study mechanisms that orchestrate the regulation of gene expression, the discussion here is divided into two sections, with focus on the *in vivo* (section 8.2) and *in vitro* models (section 8.3). Further, section 8.4 summarizes the main findings related to the regulation of transcription and 3D chromatin structure in development and cell differentiation, and also provide a selection of future perspectives.

8.2. Candidate enhancers in synovial joint formation

8.2.1. The results of this research

As discussed in **Chapter 1**, promoters and enhancers, together with TFs, orchestrate spatio-temporal gene expression and subsequently control cell fate during embryogenesis. Therefore, comprehensive characterization of enhancer activity and associated epigenetic signatures (e.g., histone H3 modifications) is crucial for deciphering the regulatory mechanisms governing tissue patterning and principles of embryonic organogenesis, including in limb development and synovial joint formation. Recently, Cheung *et al.* (2020), using ChIP-seq for mapping enhancer-associated histone modifications (H3K27ac and H3K4me1), identified a set of active enhancers crucial for establishment of cell-type specificity during chondrogenesis. This prompted us to generate a genome-wide candidate enhancer (CE) atlas of cells, which were carefully removed by dissection from two regions of the distal part of the developing limb: (1) the interzone region, containing progenitor cells of future synovial joint structures, and (2) the adjacent phalange region, enriched in proliferating and pre-hypertrophic chondrocytes. In **Chapters 3 and 5**, we describe and use this manual micro-dissection. In **Chapter 3**, we characterized for the first time the CEs in developing synovial joint using ChIP-seq data (H3K27ac and H3K4me1), which we further correlated with the transcriptomes of interzone and phalange, respectively.

Various research showed that enhancer activity can be conserved among multiple vertebrate species (Yang *et al.*, 2015; Hirsch *et al.*, 2018; Yuan *et al.*, 2018). Thus, such genomic regulatory elements (REs) often co-steer biological processes that are common or even universal in different animals and phyla. Here, we narrowed down our studies of CEs to limb development in chicken embryos, historically used as a classical system to study organogenesis, including patterning and cell differentiation, in particular of skeletal tissue cells, and also extend our results to the mouse and human situation. These sets of experiments resulted in the identification of well-studied enhancers involved in regulation of genes important for interzone and chondrocyte identity, as well as novel CEs. The latter have also been linked to biological processes unique for interzone or phalange samples and to novel candidate marker, i.e., differentially expressed genes (DEGs). Further, the CEs have also been linked to genes associated with joint disorders and congenital phalange malformations, which supports the principle that deviation of normal gene expression levels at early developmental stages can cause limb defects. This is in line with a growing number of reports showing that structural variations (SVs) or SNPs, located in an enhancer, results in misexpression of the cognate gene(s), and can also lead to congenital limb defects abnormalities and beyond (*see* **Chapter 1b**; Nowosad *et al.,* 2020).

To further investigate the potential role of enhancer alterations in the molecular etiology of joint/phalange disorders, we utilized the GWAS dataset containing the studies that focus on SNPs in individuals with limb abnormalities and joint dysfunctions. Among all, this analysis led to identification of multiple CEs associated with osteoarthritis (OA), one of the most common joint disorders (Xia *et al.,* 2014). Many of these SNPs have been mapped to TF binding-motifs suggesting that modification of only one bp can lead to changes in TF binding, and/or the dynamics thereof, and subsequently result also in modification of enhancer activity. This can also manifest as congenital abnormality or increased susceptibility to OA, including its form caused most likely by alteration in homeostasis of the adult synovial joint.

The atlas of CEs active in interzone vs. adjacent phalange that was generated through this PhD research, is of high informative value. It serves not only the experimental biologist, who can utilize such atlas for further studies of gene regulation in these tissues. Also clinicians will benefit from it, for they can investigate potential genomic alteration in non-coding DNA in the context of enhancer activity, in particular in the field of congenital limb malformations.

One of the limitations of the ChIP-seq based enhancer atlas remains the further annotation of each CE or sets of CEs to potential target gene(s), which is often done based on genomic distance (McLean *et al.,* 2010). This approach leads to generation of multiple false positives, subsequently affecting downstream analysis. In order to reduce the error rate of enhancer - gene mapping the 3C technique and/or its derivatives can be applied to investigate the enhancer-promoter (E-P) connections (Golov *et al.,* 2020). However, one of the challenges when studying 3D chromatin architecture and E-P interactions in joint interzone is the low amount of cellular material collected from these sites of developing embryos. One interphalangeal interzone dissected from chicken hindlimb digit-3 (which we opted for in our work) contains around 10,000 cells. As an illustration, many 3C-based protocols require millions of cells as input (Belaghzal *et al.,* 2017). We dealt with this issue in two ways. First, we decided to use our in-house developed targeted chromatin capture protocol (T2C) and second, further optimized it for low-input (low-T2C) (*see* **Chapter 4**). For the purpose of protocol optimization, we used cultured cells, which also substantially reduced the time needed for sample collection.

We also decided to focus on one genomic region of interest, i.e., the locus encompassing *DACT2* and *SMOC2*. We were intrigued by the co-expression and neighboring chromosomal location of these two genes. Both are expressed in joint interzones and encode proteins that modulate Wnt signaling (Diez-Roux *et al.,* 2011; Sensiate *et al.,* 2014; Wang *et al.,* 2014; Lu *et al.,* 2019). We successfully obtained high-resolution T2C proximity maps. This T2C protocol together with our identification of CEs (by ChIP-seq) allowed us to deepen existing studies of regulation of *DACT2* and *SMOC2* in particular. In **Chapter 5**, we report for the first time the role of 3D chromatin architecture and enhancers in the expression of these genes during joint interzone formation and adjacent phalange development. For this, we analyzed a ~3.45 Mb-long chicken genomic region (chr3:40,15-43,6) encompassing *DACT2-SMOC2*. We have detected DNA-loops within their TADs, which were predominantly tissue-specific, in contrast to the TAD structure conserved between interzone and adjacent phalange. Further, we integrated T2C data with mapped H3 signatures (H3K27ac and H3K4me1), and identified seven *DACT2-SMOC2* CEs whose changes in chromatin states in interzone and phalange, respectively, were documented. Importantly, some of these changes correlate with mRNA expression and *cis*-proximities between enhancers and promoters (likely by DNA-looping). Further, we started to functionally validate CEs in developing zebrafish larvae. We were able to show that 4 out of 7 did present enhancer activity within the embryonic *DACT2*+ expression domains, suggesting that these enhancers can regulate *DACT2*.

8.2.2. Future perspectives

Application of ChIP-seq for investigation of histone marks (in particular of H3) revolutionized research projects aiming at identification of enhancers. However, characterization of CEs based only on histone signatures has limitations. First, peaks obtained from ChIP-seq experiments that use antibodies against histone marks (e.g., H4K4me1) are often broader than 1 kb, which does not allow to narrow down the CEs to core active regions. Second, the CEs defined as such cannot

be firmly annotated to a target gene (or target genes) yet. Third, and importantly, functional validation of CEs proves prone to error. A striking and recent example comes from the impressive work by the Pennacchio lab (Berkeley, CA, USA), and that calls for caution in this. Gorkin *et al.* (2020) reported the identification of 300,000 H3K27ac enhancers in 12 tissues and at 8 time points. That was an ideal starting point for their subsequent validation experiments. They first improved the high-throughput validation by site-directed integration in transgenic mice. For this, they adapted the Kothary *et al.* (1989a, 1989b) protocol. The end result is that they achieve a 50% transgenesis rate (and no longer maximally 12%), get higher reproducibility in the observed patterns and have no ectopic staining, and need less injections (Kvon *et al.*, 2020), meaning they could test many more candidate enhancers and even re-test significant numbers of previously acknowledged enhancers. Using this approach, 35 polydactyly mutations (of a well-characterized enhancer 1 Mb away from *SHH*) could be validated, many of which being re-validated. Remarkably, about 40% of these variations/mutations that were accepted pathogenic in the field thus far, turned out to have no effect in their functional assay, whereas 60% did.

In a next step, using VISTA for forebrain, midbrain, hindbrain, craniofacial, limb and heart, and integrating ChIP-seq, ATAC-seq and DNase-seq, and considering evolutionary conservation and using human genetic evidence, they then asked the question how often enhancers are missed by these current approaches or, vice versa, asked in how many cases they would find an enhancer *in vivo* that has no molecular mark. Pennacchio and co-workers reported on this work (at a recent conference, IMPC, Seoul, 2022) that 13% of the positively-detected sequences in the assay lack markers (H3K27Ac, H3K4me1, ATAC-seq signal), which prompted them to launch the term "hidden" enhancers.

This prompted them to do tiling analysis of a locus of interest and use their high-throughput system in an unbiased way, probing >1.5 Mb of mouse DNA. The stunning result was that 26% of the positive enhancers in this case had no signature! Of course, they then asked whether perhaps the biochemical data for those hidden enhancers are positive at earlier or later time points. That turned out to be the case, and explains about half of the hidden enhancers (so, this yields them again the aforementioned and estimated 13% of hidden enhancers). Last but not least, hidden enhancers are as conserved as non-hidden enhancers, and contain similar TF binding-motifs.

This raises questions on how to (1) improve the identification of enhancer core regions, (2) accurately annotate enhancers to target genes, and (3) identify CEs that are functionally active – in our case – in developing synovial joints. We have addressed some of these limitations when generating our CE atlas. For example, our exclusive selection of conserved regions allowed us to narrow down substantially the enhancer regions (**Chapter 3**), and we tested well-defined, stringently selected CEs of the *DACT2-SMOC2* region (**Chapter 5**). Further studies should be carried out to improve the accuracy of the existing atlas. The use of ATAC-seq would narrow down the enhancer coordinates to open chromatin (Bozek *et al.*, 2019), while the combination of ChIP-seq based CEs and massive parallel reporter assays (MPRAs, but after careful selection of the used cells) would provide genome-wide (or close to that) information on functionally active enhancers (Inoue and Ahituv, 2015). Further, integrating our proposed CEs with promoter capture-C type of experiments would allow for annotation of identified enhancers to all gene promoters and that are based on *cis*-proximities (Lu *et al.,* 2020).

Integration of transcriptome data from interzone/phalange with our CE atlas has provided us solid ground for further investigation of these DEGs in developing synovial joints, also in a more

clinical context. For instance, alterations in regulatory elements of *GFD5*, one of the most prominent marker genes, have been previously associated with increased susceptibility of OA (Capellini *et al.,* 2017). Therefore, it suggests that modifications of mRNA levels of genes expressed at onset of joint formation or in adult synovial joints may increase the probability of OA development. In **Chapter 3** we identified various DEGs of interzone and that have previously been linked to OA. Among all, the TF-encoding gene *ERG* factor was significantly upregulated in interzone as compared to phalange. Interestingly, Ochta *et al.* (2015) showed that conditional genetic inactivation of *ERG* leads to spontaneous degradation of synovial joints and OA-like phenotypic defects. This is linked to the role of *ERG* in endurance of articular cartilage in adult life. Also, *ERG* is upregulated in articular cartilage of patients with OA (most likely as an attempt to stabilize/repair cell phenotype), in contrast to less affected cartilage where *ERG* expression is barely detected. Taking into consideration our data from the developing synovial joint, it is tempting to suggest that *ERG* is one of the crucial TFs that controls both the development of the joint at the onset of interzone formation, and the maintenance of adult synovial joint. Therefore, a project aiming at detailed studies of *ERG* regulation and its role in joint development, including by functional characterization of the CEs described in **Chapter 3**, as well as genome-wide binding of ERG (a ChIP-seq approach being used in **Chapter 7**), would be very valuable.

8.3. On Zeb2 DNA-binding sites, including to *Zeb2* itself

8.3.1. The results of this research

The T2C protocol allowed us to investigate *ZEB2* regulation in PSC-derived neuroprogenitors. Importantly, *ZEB2* expression increases at the exit from primed pluripotency concomitant with start of cell differentiation (Stryjewska *et al.,* 2017). We hypothesized that the 3D chromatin landscape of the *ZEB2* locus (including its conserved ~3.5 Mb-long gene desert, *see* **Chapter 6**) changes during this differentiation, including reorganization of *cis*-proximities and E-P proximities. Gene deserts that flank developmentally significant genes often contain multiple enhancers, and we expected that these regulatory elements would control *ZEB2*. For the first time, we described the dynamics of 3D chromatin architecture of a 7.5 Mb-long region encompassing the entire *ZEB2* locus, including the aforementioned downstream gene desert, during neural differentiation of PSCs to NPCs. We observed enrichment of *cis*-proximities within the analyzed region after onset of differentiation, when *ZEB2* steady-state mRNA levels reach high level. We also identified three new enhancers of *ZEB2* located in the gene desert whose activities correlate nicely with changes in level of *ZEB2* mRNA. These enhancers act co-operatively in transfected cells, including in NPCs themselves, in reporter-based assays. Further, we propose candidate TFs (in particular Hoxb2 and Sox10) involved in the regulation of *ZEB2* via their control of its new enhancers.

Binding of TFs to enhancers/promoters regulates expression of genes and controls cell fate (Takahashi and Yamanaka, 2006). However, it is still unclear how different levels of TFs affect dynamics of binding and whether this would at all contribute to selectivity of TF action on certain, but not all bound target genes, for the latter may also depend on recruited co-factors and/or local PTM of the TF that complexes with PTM-enzymes. Mapping of ZEB2 binding-sites has thus far in the few cases reported been studied using high-ZEB2 hepatocellular carcinoma and leukemia cell lines (Balcik-Ercin *et al.*, 2018). Abnormally high levels of ZEB2 could change the dynamics

of binding too. To avoid this, we invested in the development of a Zeb2-V5-tag mouse ESC line, which maintains normal neuroprogenitor and mesendodermal differentiation, respectively. I co-investigated by application of bio-informatics the genome-wide binding sites of this tagged-Zeb2, produced at its normal endogenous mRNA/protein level, using ChIP-seq (*see* **Chapter 7**). This resulted in confirmation of already described Zeb2 targets (*Cdh1*), novel strong candidate target genes (such as *Tcf4*) that are linked to neurodevelopmental disorders, such as Pitt-Hopkins Syndrome (PTHS).

Furthermore, we discovered that *Zeb2* binds to its own promoter region, thus revealing autoregulation, which is critical for ESC differentiation towards NPCs and mesendodermal cells, respectively. Altogether, the work described in **Chapters 6** and **7** provides better understanding of *Zeb2* regulation and actions. Moreover, the identification of genome-wide binding of Zeb2 at its normal level provides solid ground for further investigation of ZEB2-orchestrated GRNs and its role in molecular etiology of not only MOWS (now expanding into non-coding mutations), but also as candidate modifier of neurodevelopmental syndromes caused by mutation in other TF-encoding genes that are bound by ZEB2. This type of research goes beyond the field of developmental biology, to clinical studies.

8.3.2. Future perspectives

Identification of novel mechanisms of *Zeb2* locus regulation provides a great opportunity to investigate the role of *Zeb2* during embryonic development. As shown in **Chapter 6**, the *Zeb2* enhancers present dynamic activities when studied separately or together. Therefore, CRISPR/Cas-based deletion of the abovementioned enhancers or Zeb2 binding-site in its promoter-proximal region (as described in **Chapter 7**) could provide novel mouse models. Further, combination of the deep knowledge of *Zeb2* mutant mouse models (including cell-type specific KOs, cKOs) with ChIP-seq (in the same cells as in cKOs) of a still to-be-made *Zeb2*-tag mouse line would be a clear asset to understand intact-Zeb2 dependent direct target genes. Such studies could then be further extended to investigate Zeb2 actions and targets post-natally, including in challenged mice and in chronic disease, such as cancer. Zeb2 indeed plays an important role in normal and pathologic EMT (DaSilva-Arnold *et al.,* 2019).

8.4. Discussion and future perspectives in general

The precise spatio-temporal regulation of gene expression is orchestrated at multiple levels, including among all the dynamic regulation of 3D chromatin organization, and also complex transcription control obtained by combination of *cis*-RE and *trans*-REs. In this PhD research, application of multi-omics techniques such as developed T2C protocol combined with integrative analysis of ChIPseq and RNAseq data, and also functional enhancer assays was used to investigate dynamic changes of 3D chromatin structure, histone profiles associated with enhancers, and changes in enhancer activities during differentiation and development. Importantly, application of T2C protocol for *in vivo* samples as well as cultured cells highlines the utility of this technique. The T2C was used to characterize 3D chromatin structure during development (**Chapter 5**) and differentiation (**Chapter 6**) revealing dynamic changes of inter-TAD *cis*-proximities.

Importantly, the identification of DNA-loops using T2C data revealed that some of the E-P *cis*-proximities correlate with mRNA expression and changes in H3K37ac histone mark in both, *Dact2-Smoc2* locus (**Chapter 5**) and *Zeb2* locus (**Chapter 6**). Also, functional characterization of *ZEB2* enhancers revealed that changes in enhancer activity occurs together with reorganization of 3D chromatin architecture during differentiation, as shown by T2C interaction maps (**Chapter 6**). This PhD research also characterized differences in histone profiles within candidate enhancer regions, which most likely precedes changes of enhancers activity (**Chapter 3** and **Chapter 5**), and revealed autoregulatory mechanism of *Zeb2* upregulation via binding of Zeb2 to its own promoter (**Chapter 7**). To summarize, application of *in vitro* and *in vivo* models, multi-omics methods, and functional assays provides a powerful combination of tools, which allows to investigate multiple regulatory mechanisms, together with their crosstalk, in genome-wide and locus-specific studies to reveal the principles of strict and precise control of gene expression.

Despite the identification of tens of thousands of CEs, with the numbers growing every year (Gao *et al.,* 2016; Gao and Qian, 2020), the field still lacks more precise and above all user-friendly tools for annotation of such regulatory elements to target genes and further connect this to patient phenotypes. GREAT is one of the most popular tools for such annotation and interpretation of *cis*-regulatory elements. However, it links distal enhancers to target genes based on genomic distances, leading to generation of multiple false positives (McLean *et al.,* 2010). Therefore, the generation of a web-server tool, which combines ChIP-seq/ATAC-seq derived CEs with interactome data generated by 3C-based technique(s), would allow to annotate the CEs to target genes with relatively higher precision, and subsequently reduce error rates. Such tool should also enable to associate CEs with potential phenotypes and could definitely be used intensely in future clinical studies.

It is precisely the hitherto lack of such tool, which combines the mentioned utilities and is characterized by high precision, that prompted me to participate in another project outside of my PhD thesis. This project involved the development of TADeus2, a web-server tool for the clinical diagnosis of genomic alterations affecting 3D chromatin structure (Poszewiecka *et al.,* 2022). Among all, TADeus2 allows to analyze the CEs in the context of loss/gain-of-function via *in silico* perturbation (such as deletion/duplication) of CEs and association of copy number variations with potentially affected genes, which are further ranked based on pathogenicity score and patient phenotype(s).

To yet again improve TADeus2 precision for enhancer annotation in this next, ongoing project, we decided to focus on the development of an additional mapping tool based on various 3C-based data. This “suite” of tools will be further upgraded by additional utilities designed for identification of binding motifs and hence putative binding sites of TFs.

Literature references

Balcik-Ercin P, Cetin M, Yalim-Camci I, Odabas G, Tokay N, Sayan AE, Yagci T. Genome-wide analysis of endogenously expressed ZEB2 binding sites reveals inverse correlations between ZEB2 and GalNAc-transferase GALNT3 in human tumors. Cell Oncol (Dordr). 2018 Aug;41(4): 379-93.

Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods. 2017 Jul 1;123:56-65.

Benito-Kwiecinski S, Giandomenico SL, Sutcliffe M, Riis ES, Freire-Pritchett P, Kelava I, Wunderlich S, Martin U, Wray GA, McDole K, Lancaster MA. An early cell shape transition drives evolutionary expansion of the human forebrain. Cell. 2021 Apr 15;184(8):2084-2102.e19.

Birkhoff JC, Huylebroeck D, Conidi A. ZEB2, the Mowat-Wilson Syndrome Transcription Factor: Confirmations, Novel Functions, and Continuing Surprises. Genes (Basel). 2021 Jul 3;12(7):1037.

Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. Genome Res. 2019 May;29(5):771-783.

Burgess DJ. Spatial transcriptomics coming of age. Nat Rev Genet. 2019 Jun;20(6):317.

Capellini TD, Chen H, Cao J, Doxey AC, Kiapour AM, Schoor M, Kingsley DM. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat Genet. 2017 Aug;49(8):1202-1210.

Conidi A, Cazzola S, Beets K, Coddens K, Collart C, Cornelis F, Cox L, Joke D, Dobreva MP, Dries R, Esguerra C, Francis A, Ibrahimi A, Kroes R, Lesage F, Maas E, Moya I, Pereira PN, Stappers E, Stryjewska A, van den Berghe V, Vermeire L, Verstappen G, Seuntjens E, Umans L, Zwijsen A, Huylebroeck D. Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFβ/BMP signaling *in vivo*. Cytokine Growth Factor Rev. 2011 Oct-Dec;22(5-6):287-300.

DaSilva-Arnold SC, Kuo CY, Davra V, Remache Y, Kim PCW, Fisher JP, Zamudio S, Al-Khan A, Birge RB, Illsley NP. ZEB2, a master regulator of the epithelial-mesenchymal transition, mediates trophoblast differentiation. Mol Hum Reprod. 2019 Feb 1;25(2):61-75.

De Bari C, Roelofs AJ. Stem cell-based therapeutic strategies for cartilage defects and osteoarthritis. Curr Opin Pharmacol. 2018 Jun;40:74-80.

Deng Y, Bartosovic M, Ma S, Zhang D, Kukanja P, Xiao Y, Su G, Liu Y, Qin X, Rosoklija GB, Dwork AJ, Mann JJ, Xu ML, Halene S, Craft JE, Leong KW, Boldrini M, Castelo-Branco G, Fan R. Spatial profiling of chromatin accessibility in mouse and human tissues. Nature. 2022 Aug 17. doi: 10.1038/s41586-022-05094-1. Online ahead of print.

Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, Lin-Marq N, Koch M, Bilio M, Cantiello I, Verde R, De Masi C, Bianchi SA, Cicchini J, Perroud E, Mehmeti S, Dagand E, Schrinner S, Nürnberger A, Schmidt K, Metz K, Zwingmann C, Brieske N, Springer C, Hernandez AM, Herzog S, Grabbe F, Sieverding C, Fischer B, Schrader K, Brockmeyer M, Dettmer S, Helbig C, Alunni V, Battaini MA, Mura C, Henrichsen CN, Garcia-Lopez R, Echevarria D, Puelles E, Garcia-Calero E, Kruse S, Uhr M, Kauck C, Feng G, Milyaev N, Ong CK, Kumar L, Lam M, Semple CA, Gyenesei A, Mundlos S, Radelof U, Lehrach H, Sarmientos P, Reymond A, Davidson DR, Dollé P, Antonarakis SE, Yaspo ML, Martinez S, Baldock RA, Eichele G, Ballabio A. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol. 2011 Jan 18;9(1):e1000582.

Dries R, Stryjewska A, Coddens K, Okawa S, Notelaers T, Birkhoff J, Dekker M, Verfaillie CM, Del Sol A, Mulugeta E, Conidi A, Grosveld FG, Huylebroeck D. Integrative and perturbation-based analysis of the transcriptional dynamics of TGFβ/BMP system components in transition from embryonic stem cells to neural progenitors. Stem Cells. 2020 Feb;38(2):202-217

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495-501.

Golov AK, Abashkin DA, Kondratyev NV, Razin SV, Gavrilov AA, Golimbet VE. A modified protocol of Capture-C allows affordable and flexible high-resolution promoter interactome analysis. Sci Rep. 2020 Sep 23;10(1):15491.

Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, Zhang B, Wang M, Strattan JS, Davidson JM, Qiu Y, Afzal V, Akiyama JA, Plajzer-Frick I, Novak CS, Kato M, Garvin TH, Pham QT, Harrington AN, Mannion BJ, Lee EA, Fukuda-Yuzawa Y, He Y, Preissl S, Chee S, Han JY, Williams BA, Trout D, Amrhein H, Yang H, Cherry JM, Wang W, Gaulton K, Ecker JR, Shen Y, Dickel DE, Visel A, Pennacchio LA, Ren B. An atlas of dynamic chromatin landscapes in mouse fetal development. Nature. 2020 Jul;583(7818):744-751. With author correction: Nature. 2020 Oct;586(7831):E31.

Hegarty SV, Wyatt SL, Howard L, Stappers E, Huylebroeck D, Sullivan AM, O'Keeffe GW. Zeb2 is a negative regulator of midbrain dopaminergic axon growth and target innervation. Sci Rep. 2017 Aug 17;7(1):8568.

Hong YJ, Do JT. Neural Lineage Differentiation From Pluripotent Stem Cells to Mimic Human Brain Tissues. Front Bioeng Biotechnol. 2019 Dec 6;7:400.

Hirsch N, Eshel R, Bar Yaacov R, Shahar T, Shmulevich F, Dahan I, Levaot N, Kaplan T, Lupiáñez DG, Birnbaum RY. Unraveling the transcriptional regulation of TWIST1 in limb development. PLoS Genet. 2018 Oct 29;14(10):e1007738.

Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. Genomics. 2015 Sep;106(3):159-164.

Kania K, Colella F, Riemen AHK, Wang H, Howard KA, Aigner T, Dell'Accio F, Capellini TD, Roelofs AJ, De Bari C. Regulation of Gdf5 expression in joint remodelling, repair and osteoarthritis. Sci Rep. 2020 Jan 13;10(1):157.

Kothary RK, Allen ND, Surani MA. Transgenes as molecular probes of mammalian developmental genetics. Oxf Surv Eukaryot Genes. 1989a;6:145-78.

Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development. 1989b Apr;105(4):707-14.

Kvon EZ, Zhu Y, Kelman G, Novak CS, Plajzer-Frick I, Kato M, Garvin TH, Pham Q, Harrington AN, Hunter RD, Godoy J, Meky EM, Akiyama JA, Afzal V, Tran S, Escande F, Gilbert-Dussardier B, Jean-Marçais N, Hudaiberdiev S, Ovcharenko I, Dobbs MB, Gurnett CA, Manouvrier-Hanu S, Petit F, Visel A, Dickel DE, Pennacchio LA. Comprehensive *In vivo* Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Cell. 2020 Mar 19;180(6):1262-1271.e15.

Lu H, Ju DD, Yang GD, Zhu LY, Yang XM, Li J, Song WW, Wang JH, Zhang CC, Zhang ZG, Zhang R. Targeting cancer stem cell signature gene SMOC-2 Overcomes chemoresistance and inhibits cell proliferation of endometrial carcinoma. EBioMedicine. 2019 Feb;40:276-289.

Lu L, Liu X, Huang WK, Giusti-Rodríguez P, Cui J, Zhang S, Xu W, Wen Z, Ma S, Rosen JD, Xu Z, Bartels CF, Kawaguchi R, Hu M, Scacheri PC, Rong Z, Li Y, Sullivan PF, Song H, Ming GL, Li Y, Jin F. Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. Mol Cell. 2020 Aug 6;79(3):521-534.e15.

Mommaerts H, Esguerra CV, Hartmann U, Luyten FP, Tylzanowski P. Smoc2 modulates embryonic myelopoiesis during zebrafish development. Dev Dyn. 2014 Nov;243(11):1375-90.

Nowosad K, Hordyjewska-Kowalczyk E, Tylzanowski P. Mutations in gene regulatory elements linked to human limb malformations. J Med Genet. 2020 Jun;57(6):361-370.

Ohta Y, Okabe T, Larmour C, Di Rocco A, Maijenburg MW, Phillips A, Speck NA, Wakitani S, Nakamura T, Yamada Y, Enomoto-Iwamoto M, Pacifici M, Iwamoto M. Articular cartilage endurance and resistance to osteoarthritic changes require transcription factor Erg. Arthritis Rheumatol. 2015 Oct;67(10):2679-90.

Peeters T, Monteagudo S, Tylzanowski P, Luyten FP, Lories R, Cailotto F. SMOC2 inhibits calcification of osteoprogenitor and endothelial cells. PLoS One. 2018 Jun 13;13(6):e0198104.

Poszewiecka B, Pienkowski VM, Nowosad K, Robin JD, Gogolewski K, Gambin A. TADeus2: a web server facilitating the clinical diagnosis by pathogenicity assessment of structural variations disarranging 3D chromatin structure. Nucleic Acids Res. 2022 May 7;50(W1):W744–52.

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N; Human Cell Atlas Meeting Participants. The Human Cell Atlas. Elife. 2017 Dec 5;6:e27041.

Sensiate LA, Sobreira DR, Da Veiga FC, Peterlini DJ, Pedrosa AV, Rirsch T, Joazeiro PP, Schubert FR, Collares-Buzato CB, Xavier-Neto J, Dietrich S, Alvares LE. Dact gene expression profiles suggest a role for this gene family in integrating Wnt and TGF-β signaling pathways during chicken limb development. Dev Dyn. 2014 Mar;243(3):428-39.

Settle SH Jr, Rountree RB, Sinha A, Thacker A, Higgins K, Kingsley DM. Multiple joint and skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5 genes. Dev Biol. 2003 Feb 1;254(1):116-30.

Stryjewska A, Dries R, Pieters T, Verstappen G, Conidi A, Coddens K, Francis A, Umans L, van IJcken WF, Berx G, van Grunsven LA, Grosveld FG, Goossens S, Haigh JJ, Huylebroeck D. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells. 2017 Mar;35(3):611-625.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug 25;126(4):663-76.

Teufel, S., P. Köckemann, U. König, and C. Hartmann. Loss of Wnt9a and Wnt4 causes degenerative joint alterations. Osteoarthritis and Cartilage. 2018 26:S94-S95.

Wang S, Dong Y, Zhang Y, Wang X, Xu L, Yang S, Li X, Dong H, Xu L, Su L, Ng SS, Chang Z, Sung JJ, Zhang X, Yu J. DACT2 is a functional tumor suppressor through inhibiting Wnt/β-catenin pathway and associated with poor survival in colon cancer. Oncogene. 2015 May 14;34(20):2575-85.

Wu LM, Wang J, Conidi A, Zhao C, Wang H, Ford Z, Zhang L, Zweier C, Ayee BG, Maurel P, Zwijsen A, Chan JR, Jankowski MP, Huylebroeck D, Lu QR. Zeb2 recruits HDAC-NuRD to inhibit Notch and controls Schwann cell differentiation and remyelination. Nat Neurosci. 2016 Aug;19(8):1060-72.

Xia B, Di Chen, Zhang J, Hu S, Jin H, Tong P. Osteoarthritis pathogenesis: a review of molecular mechanisms. Calcif Tissue Int. 2014 Dec;95(6):495-505.

Yang S, Oksenberg N, Takayama S, Heo SJ, Poliakov A, Ahituv N, Dubchak I, Boffelli D. Functionally conserved enhancers with divergent sequences in distant vertebrates. BMC Genomics. 2015 Oct 30;16:882.

Yuan X, Song M, Devine P, Bruneau BG, Scott IC, Wilson MD. Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. Nat Commun. 2018 Nov 26;9(1):4977.

Appendix

Summary

Our genome contains an estimated number of 400,000 to 1.4 million enhancers. These are regulatory sequences mostly located outside exons of protein-coding genes either close to or far away from the transcription start site(s) of genes on the chromosome. Together with gene promoters, they are principal cell-intrinsic genomic elements assuring the activity and accurate regulation of transcription of genes, in terms of the steady-state mRNA levels and spatio-temporal precision. During early embryogenesis and subsequent organogenesis, enhancers co-regulate decisions in cell fate and differentiation, and thus creation of cell diversity. Such regulatory sequences operate often in a context of extrinsic cell stimulation by polypeptide growth factors, the signaling cascade of which is interpreted and further executed in the nucleus by transcription factors, assuring RNAPol2-mediated activation of sets of genes by their binding to promoter and enhancer sequences.

These precise molecular actions must take place in a cell nucleus of 6-µm diameter average, and wherein 2m of DNA is highly compacted and present as architecturally ordered chromatin (DNA and bound proteins), as individual chromosomes. This chromatin in general, as well as its typical nucleosomes, but also enhancer and promoter sequences, are epigenetically marked (biochemically modified, in various ways). This collectively assures that the chromatin opens locally when gene transcription activation is needed, and gene-specific enhancers must thereby achieve physical proximity, which can be documented by chromatin conformation capture (3C) analyses, with the promoter of their cognate gene. At the same time, a set of biochemical marks, including those modifying histone-3 (H3) in the nucleosome, facilitates the identification of candidate enhancers.

Importantly, like key protein-coding genes, sequence variation or mutation of enhancers can lead or contribute to congenital syndromes and chronic disease. Hence, both experimental studies of enhancers (including their identification in small population of embryonic cells) and key developmental transcription factors (including how their own level and activity are regulated, but also which genes and regulatory sequences they bind to), are fundamental to understanding cell-based health and disease of the entire organism.

The research in this PhD thesis addresses these fundamental aspects of genetic and molecular control of embryogenesis, and at the same time investigates the mechanisms of gene regulation. It presents two lines of experimental as well as bio-informatics work, one *in vivo* and one in cell culture, specifically in the formation of flexible joints in developing limbs and chondrogenesis in the latter, and neural differentiation of pluripotent embryonic cells, respectively. The mechanisms of gene regulation most relevant to this PhD research are described in Chapter 1, which also includes an overview of these molecular regulations as well as mutations in regulatory elements that link to selected human limb malformations.

The experimental part of research line-1 reports on the production, for the first time, of a genome-wide candidate-enhancer atlas of the joint interzone and adjacent phalanges, respectively. This work includes integrative analysis of transcriptomic data from RNA-sequencing together with H3K27ac and H3K4me1 signatures obtained by ChIP-sequencing (Chapter 3). It then reports on contribution to the establishment of a low-T2C (a targeted 3C-assay) protocol applicable to cell populations or *in vivo* samples available as low cell numbers, like for synovial joints (Chapter 4). Then, this low-T2C protocol is used to investigate the genomic region encompassing the *Dact2 and Smoc2* genes, and identify and characterize their enhancers in the

interzone during synovial joint development, which were also validated using a zebrafish larvae enhancer assay (Chapter 5).

In research line-2, using neural differentiation of embryonic stem cells (ESCs), the thesis reports on the demonstration of dynamic DNA-loops in and around the human *ZEB2* locus (including its 3.5 Mb-long gene desert), and co-operation between the newly identified enhancers, including in human neuroprogenitor cells (NPCs) (Chapter 6). The DNA-binding transcription factor ZEB2 is studied by many teams in different fields, but maps of its genome-wide binding sites are urgently needed. This PhD thesis therefore also includes ChIP-sequencing of Zeb2 (endogenously tagged in mouse ESCs, guaranteeing normal levels of Zeb2 production) in ESC-derived cultures of NPCs (Chapter 7). This work aims at identifying Zeb2-dependent, directly controlled target genes, as well as candidate TFs that regulate *Zeb2* gene expression via its identified enhancers, and illustrates for the first time how critical the identified autoregulation of *Zeb2*.

This PhD research as a whole combines two experimental models, multi-omics methods, and functional assays to investigate multiple regulatory mechanisms, in genome-wide as well as locus-specific studies. It documents dynamic changes in 3D chromatin architecture, enhancer signatures and activity, to reveal the underlying principles of precise control of gene expression during development and differentiation.

Samenvatting

Ons genoom bevat naar schatting 400,000 tot 1,4 miljoen enhancers. Dat zijn regulatorische sequenties die, meestal buiten de exons van eiwit-coderende genen, dichtbij tot ver van de tramscriptie startplaats(en) van genen op het chromosoom zijn gelokaliseerd. Samen met genpromoters behoren enhancers tot de voornaamste celintrinsieke genomische elementen die de activiteit en de accurate regulatie van de transcriptie van genen verzekeren, vooral wat stabiele mRNA niveaus en de spatio-temporele precisie betreffen. Gedurende de vroege embryogenese en organogenese co-reguleren enhancers beslissingen over het lot en de differentiatie van cellen, en dus de creatie van celdiversiteit. Dikwijls werken dergelijke regulatorische sequenties in een context van extrinsieke stimulatie van cellen door polypeptide groeifactoren, waarvan de signalisatiecascade wordt geïnterpreteerd en verder bewerkstelligd door transcriptiefactoren in de celkern, die hiermee RNAPol2-gemedieerde activatie van sets van genen verzekeren door te binden aan promoter- en enhancersequenties.

Deze nauwkeurige moleculaire acties moeten wel plaats grijpen in een celkern van gemiddeld 6 µm diameter, en waarin 2m DNA dus in hoge mate is samengetrokken en aanwezig als architecturaal geordend chromatine (DNA en gebonden eiwitten), als individuele chromosomen. In het algemeen, alsook de typische nucleosomen ervan, zijn chromatine, maar ook enhancer- en promotersequenties, epigenetisch gestempeld (biochemisch divers gemodifieerd). Alles samen verzekert dit dat het chromatine lokaal opent wanneer gentranscriptie moet worden geactiveerd, en genspecifieke enhancers hierbij ook in de fysische nabijheid van de promoter van het cognate gen komt, wat kan gedocumenteerd worden door chromatin conformatie capture (3C) analyses. Ook faciliteren sets van biochemische stempels, inbegrepen die die histone-3 (H3) modifiëren in het nucleosoom, de identificatie van kandidaat-enhancers.

Belangrijk, en net zoals voor eiwit-coderende genen, is dat variatie of mutatie van enhancersequenties kan leiden of bijdragen tot aangeboren afwijkingen en chronische ziekte. Bijgevolg zijn zowel experimentele studies van enhancers (inbegrepen hun identificatie in kleine populaties van embryonale cellen) en voor de ontwikkeling belangrijke transcriptiefactoren (inbegrepen hoe hun eigen hoeveelheid en activiteit zijn gereguleerd, maar ook welke genen en regulatorische sequenties zij binden), van het hoogste fundamenteel belang voor het begrijpen van celgebaseerde gezondheid en ziekte van het volledige organisme.

Het onderzoek in deze doctoraatsthesis richt zich op deze fundamentele aspecten van de genetische en de moleculaire controle van embryonale ontwikkeling, en bestudeert terzelfdertijd ook mechanismen van genregulatie. Het presenteert zich in twee lijnen van experimenteel zowel als bio-informatica werk, één *in vivo* en één in celcultuur, meer in het bijzonder en respectievelijk de vorming van flexibele gewrichten in ontwikkelende extremiteiten en de chondrogenese in deze laatste, en neurale differentiatie van pluripotente embryonale cellen. De genregulatorische mechanismen die het merest relevant zijn voor dit PhD onderzoek wordenbesproken in hoofdstuk 1. Dit hoofdstuk bevat ook een geselecteerd overzicht van deze moleculaire regulaties zowel als mutaties in regulatorische elementen die verband houden met aangeboren afwijkingen van extremiteiten bij de mens.

Het experimentele deel van onderzoekslijn-1 gaat over de aanmaak, voor de eerste keer, van een genoomwijde kandidaat-enhancer atlas van het vormende gewricht, nl. in cellen van respectievelijk zijn interzone zelf en de naburige, toekomstige kootjes. Dit werk omvat integratieve analyse van transcriptoomdata uit RNA-sequencing samen met H3K27ac en

K3K4me1 stempels bekomen door ChIP-sequencing (hoodfstuk 3). Nadien volgt dan de bijdrage tot het optimaliseren van een zgn. low-T2C (een doelgerichte 3C-test) labprotocol toepasbaar voor celpopulaties of *in vivo* cellen die slechts in zeer kleine aantallen kunnen worden bekomen, zoals trouwens het geval is voor het synoviale gewricht (hoofdstuk 4). Dit T2C protocol wordt dan gebruikt om de genomische regio te onderzoeken dat de *Dact2* en *Smoc2* genen bevat, en enhancers te identificeren en te karakteriseren in de interzone gedurende de ontwikkeling van het synoviale gewricht, en die verder werden gedocumenteerd in een enhancertest in zebravislarven (hoofdstuk 5).

In onderzoekslijn-2 wordt gebruikt gemaakt van neurale differentiatie van embryonale stamcellen (ESCs). Hierin tonen wij dynamische DNA-lussen in en rond de humane *ZEB2* locus aan (inbegrepen zijn 3.5 Mb-lange genwoestijn), en samenwerking tussen de nieuw geïdentificeerde enhancers, inbegrepen in neurale voorlopercellen (NPCs) (hoofdstuk 6). De DNA-bindende transcriptiefactor ZEB2 wordt intensief bestudeerd door vele onderzoeksgroepen in meerdere onderzoeksvelden, en een kaart van zijn genoomwijde bindingsplaatsen is dringend nodig. Daarom bevat deze doctoraatsthesis ook experimenteel werk, met name ChIP-sequencing van Zeb2 (endogeen van een heteroloog epitoop voorzien in muis ESCs, en gegarandeerd met normale niveaus van Zeb2 productie) in ESC-afgeleide NPCs (hoofdstuk 7). Alles samen zal dit werk dus toelaten om Zeb2-afhankelijke, directe doelwitgenen te identificeren die Zeb2 controleert, zowel als kandidaat-transcriptiefactoren die *Zeb2* genexpressie reguleren via zijn geïdentificeerde enhancers, en bracht dit ook voor het eerst aan het licht hoe kritisch de eveneens geïdentificeerde autoregulatie van *Zeb2* is voor celdifferentiatie.

Dit PhD onderzoek in zijn geheel combineert twee experimentele modellen, multi-omics methoden, en functionele testen voor het onderzoeken van meerdere regulatorische mechanismen, in zowel genoomwijde als locusspecirieke studies. Het documenteert dynamische veranderingen in 3D chromatine architectuur, enhancermarkering en activiteit, om de onderliggende principes van de precieze controle van genexpressie in ontwikkeling en differentiatie.

Streszczenie

Genom człowieka zawiera od 400,000 do 1.4 miliona sekwencji wzmacniających (enhancerów). Enhancery to sekwencje regulatorowe zlokalizowane głównie poza eksonami. Enhancery mogą znajdować się zarówno w pobliżu, jak też w dużej odległości od regionów rozpoczęcia transkrypcji. Wraz z promotorami enhancery są głównymi wewnątrzkomórkowymi czynnikami genetycznymi zapewniającymi aktywność i właściwą regulację transkrypcji, również w kontekście czasowo-przestrzennej ekspresji genów. Podczas wczesnych etapów embriogenezy, a następnie organogenezy, enhancery koordynują decyzje dotyczące losu i różnicowania komórek, a tym samym biorą udział w tworzeniu różnorodności komórkowej. Sekwencje regulatorowe często działają w kontekście zewnętrznej stymulacji komórek poprzez polipeptydowe czynniki wzrostu, których kaskada sygnałowa jest interpretowana, a następnie uruchamiana w jądrze komórkowym przez czynniki transkrypcyjne, gwarantując aktywację zestawów genów za pośrednictwem RNAPol2 poprzez ich wiązanie z sekwencjami promotorów i enhancerów.

Te precyzyjne procesy molekularne zachodzą w jądrze komórkowym o średnicy 6 µm, w którym 2 m DNA jest upakowane w postaci uporządkowanej struktury chromatyny (DNA i związane białka) jako pojedyncze chromosomy. Sama chromatyna, jak też zawarte w niej nukleosomy, a także sekwencje wzmacniające (enhancery) i promotorowe, są oznaczone epigenetycznie (poprzez różnorodne modyfikacje biochemiczne). To sprawia, że chromatyna otwiera się lokalnie, gdy potrzebna jest aktywacja transkrypcji określonego genu. Aktywacja transkrypcji wymaga również aby enhancery znajdowały się w odpowiedniej odległości od genów, co można udokumentować za pomocą analiz konformacji chromatyny (3C). Dodatkowo, analiza modyfikacji biochemicznych, w tym modyfikacji histonu-3 (H3) w nukleosomie, umożliwia identyfikację potencjalnych enhancerów.

Co ważne, jak w przypadku genów kodujących białka, zmienność sekwencji lub mutacje w obszarze enhancerów mogą przyczyniać się do powstania licznych zespołów wad wrodzonych oraz chorób przewlekłych. Stąd zarówno badania eksperymentalne dotyczące enhancerów (w tym ich identyfikacja w małej populacji komórek embrionalnych), jak też kluczowych czynników transkrypcyjnych dla rozwoju embrionalnego (w tym analiza procesów regulujących ich poziom ekspresji i aktywność, a także ocena, z którymi genami i sekwencjami regulującymi się wiążą), mają istotne znaczenie w zrozumieniu podstaw powstawania chorób rozwojowych.

Badania zawarte w rozprawie doktorskiej dotyczą podstawowych aspektów genetycznej i molekularnej kontroli embriogenezy oraz skupiają się na opisie mechanizmów zaangażowanych w regulację ekspresji genów. Badania te mogą być podzielone na dwa główne nurty. Pierwszy obejmuje prace eksperymentalne w modelu *in vivo* w celu analizy rozwoju stawów i chondrogenezy. Drugi wykorzystuje model *in vitro*, w celu analizy różnicowania embrionalnych komórek pluripotentnych w stronę neuronów. Mechanizmy regulacji ekspresji genów zostały opisane w rozdziale 1, który zawiera ponadto przegląd mutacji występujących w elementach regulatorowych, wpływających na rozwój wybranych wad wrodzonych obejmujących zmiany w układzie kostnym człowieka.

Pierwsza część badań obejmująca analizy w modelu *in vivo* pozwoliła na utworzenie (po raz pierwszy) atlasu potencjalnych enhancerów występujących w całym genomie, odpowiednio w interzonach stawów i przyległych do nich paliczkach. Część ta obejmuje analizę danych transkryptomicznych z sekwencjonowania RNA wraz z sygnaturami H3K27ac i H3K4me1

uzyskanymi metodą sekwencjonowania ChIP (Rozdział 3). Rozprawa doktorska zawiera ponadto opracowany protokół low-T2C (w oparciu o technologię 3C), który może być wykorzystany do analizy 3D struktury chromatyny małych populacji komórek lub próbek *in vivo* o stosunkowo niskiej liczebności, takich jak m.in. stawy maziowe (Rozdział 4). Opracowany low-T2C protokół został wykorzystany do analizy regionu genomowego obejmującego geny *DACT2* i *SMOC2* oraz identyfikacji, a także charakterystyki ich enhancerów w interzonach podczas rozwoju stawu maziowego. Zidentyfikowane enhancery zostały zwalidowane za pomocą analizy aktywności enhancerów w trakcie rozwoju larw danio pręgowanego (Rozdział 5).

W drugiej części badań, wykorzystującej protokoły do różnicowania embrionalnych komórek macierzystych (ESC) w stronę neuronów, zanalizowana została struktura 3D chromatyny oraz dynamika pętli DNA w genomie człowieka w regionie obejmującym gen *ZEB2* (3,5 Mb, w tym region nie zawierający genów - z ang. "gene desert"). Dodatkowo badania obejmowały analizę zidentyfikowanych enhancerów *ZEB2* w ludzkich komórkach neuroprogenitorowych (NPC) (Rozdział 6). Czynnik transkrypcyjny *ZEB2* wiążący się do DNA jest szeroko badany przez wiele zespołów z różnych dziedzin naukowych, dlatego analiza miejsc wiązania *ZEB2* w całym genomie jest obecnie pożądana. W związku z powyższym badania zawarte w pracy doktorskiej obejmują również sekwencjonowanie ChIP *Zeb2* (endogennie wyznakowanego w mysich ESC, co gwarantuje prawidłowy poziom ekspresji *Zeb2*) w populacji NPC pochodzących z ESC. Ta część rozprawy doktorskiej miała na celu identyfikację zarówno genów kontrolowanych przez *Zeb2*, jak również czynników transkrypcyjnych, które wpływają na regulację ekspresji *Zeb2* poprzez kontrolujące go enhancery. Ponadto uzyskane wyniki ilustrują po raz pierwszy, jak ważna jest autoregulacja *Zeb2* w procesie różnicowania komórek (Rozdział 7).

Badania zawarte w rozprawie doktorskiej wykorzystują dwa modele eksperymentalne, metody oparte o sekwencjonowanie następnej generacji (ang. NGS) oraz techniki do badań funkcjonalnych, w celu analizy mechanizmów regulacji genów z uwzględnieniem ich złożoności oraz interakcji. Uzyskane wyniki pozwoliły na charakterystykę dynamiki 3D struktury chromatyny, profilu enhancerów oraz ich aktywności, umożliwiając głębsze poznanie proceów zaangażowanych w regulację genów w trakcie rozwoju embrionalnego oraz różnicowania komórkowego.

Curriculum vitae

Personal information

Karol Nowosad

Date of Birth: March 11, 1993
Legal address: 6/25 Kotlarska Street,
22-200 Włodawa, Poland
E-mail: karol.nowosad@umlub.pl
k.nowosad@erasmusmc.nl

Education

- October 1, 2017 – present:

  Postgraduate School of Molecular Medicine, Medical University Lublin, operating as part of an overarching Postgraduate School coordinated at University of Warsaw, Poland

- October 1, 2017 – present:

  PhD student at the Department of Cell Biology, Erasmus University Medical Center, Rotterdam, The Netherlands. Promoter: D. Huylebroeck

  PhD student at the Department of Biochemistry and Molecular Biology, Medical University of Lublin, Poland. Promoter: Prof. P. Tylzanowski

- October 1, 2015 – May 29, 2017:

  Master degree, with thesis: "Analysis of interaction of partitioning protein RepA and RepB of *Rhizobium leguminosarum* pRleTA1d and pRleTA1c plasmids using bacterial two-hybrid system" under supervision of Andrzej Mazur, PhD, Department of Genetics and Microbiology, University of Maria Curie-Skłodowska, Poland. Major subject: biotechnology;

- October 1, 2012 – July 10, 2015:

  Bachelor degree, with thesis: "*Cryptococcus neoformans* – characterization" under supervision of Monika Marek-Kozaczuk, PhD, Department of Genetics and Microbiology, University Marie Curie-Sklodowska in Lublin, Poland. Major subject: biotechnology

Work experience / Fellowships

- Apr. 2021-May 2021: *NAWA PROM fellowship* at the Department of Cell Biology, Erasmus University Medical Centre Rotterdam, NL. Project description: Integrated analysis of RNA-seq, ChIP-seq and Targeted Chromatin Capture (T2C) datasets for identification of cis-regulatory elements in developing chicken hindlimb digit. Under supervision of Danny Huylebroeck.

- Nov. 2019–Feb. 2020: *EMBO short-term fellowship* at the Arthritis and Regenerative Medicine Laboratory, Institute of Medical Sciences, University of Aberdeen, UK. Project description: Investigation of gene expression during synovial joint formation and regeneration of articular cartilage using bulk and single cell (sc)RNAseq. Under supervision of Cosimo de Bari.

- Jun. 2017-Sep. 2017: *Erasmus+ fellowship* at the Department of Cell Biology, Erasmus University Medical Centre Rotterdam, NL. Project description: Optimization of the Targeted Chromatin Capture (T2C) protocol for low input samples. Under supervision of Danny Huylebroeck and Frank Grosveld.
- Jul. 2016 –Oct. 2016: *UMCS Dean Fellowship* at the Department of Plant, Soil and Entomological Sciences, University of Idaho, USA; Project description: Identification and analysis of NAC transcription factor candidate targets. Under supervision of Fangming Xiao.

Courses

- Course "Alignment, Visualisation and Variant Calling", The Centre for Genome-Enabled Biology and Medicine (CGEBM), University of Aberdeen, Aberdeen, February 25, 2020
- Course "RNA Sequencing and Differential Expression", The Centre for Genome-Enabled Biology and Medicine (CGEBM), University of Aberdeen, Aberdeen, January 29, 2020
- Course "From Gene to Phenotype – Advances in Molecular Medicine", School of Molecular Medicine, Medical University of Warsaw, Warsaw, March 11-23, 2019
- "Training in methods of scarifying laboratory animals", Experimental Medicine Centre, Lublin, January 7, 2019
- Course "Life Science Imaging – Workshop on Visualization of Molecules, Interactions and Biological Processes", School of Molecular Medicine, Medical University of Warsaw, Warsaw, June 4-6, 2018
- Course "From Gene to Phenotype – Advances in Molecular Medicine", School of Molecular Medicine, Medical University of Warsaw, Warsaw, March 19-21, 2018
- "Training in planning and conducting procedures and experiments on animals", Experimental Medicine Centre, Lublin, January 25, 2018
- Course „Good Laboratory Practice (DPL/GLP)", Lublin, November 16, 2017

Membership

- Polish Biochemical Society

Publications

Published

Nowosad K, Brouwer RWW, Odrzywolski A, Korporaal AL, Gielniewski B, Wojtaś B, van IJcken WFJ, Grosveld F, Huylebroeck D, Tylzanowski P. Identification of candidate enhancers controlling the transcriptome during the formation of interphalangeal joints. ***Sci Rep.*** 2022 Jul 27;12(1):12835. doi: 10.1038/s41598-022-16951-4.

Poszewiecka B, Pienkowski VM, **Nowosad K**, Robin JD, Gogolewski K, Gambin A. TADeus2: a web server facilitating the clinical diagnosis by pathogenicity assessment of structural variations disarranging 3D chromatin structure. ***Nucleic Acids Res.*** 2022 May 7;50(W1):W744-52. doi: 10.1093/nar/gkac318.

Hordyjewska-Kowalczyk E, **Nowosad K**, Jamsheer A, Tylzanowski P. Genotype-phenotype correlation in clubfoot (*Talipes equinovarus*). ***J Med Genet.*** 2022 Mar;59(3):209-219. doi: 10.1136/jmedgenet-2021-108040.

Boltsis I, **Nowosad K**, Brouwer RWW, Tylzanowski P, van IJcken WFJ, Huylebroeck D, Grosveld F, Kolovos P. Low Input Targeted Chromatin Capture (Low-T2C). ***Methods Mol Biol.*** 2021;2351:165-179. doi: 10.1007/978-1-0716-1597-3_9.

Birkhoff JC, Brouwer RWW, Kolovos P, Korporaal AL, Bermejo-Santos A, Boltsis I, **Nowosad K**, van den Hout MCGN, Grosveld FG, van IJcken WFJ, Huylebroeck D, Conidi A. Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. ***Hum Mol Genet.*** 2020 Aug 29;29(15):2535-2550. doi: 10.1093/hmg/ddaa141.

Nowosad K, Hordyjewska-Kowalczyk E, Tylzanowski P. Mutations in gene regulatory elements linked to human limb malformations. ***J Med Genet.*** 2020 Jun;57(6):361-370. doi: 10.1136/jmedgenet-2019-106369.

Posted preprints of manuscripts in preparation

Birkhoff JC, Korporaal AL, Brouwer RWW, **Nowosad K**, Milazzo C, Mouratidou L, van den Hout MCGN, van IJcken WFJ, Huylebroeck D*, Conidi A*. Zeb2 DNA-binding sites in neuroprogenitor cells reveal autoregulation and affirm neurodevelopmental defects, including in Mowat-Wilson Syndrome. *In preparation.* Posted as preprint

https://www.biorxiv.org/content/10.1101/2021.07.06.451350v2

doi: 10.1101/2021.07.06.451350

Nowosad K, Malesa A, Hordyjewska-Kowalczyk E, Brouwer RWW, Odrzywolski A, Gielniewski B, Wojtaś B, Boltsis I, Birkhoff JC, van IJcken WFJ, Grosveld FG, Huylebroeck D, Conidi A, Tylzanowski P*. Cromatin architecture and *cis*-regulatory landscape of the *DACT2-SMOC2* genomic region in the developing synovial joint. Posted as preprint

https://www.biorxiv.org/content/10.1101/2022.10.06.511134v1

doi: 10.1101/2022.10.06.511134

PhD portfolio

Courses

- OMICS DATA SCIENCE traineeship (200h) – "Bioinformatics and analysis of large-scale biomedical data" organized by Interdisciplinary Center for Mathematical and Computational Modeling at the University of Warsaw (ICM UW), PL, zoom-online Oct. 2020-Feb. 2021
- "Code management with Git"; Graduate school Medical Genetics Centre South-West Netherlands (MGC), Leiden University Medical Center, Leiden, NL 6.10.2021
- "Next Generation Sequencing data analysis"; Graduate school Medical Genetics Centre South-West Netherlands (MGC), zoom-online 30.08-3.09.2021
- Course "Alignment, Visualisation and Variant Calling", The Centre for Genome-Enabled Biology and Medicine (CGEBM), University of Aberdeen, Aberdeen, February 25, 2020
- Course "RNA Sequencing and Differential Expression", The Centre for Genome-Enabled Biology and Medicine (CGEBM), University of Aberdeen, Aberdeen, January 29, 2020
- Course "From Gene to Phenotype – Advances in Molecular Medicine", School of Molecular Medicine, Medical University of Warsaw, Warsaw, March 11-23, 2019
- "Training in methods of scarifying laboratory animals", Experimental Medicine Centre, Lublin, January 7, 2019
- Course "Life Science Imaging – Workshop on Visualization of Molecules, Interactions and Biological Processes", School of Molecular Medicine, Medical University of Warsaw, Warsaw, June 4-6, 2018
- Course "From Gene to Phenotype – Advances in Molecular Medicine", School of Molecular Medicine, Medical University of Warsaw, Warsaw, March 19-21, 2018
- "Training in planning and conducting procedures and experiments on animals", Experimental Medicine Centre, Lublin, January 25, 2018
- Course „Good Laboratory Practice (DPL/GLP)", Lublin, November 16, 2017

Seminar and workshops

- Weekly departmental Monday Morning Meetings (Rotterdam) 2020-2022
- MGC PhD workshop (Oldenburg) 2022
- Introduction to single-cell sequencing and computational analysis (Leuven-online) 2021
- NGSprint Hackathon (online) 2021

Conferences

Poster presentation

- Biomedical Science PhD Day, Rotterdam 2022
- Young Scientists Conference on Molecular and Cell Biology, Warsaw 2019
- The Nuclear Landscapes, Warsaw 2018

Teaching (co-supervision of master students)

- 2019-2020: Ewa Berlinska, Thesis title: "*Study of conserved regulatory elements using zebrafish model organism"*.
- 2020-2021: Aneta Malesa, Thesis title: "*In vitro and in vivo studies of synovial joint enhancers"*.

Acknowledgments – Podziękowania

My scientific journey can be divided into multiple phases including (i) research at EUR, (ii) research at MUL, and (iii) all other quests I had to accomplish to finish my PhD research. Thus, I decided to divide the acknowledgments into three sections.

ERASMUS UNIVERSITY ROTTERDAM:

Dear **Danny**, once I heard in my life: "… If you want to be successful, find a mentor, not for weeks, but for years …". I was very lucky to find you, a true mentor who shaped me into the scientist I always wanted to be. You gave me a lot of freedom, but you also guided me through this scientific journey and helped me to focus on important things. I learned a lot during our discussions trying to 'sponge' as much as I could. Among all, you inspired me about Zeb2, which I plan to further study in Poland. Your scientific excitement encouraged me to go outside of my research area and discover new opportunities which helped me to become an interdisciplinary scientist. You always had time to talk and you supported me when it was needed. You created an atmosphere that cultivated the scientific spirit and motivated me (and others) to work as hard as possible to achieve the biggest goals, and this with a lot of fun. Your contribution to my PhD research and preparation of this booklet is invaluable. It is a big honor to learn from you. I hope you will still be my mentor in the next steps of this exciting scientific journey. Thank you for everything!

Dear **Frank**, during my last day of Erasmus+ internship you told me: "To hit a goal you need to shoot". This sentence encouraged me to pursue my scientific goals. Thank you for your wisdom, support and knowledge that you shared with me. You inspired me to study 3D chromatin organization and gene regulation, which is now my hobby. Dear **Petros**, I had the pleasure to work with you in multiple projects. I enjoyed our scientific discussions and learning from you. I am looking forward to our further collaboration! Dear **Raymond**, thank you for the opportunity to participate in your projects. You gave me the chance to discover a new field and expand my knowledge. Dear **Rutger**, thank you for training me in bioinformatics and sharing the scripts with me. Dear **Andrea**, it was a pleasure to learn from you. It is pity that you quit academia. Dear **Marike van Geest**, thank you for all the help and assistance, without you the department would become a pure chaos.

Dear **Ilias**, 'my masta' and dear friend, I started my PhD research working with you. Thank you for sharing with me the knowledge and lab tricks that made my life easier. We had a great time together! I would like to express my deepest gratitude for your support outside the lab, which helped me to start my PhD research in Netherlands. I hope to see you soon in Greece. Dear **Jente**, you are my biggest inspiration among all PhD candidates I met in my life. You were born to be a great scientist and good leader. I hope you will not quit academia! It was a lot of fun to discuss with you and **Ilias** all the projects. Wish you all the best, man! Dear **Anne**, I miss our chats in the office! It was a pleasure to work with you. You are a beautiful person with pure heart. Thank you for all the help and support. Dear **Iris**, we had a lot of fun discussing our projects. Thanks for the interesting ideas and great atmosphere at work. Dear **Judith**, you taught me a lot at the beginning of my PhD, which helped me to establish good habits in the lab. It was fun to work with you. Dear **Ridvan**, I do not know better strategist than you. I enjoyed our stimulating discussions and games. I hope you will visit me soon! Dear **Lucas**, you have always been the first person in the lab. I still try to learn this from you! Dear **Giulia**, you were responsible for the great atmosphere in the lab. Thank you for that. Dear **Manos**, είσαι ο αδερφός μου, we had a lot of fun together in Rotterdam,

"… very good, very good, bravo!". I am sure that we will reunite soon. Thank you for showing me Greece, which is now my second home. You will be a great scientist, I know it. Dear **Memnia**, I was very lucky to meet you. You are very wise and helpful person. Thank you for explaining me how my brain works. Dear **Roos**, without you the lab would be completely different. Discussions with you were unique, I loved them. I hope to see you soon! Also, I would like to express my gratitude to a number of other members of Cell Biology for a great atmosphere at work.

I would like to thank my inner doctorate committee: **G. Jansen, M.K. Koblowska** and **G.J.V.M. van Osch** for their time and valuable suggestions which helped to improve my PhD book. I would also like to thank the plenary committee: **T.S. Barakat, J. Vermeesch, K. Wendt** for their time.

I would like to thank **Wilfred van IJcken** and colleagues from the Biomics center of Erasmus MC, for valuable feedback and sequencing experiments.

MEDICAL UNIVERSITY OF LUBLIN:

Szanowny **Prof. Przemko Tylżanowski**, chciałbym uprzejmie podziękować za mentoring oraz możliwość realizacji rozprawy doktorskiej w ramach łączonego doktoratu. Koordynowanie interdyscyplinarnego projektu we współpracy z wieloma jednostkami naukowymi pozwoliła mi przygotować się na następny etap mojej kariery naukowej.

Szanowny **Prof. Andrzej Stepulak**, chciałbym uprzejmie podziękować za możliwość realizacji badań naukowych w zakładzie Biochemii i Biologii Molekularnej oraz za wsparcie, które otrzymałem w trakcie studiów doktoranckich.

Dear **Adolek**, I would like to thank you for the hours of discussions about scientific and non-scientific yet exciting topics. You are an inspiration for many employees and students at MUL.

Ewa, dziękuję Ci za cierpliwość oraz wspólnie przepracowane godziny w labie. Pokazałaś mi jak profesjonalnie podchodzić do rozwiązywania problemów. **Adrian,** dzięki Tobie moja wiedza z programowania poszerzyła się wielokrotnie. Twoja pomoc była kluczowa w moim przekwalifikowaniu. **Michał**, nasz czat to kopalnia złota, dziękuje za tysiące wiadomości. **Aneta**, dziękuję za pomoc w badaniach oraz przygotowaniu pięknej okładki. **Ania**, dziękuję za pomoc w pisaniu grantów w języku polskim. **Agnieszka Styczyńska**, dziękuję za pomoc w sprawach organizacyjnych. Chciałbym również podziękować **Ilonie, Kubie, Paulinie, Alicji, Asi, Marzenie, Karolinie, Lidce, Ewelinie, Pani Magdalenie, Pani Agnieszce** i wielu innym za świetną atmosferę w pracy.

OTHERS:

Bartosz Wojtaś, jesteś jedną z kluczowych osób, które pomogły mi stać się interdyscyplinarnym naukowcem. Dziękuję Ci za motywację, która napędzała mnie do dalszego rozwoju oraz wiedzę, którą mi przekazałeś.

Victor, dziękuję Ci za nieocenioną pomoc w procesie przekwalifikowania z „wet labu" w stronę „computational biology". Za kilka lat otworzymy wspólnie lab! **Basia, Krzysiek i Prof. Anna Gambin**, chciałbym wam serdecznie podziękować za możliwość współpracy nad TADeus2. Bardzo dużo się od was nauczyłem. Mam nadzieje, że uda nam się opracować wiele narzędzi, które będą ułatwiać pracę naukowcom.

Kochani **rodzice**, chciałbym bardzo Wam podziękować za ogrom pracy, który włożyliście w wychowanie i przygotowanie mnie do dorosłego życia. **Mamo**, zawsze byłaś moją siłą, moją przewodniczką. Nauczyłaś mnie nie ustępować i pokazałaś jak pokonać strach. Dałaś mi wszystko

czego potrzebowałem aby móc się rozwijać, poszerzać swoją wiedzę i umiejętności. Zawsze wierzyłaś we mnie i moje możliwości. Dziękuję Ci za to jak mnie wychowałaś, możesz być dumna z tego powodu. **Tato**, niejednokrotnie odmawiałeś sobie przyjemności, aby zapewnić mi możliwość studiowania czy wyjazdów za granicę. Bardzo to doceniam. Daliście mi wspaniały dom, do którego zawsze mogę wracać. Dzięki Waszemu zaangażowaniu i wsparciu mogłem realizować kolejne cele i osiągać sukcesy a teraz Wy możecie czytać te słowa w mojej rozprawie doktorskiej, która jest zwieńczeniem ostatnich lat mojej pracy. **Marcin**, dziękuję Ci za pomoc w wyjeździe do USA.

Ta część to podziękowania dla przyjaciół oraz znajomych, którzy towarzyszyli mi cały czas i od których miałem okazję nauczyć się wielu rzeczy. **Łukasz**, moja bratnia duszo, nikt nie rozumie mnie lepiej niż Ty. Dzięki za wszystkie wspólne chwile, dyskusje oraz za to, że zawsze jesteś szczery. Trzeba w końcu pojechać do Ameryki Południowej. **Piotrek**, potrafisz mnie zmotywować nawet na odległość. Setki godzin spędzonych na dyskusji miały duży wpływ na przygotowanie moich eksperymentów. **Danielu**, zawsze mogłem na ciebie liczyć. Wierzyłeś we mnie nawet w momentach, kiedy ja w siebie nie wierzyłem. Dzięki za wszystko! **Karolina**, byłaś dla mnie zawsze, kiedy tego potrzebowałem. Stymulowałaś mnie do zadawania pytań oraz przyczyniłaś się do tego, że lepiej rozumiem siebie. Byłaś dla mnie wsparciem, bez którego nie dałbym rady ukończyć doktoratu. Dziękuję Ci bardzo za wszystko! **Erbelku i Paweł**, uwielbiam grać z wami w planszówki. Pamiętajcie, „… tylko dzieci grają dla przyjemności …". **Bartek**, chętnie bym wrócił do czasu na studiach, kiedy siedzieliśmy razem w ławce na zajęciach. **Dominika**, dziękuję za wsparcie na wczesnym etapie mojego doktoratu i za to, że nigdy we mnie nie zwątpiłaś. **Damian**, mam nadzieje, że uda nam się reaktywować nasze hobby i że znajdziesz więcej czasu dla starego przyjaciela. Panie **Robercie**, od Pana nauczyłem się sprytu co pozwoliło mi się „przebić" z małej miejscowości do „wielkiego świata".