Review

Identification of the gene-regulatory landscape in skeletal development and potential links to skeletal regeneration

Hironori Hojo [a, *], Ung-il Chung [a, b], Shinsuke Ohba [a, b]

[a] Department of Bioengineering, The University of Tokyo Graduate School of Engineering, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[b] Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

A B S T R A C T

A class of gene-regulatory elements called enhancers are the main mediators controlling quantitative, temporal and spatial gene expressions. In the course of evolution, the enhancer landscape of higher organisms such as mammals has become quite complex, exerting biological functions precisely and coordinately. In mammalian skeletal development, the master transcription factors Sox9, Runx2 and Sp7/Osterix function primarily through enhancers on the genome to achieve specification and differentiation of skeletal cells. Recently developed genome-scale analyses have shed light on multiple layers of gene regulations, uncovering not only the primary mode of actions of these transcription factors on skeletal enhancers, but also the relation of the epigenetic landscape to three-dimensional chromatin architecture. Here, we review findings on the emerging framework of gene-regulatory networks involved in skeletal development. We further discuss the power of genome-scale analyses to provide new insights into genetic diseases and regenerative medicine in skeletal tissues.

© 2017, The Japanese Society for Regenerative Medicine. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Framework of gene regulation

The sequence from gene expression to protein translation (DNA to RNA to protein) comprises the central dogma of biology, as these phenomena collectively mediate the biological actions of cells. Gene expression is initiated through activation of the basal promoter, generally a DNA sequence less than 100 bp, in conjunction with the transcription start site (TSS). The basal promoters recruit RNA polymerases and basal transcription factors [1]. This action is necessary to initiate the process of transcription; however, it is not sufficient for the proper expression of the genes. Additional regulatory elements, called enhancers, are the main mediators specifying quantitative, spatial, and temporal regulation of the gene expression.

The first enhancer described was a 72-bp tandem repeat of SV40 DNA, which was shown to function as a cis-regulatory element activating the transcription of a cloned β-globin gene. This activation was observed even after changing the orientation or location of the sequence among several positions [2]. Currently, an enhancer is defined as a noncoding DNA sequence that can drive the target gene expression, regardless of the distance, location, or orientation of the sequence from the target basal promoter [3]. The size of enhancers ranges from 50 bp to 2 kb, and transcription factors bind to enhancers through specific DNA sequences called motifs, resulting in activation of the enhancers.

Although the basic machinery of the enhancer action is well conserved, the enhancer landscapes differ widely among species (Fig. 1; see the detail in Ref. [4]). In organisms with small genomes, such as bacteria and yeast, local regulatory controls are predominant, primarily through activation of the basal promoter and/or very few enhancers located mainly within 1 kb of the respective genes. Invertebrate species, such as *Drosophila*, have relatively complex machinery. Multiple enhancers targeting individual genes are often observed, although the distance of enhancer–promoter interactions is usually less than 10 kb. In contrast to these species, vertebrae have much more complex regulation for transcription. Enhancers can interact with target promoters over a range of 100 or even 1000 kb. Multiple enhancers can target genes even beyond the nearest genes. This complexity is probably due to the two rounds of whole genome duplication, which generate paralogous genes and additional regulators. Thus, *in vitro* studies such as reporter assays using several-kilobase-long genomic fragments around the basal promoter could cover gene-regulatory elements in invertebrates; however, capturing the vertebrate gene-regulatory landscape was difficult until next-generation sequencers became available.

Genome-scale studies using next-generation sequencers have provided novel insights into gene regulations, shifting interest from local gene regulations to multiple dimensions of the gene-regulatory landscape that integrates primary actions of the transcriptional regulators, the enhancer landscape, and the three-dimensional chromatin architecture (Fig. 2). Uncovering this landscape will provide new insights into the development,
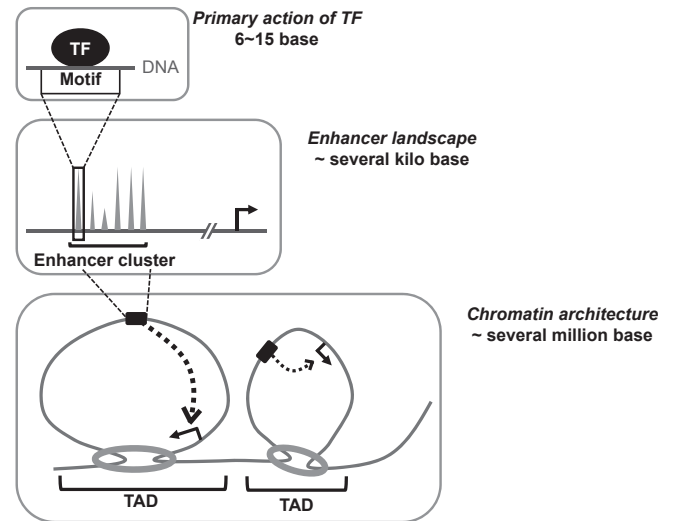

Fig. 2. Emerging framework of gene regulations. Transcription factors (TFs) prefer to bind to specific DNA sequences called motifs. Multiple associations of TFs to DNA occur at enhancers. On the genome, multiple enhancers form clusters; their distributions differ among cell types, representing cell type-distinct signatures. Topologically associating domains (TADs), which are defined as a three-dimensional chromatin landscape, have emerged as a new concept. Within a TAD, physical interactions of enhancers and promoters frequently occur, whereas these interactions are generally not observed across the boundary of TADs. In the boundary regions, CTCF and cohesin are thought to be involved in the looping of the chromatin structure.

evolution and pathologies of organisms. In this review, we summarize the emerging gene-regulatory networks involved in the development of skeletal tissues, which both create a supportive framework and systematically control the supply of minerals for the body.

2. Primary mode of action of key transcription factors in skeletal development

The mammalian skeleton is derived from cells of three origins: the neural crest, paraxial mesoderm, and lateral plate mesoderm. It is formed through two distinct modes of ossification: intermembrane ossification and endochondral ossification (see review [5]). Regardless of the origin or the mode of ossification, key transcription regulators are crucial for the specification and differentiation of the skeletal cells. So far, three transcription factors, Sox9, Runx2 and Sp7/Osterix, have been identified as master regulators in skeletal development [6–10], whereas others were identified mainly as co-regulators of these master regulators that modify their functions or expressions at either the transcriptional or protein level [11–13]. In this section, we briefly summarize the biological roles of these master regulators in skeletal development and discuss the gene-regulatory networks that have emerged from recent genome-scale studies, particularly focusing on how the master regulators program the regulatory networks cooperatively with co-regulators.

2.1. Sox9-mediated chondrocyte regulatory program

Sox9 is a high-mobility group (HMG) domain-containing transcription factor, closely related to the Y chromosome-encoded testis-determining factor SRY. Mutations in *SOX9* are associated with campomelic dysplasia, in which both sex and skeletal development are affected [14]. During skeletal development, Sox9 is expressed in mesenchymal condensation; the Sox9-positive cells in the condensation give rise to osteoblasts, chondrocytes, tendon
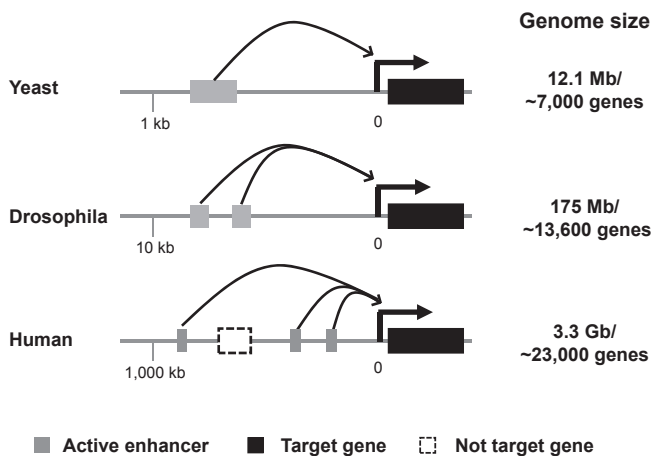
Fig. 1. Enhancer landscape in various species. A schematic model of the gene-regulatory landscape and genome size information of yeast, *Drosophila*, and humans. In yeast, many genes are regulated only by activation of the basal promoter. A few enhancers are reported to control the target gene, but most of the enhancers are located within a 1-kb range of the basal promoter. In *Drosophila*, multiple enhancers work together to control a target gene, although the distance of enhancer–promoter interactions is usually less than 10 kb. In mammals, including human, enhancers can interact with target promoters over a range of 100 or even 1000 kb. The interactions can occur beyond several of the nearest genes. Multiple enhancers likely control quantitative, temporal, and spatial aspects of gene expression.

cells, and synovium cells [15]. After the stage of mesenchymal condensation, Sox9 expression is restricted to the chondrocyte lineage. Sox9 is expressed in most chondrocytes throughout their differentiation except at the terminal differentiation stage. As expected from the expression pattern in skeletal development, Sox9 activity is essential throughout the chondrocyte regulatory program: Sox9 is necessary for mesenchymal condensation and subsequent chondrocyte survival, proliferation, and hypertrophy [7,16–19].

Sox9 target genes were identified by genomic and biochemical approaches. The analyses identified Sox9-associated cis-regulatory regions flanking *Sox9* itself and genes encoding cartilage extracellular matrix proteins and cartilage growth factors [20,21]. In addition, Sox9 ChIP-seq analyses were recently reported by three independent groups including the present authors [22–24]. All of these studies provided data sets of putative chondrocyte enhancers associated with Sox9; some were verified by an *in vivo* reporter assay in zebrafish [24] or a gene expression analysis in Sox9-deficient mice [23]. The Sox9 targets included not only major cartilage matrix genes, but also signaling pathway components including *Fgfr2* and *Fgfr3*, and transcriptional regulators including *Arid5a*, *Runx2*, and *Runx3*. These findings expand our understanding of Sox9-mediated gene-regulatory networks in chondrogenesis.

The Sox9 ChIP-seq studies provided additional insights into the mode of Sox9 action in the chondrocyte regulatory program [24]. Previous analysis of the Sox9 peak distribution revealed two classes of Sox9 activity on the chondrocyte genome: Class I and Class II (Fig. 3a). Class I was defined as a TSS-associated Sox9-action within a 500-bp window from the nearest TSS [24]. In this class, Sox9 had positive impacts on the target gene expression, likely through protein–protein interactions between Sox9 and components of the basal transcriptional complex [25,26]. Class I Sox9 action is associated with highly expressed genes including housekeeping genes, not specifically with chondrocyte-related genes [24].

Class II was defined as the distal action of Sox9, beyond the ±500-bp range of the nearest TSS [24]. Several lines of evidence indicate that Class II action is the predominant mode of Sox9 action for the differentiation and maintenance of chondrocytes [24]. (1) More than 75% of Sox9 peaks are categorized as Class II. (2) Unlike Class I, the Class II action enriches peaks that are significantly associated with chondrocyte-related genes. (3) The peak intensity in Class II is higher than that in Class I. (4) Sox9 consensus motifs are highly enriched in Class II, but not in Class I. (5) Many Class II regions form super-enhancer-like clusters. The super-enhancer is defined as a cluster of enhancers that is densely occupied by transcription factors around key genes for cell identities [27] (Fig. 3b). The presence of these enhancer modules is significantly associated with highly expressed genes compared to typical enhancers in chondrocytes [24]. Taken together, these facts indicate that the distal action of Sox9 represents distinct signatures of its cell type-dependent actions, i.e., the execution of a chondrocyte program in this context.

The motif recovery analysis with Sox9 ChIP-seq data reveals that the Sox9 action occurs mainly through the Sox dimer motif oriented head-to-head with a 3- or 4-nucleotide spacer [23,24]. This DNA preference was also evidenced by several *in vitro* studies [21] including high-throughput protein-DNA interaction analyses, such as, analysis of the systematic evolution of ligands by exponential enrichment (SELEX) [28]. Interestingly, the enriched dimer motifs have more than a few variations at each nucleotide position within the predicted sequence set; many of the actual nucleotide sequences in Sox9 peaks were distinct from the sequence, perfectly matching with the predicted optical dimer motif with statistical significance [24]. Together with the formation of super-enhancer-like clusters, this finding suggests that Sox9 favors multiple, low-affinity binding sites. The low affinity of Sox9 to the endogenous binding site may provide the Sox9-DNA interaction with flexibilities, which maximize the biological outcome from Sox9
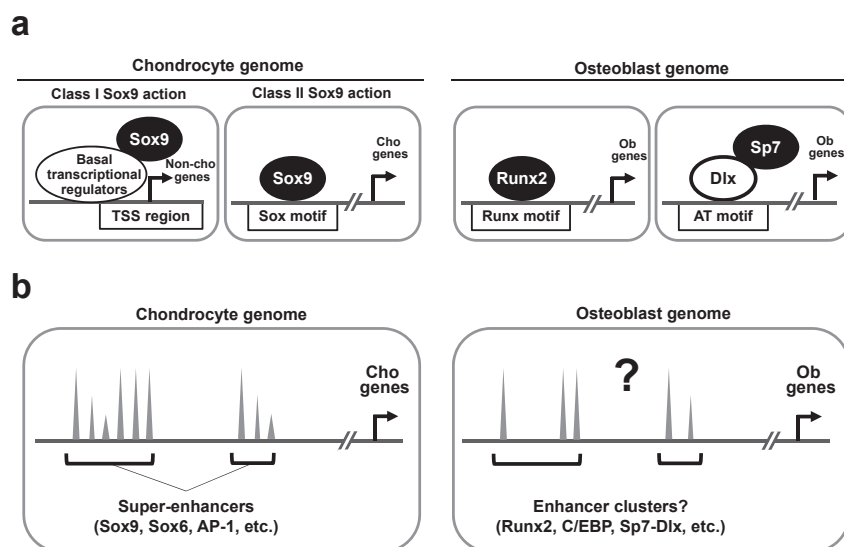

Fig. 3. Gene-regulatory networks programming skeletal development revealed by recent ChIP-seq studies. (a) Primary modes of action of master transcription factors in skeletal development. Chondrocyte determinant Sox9 has two distinct modes of action. Class I is a TSS-associated Sox9-action, in which Sox9 interacts with basal transcriptional regulators around TSS regions to broadly enhance transcription. In Class II action, Sox9 binds to Sox dimer motifs to regulate cartilage-related genes. Runx2 and Sp7/Osterix are osteoblast determinants. Runx2 mainly binds to the consensus Runx motif on the osteoblast genome. Our recent Sp7 ChIP-seq study proposed a new mode of Sp7 action; Sp7 acts as a co-factor of Dlx by indirectly associating with the genome through AT-rich motifs. TSS, transcription start site; Cho, chondrocyte; Ob, osteoblast. (b) Emerging epigenetic landscape of skeletal cells. Sox9-bound distal enhancers form super-enhancer-like clusters on the chondrocyte genome, representing chondrocyte signatures. Sox6 is involved in forming the clusters. AP-1 may also be involved in this context. In contrast, the gene-regulatory landscape comprised of Runx2 and Sp7 in osteoblasts remains to be revealed, although individual genome-scale associations of the two key determinants have been studied. It has not been clarified whether, like chondrocyte regulators, their binding regions are associated with enhancer clusters. One study revealed co-association of Runx2 and CCAAT/enhancer-binding proteins (C/EBP) in the common genomic regions in osteoblasts.

co-engagement with other transcriptional regulators. The formation of a transcriptional complex with co-regulators could change the preferences of the primary binding motifs [29].

A number of Sox9 co-regulators in chondrocytes have been identified by genetics, proteomics, and *in vitro* studies [11]. Among them, the modes of action of Sox6 and AP-1 were further examined in genome-scale analyses [18,23]. Sox6 is known to act as a "Sox-trio" with Sox9 and Sox5. The amino acid sequences of the Sox6 HMG domain have 50% identity with that of Sox9 HMG [11]. Unlike Sox9, however, Sox6 does not have a transactivation domain [11]. A Sox6 ChIP-seq study in a rat chondrosarcoma cell line (RCS cells) demonstrated that about half of the Sox6-binding regions were shared with those of Sox9 on the genome, suggesting cooperative actions in the shared regions [23]. However, the motif recovery analysis with the Sox6 ChIP-seq data revealed the enrichment of distinct motifs from Sox9; Sox6 favors the monomer Sox motif [23]. Thus, it still remains to be clarified how Sox9 and Sox6 cooperate in regulating the common target genes at the molecular level. Given that the protein–protein interaction of these factors has not yet been reported, they may have an indirect interaction through other, unknown transcriptional regulators, or they may bind to different motifs individually within common enhancers.

AP-1 was predicted to act as a co-engagement factor in conjunction with Sox9 by the analysis of motifs enriched in the Class II Sox9 ChIP-seq peak region [24]. A highly enriched AP-1 motif suggested interactions between Sox9 and members of the AP-1 family [24]. These interactions were confirmed and analyzed in our recent report [18]. The results showed that most Sox9 binding occurs in conjunction with Jun; Jun binds either directly to the AP-1 motif on the genome or indirectly in a complex with Sox9 [18]. Given that both Jun and Fosl2 are highly expressed within pre-hypertrophic and hypertrophic chondrocytes, the study suggests that these cooperative actions are involved in chondrocyte hypertrophy [18].

2.2. Runx2-mediated osteoblast regulatory program

Runx2 and Sp7/Osterix are master regulators in osteoblasts. *Runx2*- and *Sp7*-null mutant mice lack all bones [8,10]. Mutations of these genes in humans underlie genetic diseases in the skeleton. Mutations in *RUNX2* and *SP7* are associated with cleidocranial dysplasia and osteogenesis imperfecta, respectively [9,30,31]. The expression pattern of these factors defines distinct stages of the osteoblast differentiation. Skeletal progenitors are initially committed to Runx2-positive osteoblast precursors; the precursors then transition to Runx2- and Sp7-double-positive osteoblast precursors [32]. These precursors and their descendent give rise to osteoblasts and osteocytes, coupled with the invasion of blood vessels for generating bone marrow [33]. Consistent with the cell lineage hierarchy, Runx2 is genetically upstream of Sp7 [10].

Runx2 belongs to the runt-related transcription factor family. The primary action of Runx2 is mediated by direct binding to DNA through the Runx consensus motif. The mode of Runx2 action was initially linked to osteoblasts through the discovery of osteoblast-specific cis-regulatory elements in the *Bglap* promoter [34,35]. Prior to the emergence of genome-wide analyses, biochemical studies identified a number of Runx2 targets including *Runx2* itself, *Alpl*, *Ibsp*, and *Spp1* [36]. Runx2 ChIP-seq studies in osteoblastic cell lines were recently conducted by three independent groups and provided further understanding of the Runx2-mediated osteoblast gene-regulatory networks [37–39].

First, the genome-scale studies uncovered new Runx2-mediated cis-regulatory regions [37–39]. Notably, they identified distal cis-regulatory elements flanking *Mmp13* and *Spp1*. Although these genes were known to have proximal Runx2-binding regions [40,41],

the analysis revealed that the Runx2-DNA interaction signal was more intense in these distal regions than in the proximal regions [39,42]. A recently published comparative analysis of the putative *Mmp13* promoter and enhancers further highlighted the importance of the distal enhancers; the distal *Mmp13* enhancer had strong responsiveness of the reporter activation to Runx2 expression and *Mmp13* expression was dramatically suppressed by the deletion of the distal enhancer using Crispr/Cas9 technology *in vitro* [42].

Second, examining the Runx2-binding regions in various stages of osteoblast differentiation provided insight into a dynamic change of Runx2 action during differentiation. One study performed clustering analysis for Runx2 binding in conjunction with transcript profiles at various time points of an *in vitro* differentiation culture of the pre-osteoblastic MC3T3-E1 cell line [38]. The results showed that a subset of Runx2 peaks was highly associated with osteoblast-related genes that were induced upon osteoblast induction [38]. Another cluster of Runx2 peaks, which were lost upon osteoblast induction, were related to biological functions in other cell lineages, including fat cell differentiation, leukocyte migration, and erythrocyte differentiation [38]. These findings suggest that Runx2 may have broader interactions prior to osteogenesis or that binding at non-osteoblast targets may suppress non-osteogenic pathways for the cell commitment [38]. Further functional analyses are required to elucidate the inhibitory action of Runx2.

Third, *de novo* motif analyses revealed a consensus Runx motif as the most enriched sequence in the Runx2 peaks, which was consistent with the findings of a number of biochemical studies that Runx2 binds the consensus Runx motif [37–39]. One of these studies also revealed the co-engagement of C/EBPβ at the enhancer modules [39]. Runx2 and C/EBPβ form a transcriptional complex and function together in transactivation [43,44]. Given that a number of other regulators of transcription are thought to cooperate with Runx2, including Cbfb, Twist, Stat1, Schnurri3, SATB2, TAZ, and Zfp521 (see reviews [13,45,46]), further genome-scale approaches will help to illuminate these cooperative actions in the context of the gene regulatory networks in osteogenesis.

2.3. Sp7/Osterix-mediated osteoblast regulatory program

Sp7 was identified as a member of the Sp family of transcription factors based on the high conservation of the amino acid sequences in the zinc finger domain. All members of this family, including the well-studied family founder Sp1, were thought to bind to DNA through the zinc finger domain at an Sp consensus GC-box [47]. However, our recent Sp7 ChIP-seq study in primary osteoblasts demonstrates a mode of Sp7 action that is distinct from the previous model. *De novo* motif analysis identified an AT-rich motif containing a homeodomain-response element as the most enriched sequence in the Sp7 ChIP-seq data, whereas no enrichment of the GC-box was observed in the data set [48]. This contrast was further highlighted in a comparative analysis between the Sp1 ChIP-seq data set and the Sp7 data set in the pre-osteoblastic cell line MC3T3E1 [48]. The motif analysis confirmed that Sp1 favors the GC-box, whereas no AT-rich motif was enriched in the Sp1 peaks in the osteoblastic cell. In contrast, the AT-rich motif was the most enriched in the Sp7 peaks. Peak distribution analysis further revealed that Sp1 was mainly associated with proximal regions from a TSS targeting housekeeping genes, whereas Sp7 was mainly associated with distal genomic regions targeting ossification-related genes. These indicate that Sp7 has a mode of action distinct from those of other members of the Sp family, in terms of both genomic targets and the associated motif.

Several lines of evidence suggest that Sp7 binds indirectly to the AT-rich motif by forming a transcriptional complex with the Dlx

family of homeodomain-containing transcription factors, Dlx3, 5 and 6 [48]. First, these Dlx factors are highly expressed in osteoblasts and are known to regulate osteoblast differentiation [49–51]. Second, a genome-scale analysis revealed that most of the Sp7-associated genomic regions were shared with Dlx5-associated regions, where the AT-rich sequence was the most enriched motif [48]. Third, in in vitro studies, Sp7 alone did not bind to the AT-rich motif, whereas a complex of Sp7 and Dlx5 did [48]. Forth, knock-down of Dlx factors in osteoblasts attenuated Sp7 engagements with the genome at the osteoblast enhancers, resulting in suppression of the Sp7 target gene expression [48]. All of the data support the notion that Sp7 acts as a Dlx co-factor, not as a transcription factor in osteoblasts. However, it remains to be clarified whether Sp7 acts as a co-factor only for Dlx factors. Msx1/2, Satb2 and Alx4 are all highly expressed in osteoblasts [48] and possibly bind to AT-rich motifs in a similar manner based on the motif database [52]. Given that Sp7 is expressed in pre-hypertrophic chondrocytes and osteocytes as well as osteoblasts, its partner may be different in distinct cell types.

Cross-species analysis also suggests an evolutionary link between Sp7's distinct mode of action and the origin of bone-forming vertebrae. In vitro studies indicated that the zinc finger domain of mouse Sp7 has three different amino acids from those of other Sp family members, and the variation contributes to Sp7's distinct role as an osteoblast determinant [48]. Cross-species analyses comparing the sequences among different vertebrate and non-vertebrate chordate species showed that the closest non-boney vertebrates (e.g., lampreys), the cephalochordates (e.g., amphioxus), and the ascidians (e.g., tunicate) all lack an Sp7-type zinc finger variant in their genome [48]. In contrast, the Sp7 gene is present in all boney vertebrate groups [48]. Considering these facts together, the emergence of this particular Sp-family variant is likely coupled to the evolution of species with boney vertebrae. Dlx-related genes were likely already present, based on the identification of paralogs in non-vertebrate chordates. Thus, Sp7 may have functioned to stabilize or enhance the transcriptional activation of Dlx in osteoblast targets within the skeletal regulatory genome.

The above study not only provides genome-level evidence for known Sp7 targets, but also identified novel Sp7 targets in osteoblasts [48]. Sp7-associated cis-regulatory regions flanking Notch2, Runx2, and Col1a1 have been revealed as skeletal enhancers by analyzing transgenic mice that are engineered to express reporter genes under the control of those cis-regulatory regions [48,53,54]. Notably, analyses of site-directed mutations in the AT-rich motifs within these elements confirmed the suppression of the in vivo enhancer activity of the Runx2-flanking region [53] and the Col1a1-flanking region [55], and the in vitro activity of the Notch2-flanking region [48], supporting the crucial contribution of AT-rich motifs to the activity of these enhancers where Sp7 is possibly involved.

The target gene analysis also implicates Sp7 in the cross-talk involved in the regulation of Notch, Hedgehog, Fgf, and canonical Wnt pathways [12]. Putative Sp7 target regions are found around Notch2, Gli2 (a transcription factor downstream of hedgehog signaling), Fgfr2, and Kremen1 (a modulator of the canonical Wnt pathway [56]), predicting Sp7-mediated tuning of these signaling pathways at the transcriptional level.

In summary, emerging genome-scale analyses have been providing new insights into the gene-regulatory networks in skeletal development and the mode of action of key skeletal determinants (Fig. 3A). However, current knowledge is limited to the genomic landscape programmed by a few key transcriptional regulators. Knowledge of the sequential and/or developmental stage-specific activity of signaling pathways is essential to understanding the process of skeletal development (for review, see Refs. [12,57]). Important questions remain to be answered: Namely, how do these signaling pathways interface with the regulatory networks controlled by the key skeletal determinants on the genome? To answer this, additional ChIP-seq studies and integrative analyses with all of these data sets will be required.

3. Epigenetic landscape

As we described early, transcription factors favor specific DNA sequences in the enhancer regions. However, a transcription factor's accessibility to and affinity for the motif is not equal throughout the genome. One example is Sox9 binding to the consensus motif in chondrocytes: the Sox dimer motif-like sequence is located in more than one million sites throughout the mouse genome; however, statistically enriched Sox9 binding to the genome through the motif was observed in less than 1% of them based on prior Sox9 ChIP-seq data [24]. One possible mechanism explaining the discrepancy is that epigenetic regulation restricts the action of transcription factors on the genome [58,59]. The epigenetic landscape is mainly constituted by histone modifications and DNA methylation. Importantly, the landscape is variable among cell-types, tissues and contexts, defining distinct enhancer states [58,59]. The Roadmap Epigenomics Project has produced references for the epigenetic landscape obtained from many tissues and cell types in mammals, mainly from humans [60]. These data provide information on key functional elements controlling context-dependent gene expressions by a comparative analysis across cell types and tissues.

The epigenetic landscape is also informative to identify non-coding regions that possibly link to genetic diseases. One example is studies of cis-regulatory elements controlling Sox9 transcription through the genome-scale examination of epigenetic markers in mouse embryonic limbs [61,62]. Importantly, some of the identified regulatory regions coincided with regions where genomic translocation, deletion or duplication are frequently observed in campomelic dysplasia and Pierre Robin syndrome patients [61,62]. Further loss-of-function studies for the identified loci, or more specifically for the Sox9-binding motif within the loci, will provide insight into the molecular mechanism of these diseases as well as clinically relevant information.

Comparative analyses of the epigenetic landscape and transcript profiles before and after in vitro osteoblast induction revealed new transcriptional regulators in osteoblasts [37]. Specifically, distinct motifs for TEAD2 (TEA domain family member 2) and GTF2I were enriched in the H3K4me3-positive promoter regions in differentiated osteoblasts but not in those in the immature cells [37]. Loss of function of these regulators leads to impaired mineralization of osteoblasts in vitro, suggesting the involvement of these regulators in osteoblast maturation [37]. Further in vivo data will be important in understanding their contribution to physiological bone formation.

Although the epigenetic landscape features distinct cell types, this landscape is flexible in some contexts. Meyer et al. revealed that the epigenetic landscape representing osteoblast signatures can be changed to that of adipocytes consistent with the alteration of the transcript profiles from osteoblasts to adipocytes [63]. Such plasticity of the epigenetic landscape is also observed in macrophages, whose cell-type-specific epigenetic landscape is likely defined by the microenvironment [64]. Understanding the plasticity of skeletal cells could provide insight into the cellular mechanisms underlying the reprogramming of cell fates and the potency of differentiated cells as a source of tissue regeneration.

One current limitation of the field is the use of mixed cell populations in the acquisition of data sets. For example, the primary chondrocyte population obtained without any cell-sorting process possibly includes mitotic chondrocytes, post-mitotic

pre-hypertrophic chondrocytes or even some different cell types such as fibroblasts, even though the isolation procedure is well established. Flow cytometry analysis with primary osteoblasts isolated from Sp7-GFP reporter mice showed that only about 20% of cells were positive for GFP, indicating that the other 80% comprise different cell types and/or different stages of osteoblasts [48]. Thus, sorting a specific population by marking cells of interest with reporter proteins will precisely reveal the cell type-specific signatures. In addition, the recently developed single cell analysis [65,66] will be a powerful tool to examine the heterogeneity of any given cell type. Although *in vitro* cell culture helps to provide homogeneous populations, no global analysis has been performed comparing epigenetic data sets from *in vitro* cultures and those from *in vivo* primary cells. Thus, when a data set is obtained from mixed cell populations *in vivo* or from cell cultures, *in vivo* confirmation analyses are required to draw specific conclusions.

As ENCODE project pioneered and others including international consortia continue, more epigenetic data are becoming available. Given that various data sets are obtained from several species, including humans, mice and rats, comparative analysis among species will provide insights into the onset of heretofore unknown gene-regulatory networks as well as their conservation in evolution. Based on a comparative analysis of gene-regulatory networks between mice and humans, Stergachis et al. proposed an evolutionary selection of mammalian gene regulation at the level of trans-regulatory circuitry [67]. They revealed that although only 5% of individual DNA bases are conserved between the two species, more than 90% conservation was observed in the regulatory network architectures comprised of cell type-specific key transcriptional regulators [67]. Using iPS technology, Prescott et al. performed a comparative analysis between humans and chimpanzees of the epigenetic landscape in the cranial neural crest cells and revealed species-biased enhancer clusters near loci affecting intra-human facial variations [68]. Such cross-species analyses will be powerful tools for identifying the origin of skeletal gene-regulatory networks.

4. Higher-order chromatin architecture and molecular links to genetic diseases

Three-dimensional chromatin architecture modeling has provided a new framework for understanding gene-regulatory networks. Recently developed methods of chromosome conformation capture analyses such as Hi-C [69] have revealed that the chromosome is segmented into self-interacting domains called topologically associated domains (TADs; Fig. 2). TAD boundaries restrict the interaction between cis-regulatory elements and the target basal promoter within the TAD. Lupianez et al., demonstrate that the structural variant of the genome leading disruption of TAD boundaries underlie genetic diseases [70]. They identified TAD boundaries flanking *EPHA4* gene; the disruption of these was associated with congenital limb malformation in humans [70]. By utilizing CRISPR/Cas9-mediated genome editing strategy, they created mutant mice in which the genome structural variant in human were recapitulated in mice. The mutant mice showed the similar pathological phenotypes to human diseases, accompanied with the abnormality of enhancer–promoter interactions and ectopic expression of the adjacent genes [70]. One example here is that the disruption of a TAD boundary caused the ectopic interaction of the *Eph4a* enhancer with the adjacent *Ihh* promoter, not with the *Eph4a* promoter, resulting in polydactyly due to the ectopic activation of *Ihh* gene [70]. These studies highlight the molecular links between higher-order-gene regulatory network and genetic diseases.

5. Perspectives for associating the gene-regulatory landscape with human diseases and regenerative medicine in skeletal tissues

As discussed in this review, gene-regulatory networks are well organized through a different dimension of the regulatory mechanism. Since any of these variations can underlie disorders, it will be informative to investigate the gene-regulatory networks involved in disease-associated genetic variants. Genome-scale sequence studies, such as whole-genome sequencing and genome-wide association studies, are necessary to identify disease-associated genomic variations in different individuals, mainly focusing on single-nucleotide polymorphisms [71–74]. Since more than 90% of disease-associated loci identified so far by the analysis are localized outside of protein-coding regions, the regulatory genes in non-coding regions that might possibly influence enhancer-driven transcription are the most plausible targets of interest in next-generation genomics. Regulatory components in the genome are thought to contain distinct interaction sites for key regulatory factors, specific epigenetic states in these sites, or the boundary region of the genome. Thus, defining the regulatory components in the genome will facilitate the identification of non-coding sequence variants that are associated with regulatory mechanisms underlying human development and disease [75].

From the viewpoint of regenerative medicine, studies of gene regulatory network help us better understand the molecular mechanism of tissue regeneration and gain insights into the application of the network to clinical settings. For example, Kang et al. identified a regeneration-related enhancer element targeting *leptin b* gene. The enhancer element was conserved between fish and mammals and activated during tissue regeneration in injury models in zebrafish fin, heart and mouse fingers [76]. The enhancer element enabled not only to target regenerating tissues, but also to manipulate the potency of the regeneration by modulating signaling pathway components [76]. Similarly, Guenther et al. identified a distinct regulatory region of the *Bmp5* locus, which was activated upon several injury models [77]. Interestingly, a series of Bmp5 enhancers are locally activated in different tissues in conjunction with endogenous *Bmp5* expression during developmental stages [77]. However, in injury models including bone fracture, skin wound, and lung injury, *Bmp5* expression is induced through the common enhancer region that was distinct from the other embryonic enhancers [77]. Thus, identifying context-distinct gene regulatory elements during tissue regeneration may be a key to realize regenerative therapies through the manipulation of genes or signaling pathways that enhance tissue regeneration in context-distinct manners.

In addition to identifying key genomic regions, the recently developed Crispr/Cas9 technology enables the manipulation of epigenetic states at given genomic loci. For example, guide-RNA-targeting of specific loci with nuclease-null, deactivated Cas9 (dCas) fused to the catalytic core of the acetyltransferase p300 specifically activated the targeted enhancers [78], whereas that with dCas fused to histone demethylase suppressed their activity [79]. Thus, the manipulation of activities of key gene-regulatory elements, such as disease-associated loci and bone regeneration-associated loci, may be an attractive new approach to gene therapy for genetic disorders and regenerative medicine in skeletal tissues.

Acknowledgments

and the Core-to-Core Program Advanced Research Networks from the Japan Society for the Promotion of Science, and the Uehara Memorial Research Foundation.

References

[1] Zabidi MA, Stark A. Regulatory enhancer–core-promoter communication via transcription factors and cofactors. Trends Genet 2016;32:801–14.

[2] Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 1981;27:299–308.

[3] Schaffner W. Enhancers, enhancers – from their discovery to today's universe of transcription enhancers. Biol Chem 2015;396:311–27.

[4] de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 2013;502:499–506.

[5] Berendsen AD, Olsen BR. Bone development. Bone 2015;80:14–8.

[6] Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B. Sox9 is required for cartilage formation. Nat Genet 1999;22:85–9.

[7] 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;16:2813–28.

[8] Komori T, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, et al. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell 1997;89:755–64.

[9] Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, et al. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell 1997;89:765–71.

[10] Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, et al. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. Cell 2002;108:17–29.

[11] Liu CF, Samsa WE, Zhou G, Lefebvre V. Transcriptional control of chondrocyte specification and differentiation. Semin Cell Dev Biol 2017;62:34–49.

[12] Long F. Building strong bones: molecular regulation of the osteoblast lineage. Nat Rev Mol Cell Biol 2012;13:27–38.

[13] Karsenty G. Transcriptional control of skeletogenesis. Annu Rev Genomics Hum Genet 2008;9:183–96.

[14] Bi W, Huang W, Whitworth DJ, Deng JM, Zhang Z, Behringer RR, et al. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A 2011;98:6698–703.

[15] Akiyama H, Kim JE, Nakashima K, Balmes G, Iwai N, Deng JM, et al. Osteochondroprogenitor cells are derived from Sox9 expressing precursors. Proc Natl Acad Sci U S A 2005;102:14665–70.

[16] Ikegami D, Akiyama H, Suzuki A, Nakamura T, Nakano T, Yoshikawa H, et al. Sox9 sustains chondrocyte survival and hypertrophy in part through Pik3ca-Akt pathways. Development 2011;138:1507–19.

[17] Dy P, Wang W, Bhattaram P, Wang Q, Wang L, Ballock RT, et al. Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. Dev Cell 2012;22:597–609.

[18] He X, Ohba S, Hojo H, McMahon AP. AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. Development 2016;143:3012–23.

[19] Akiyama H, Lyons JP, Mori-Akiyama Y, Yang X, Zhang R, Zhang Z, et al. Interactions between Sox9 and beta-catenin control chondrocyte differentiation. Genes Dev 2004;18:1072–87.

[20] Akiyama H, Lefebvre V. Unraveling the transcriptional regulatory machinery in chondrogenesis. J Bone Min Metab 2011;29:390–5.

[21] Lefebvre V, Dvir-Ginzberg M. SOX9 and the many facets of its regulation in the chondrocyte lineage. Connect Tissue Res 2017;58:2–14.

[22] Oh CD, Lu Y, Liang S, Mori-Akiyama Y, Chen D, de Crombrugghe B, et al. SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. PLoS ONE 2014;9:e107577.

[23] Liu CF, Lefebvre V. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. Nucleic Acids Res 2015;43:8183–203.

[24] Ohba S, He X, Hojo H, McMahon AP. Distinct transcriptional programs underlie Sox9 regulation of the mammalian chondrocyte. Cell Rep 2015;12:229–43.

[25] Tsuda M, Takahashi S, Takahashi Y, Asahara H. Transcriptional co-activators CREB-binding protein and p300 regulate chondrocyte-specific gene expression via association with Sox9. J Biol Chem 2003;278:27224–9.

[26] Furumatsu T, Tsuda M, Yoshida K, Taniguchi N, Ito T, Hashimoto M, et al. Sox9 and p300 cooperatively regulate chromatin-mediated transcription. J Biol Chem 2005;280:35203–8.

[27] Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 2013;153:307–19.

[28] Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. Cell 2013;152:327–39.

[29] Deplancke B, Alpern D, Gardeux V. The genetics of transcription factor DNA binding variation. Cell 2016;166:538–54.

[30] Mundlos S, Otto F, Mundlos C, Mulliken JB, Aylsworth AS, Albright S, et al. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. Cell 1997;89:773–9.

[31] Lapunzina P, Aglan M, Temtamy S, Caparros-Martin JA, Valencia M, Leton R, et al. Identification of a frameshift mutation in Osterix in a patient with recessive osteogenesis imperfecta. Am J Hum Genet 2010;87:110–4.

[32] Rodda SJ, McMahon AP. Distinct roles for Hedgehog and canonical Wnt signaling in specification, differentiation and maintenance of osteoblast progenitors. Development 2006;133:3231–44.

[33] Maes C, Kobayashi T, Selig MK, Torrekens S, Roth SI, Mackem S, et al. Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. Dev Cell 2010;19:329–44.

[34] Ducy P, Karsenty G. Two distinct osteoblast-specific cis-acting elements control expression of a mouse osteocalcin gene. Mol Cell Biol 1995;15:1858–69.

[35] Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. Cell 1997;89:747–54.

[36] Stein GS, Lian JB, van Wijnen AJ, Stein JL, Montecino M, Javed A, et al. Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. Oncogene 2004;23:4315–29.

[37] Hakelien AM, Bryne JC, Harstad KG, Lorenz S, Paulsen J, Sun J, et al. The regulatory landscape of osteogenic differentiation. Stem Cells 2014;32:2780–93.

[38] Wu H, Whitfield TW, Gordon JA, Dobson JR, Tai PW, van Wijnen AJ, et al. Genomic occupancy of Runx2 with global expression profiling identifies a novel dimension to control of osteoblastogenesis. Genome Biol 2014;15:R52.

[39] Meyer MB, Benkusky NA, Pike JW. The RUNX2 cistrome in osteoblasts: characterization, down-regulation following differentiation, and relationship to gene expression. J Biol Chem 2014;289:16016–31.

[40] Sato M, Morii E, Komori T, Kawahata H, Sugimoto M, Terai K, et al. Transcriptional regulation of osteopontin gene in vivo by PEBP2alphaA/CBFA1 and ETS1 in the skeletal tissues. Oncogene 1998;17:1517–25.

[41] Wang X, Manner PA, Horner A, Shum L, Tuan RS, Nuckolls GH. Regulation of MMP-13 expression by RUNX2 and FGF2 in osteoarthritic cartilage. Osteoarthr Cartil 2004;12:963–73.

[42] Meyer MB, Benkusky NA, Pike JW. Selective distal enhancer control of the Mmp13 gene identified through clustered regularly interspaced short palindromic repeat (CRISPR) genomic deletions. J Biol Chem 2015;290:11093–107.

[43] Hirata M, Kugimiya F, Fukai A, Ohba S, Kawamura N, Ogasawara T, et al. C/EBPbeta Promotes transition from proliferation to hypertrophic differentiation of chondrocytes through transactivation of p57. PLoS ONE 2009;4:e4543.

[44] Tominaga H, Maeda S, Hayashi M, Takeda S, Akira S, Komiya S, et al. CCAAT/enhancer-binding protein beta promotes osteoblast differentiation by enhancing Runx2 activity with ATF4. Mol Biol Cell 2008;19:5373–86.

[45] Komori T. Requisite roles of Runx2 and Cbfb in skeletal development. J Bone Min Metab 2003;21:193–7.

[46] Long F, Ornitz DM. Development of the endochondral skeleton. Cold Spring Harb Perspect Biol 2013;5:a008334.

[47] Kadonaga JT, Jones KA, Tjian R. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem Sci 1986;11:20–3.

[48] Hojo H, Ohba S, He X, Lai LP, McMahon AP. Sp7/Osterix is restricted to boneforming vertebrates where it acts as a Dlx co-factor in osteoblast specification. Dev Cell 2016;37:238–53.

[49] Hassan MQ, Javed A, Morasso MI, Karlin J, Montecino M, van Wijnen AJ, et al. Dlx3 transcriptional regulation of osteoblast differentiation: temporal recruitment of Msx2, Dlx3, and Dlx5 homeodomain proteins to chromatin of the osteocalcin gene. Mol Cell Biol 2004;24:9248–61.

[50] Li H, Marijanovic I, Kronenberg MS, Erceg I, Stover ML, Velonis D, et al. Expression and function of Dlx genes in the osteoblast lineage. Dev Biol 2008;316:458–70.

[51] Tadic T, Dodig M, Erceg I, Marijanovic I, Mina M, Kalajzic Z, et al. Overexpression of Dlx5 in chicken calvarial cells accelerates osteoblastic differentiation. J Bone Min Res 2002;17:1008–14.

[52] Wingender E, Schoeps T, Donitz J. TFClass: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res 2013;41:D165–70.

[53] Kawane T, Komori H, Liu W, Moriishi T, Miyazaki T, Mori M, et al. Dlx5 and mef2 regulate a novel runx2 enhancer for osteoblast-specific expression. J Bone Min Res 2014;29:1960–9.

[54] Bedalov A, Salvatori R, Dodig M, Kronenberg MS, Kapural B, Bogdanovic Z, et al. Regulation of COL1A1 expression in type I collagen producing tissues: identification of a 49 base pair region which is required for transgene expression in bone of transgenic mice. J Bone Min Res 1995;10:1443–51.

[55] Dodig M, Kronenberg MS, Bedalov A, Kream BE, Gronowicz G, Clark SH, et al. Identification of a TAAT-containing motif required for high level expression of the COL1A1 promoter in differentiated osteoblasts of transgenic mice. J Biol Chem 1996;271:16422–9.

[56] Mao B, Wu W, Davidson G, Marhold J, Li M, Mechler BM, et al. Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling. Nature 2002;417:664–7.

[57] Samsa WE, Zhou X, Zhou G. Signaling pathways regulating cartilage growth plate formation and activity. Semin Cell Dev Biol 2016;62:3–15.

[58] Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell 2013;49:825–37.

[59] Plank JL, Dean A. Enhancer function: mechanistic and genome-wide insights come together. Mol Cell 2014;55:5–14.

[60] Romanoski CE, Glass CK, Stunnenberg HG, Wilson L, Almouzni G. Epigenomics: roadmap for regulation. Nature 2015;518:314−6.

[61] Yao B, Wang Q, Liu CF, Bhattaram P, Li W, Mead TJ, et al. The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. Nucleic Acids Res 2015;43: 5394−408.

[62] Gordon CT, Attanasio C, Bhatia S, Benko S, Ansari M, Tan TY, et al. Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. Hum Mutat 2014;35: 1011−20.

[63] Meyer MB, Benkusky NA, Sen B, Rubin J, Pike JW. Epigenetic plasticity drives adipogenic and osteogenic differentiation of marrow-derived mesenchymal stem cells. J Biol Chem 2016;291:17829−47.

[64] Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 2014;159:1312−26.

[65] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486−90.

[66] Li J, Luo H, Wang R, Lang J, Zhu S, Zhang Z, et al. Systematic reconstruction of molecular cascades regulating GP development using single-cell RNA-seq. Cell Rep 2016;15:1467−80.

[67] Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, et al. Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature 2014;515:365−70.

[68] Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell 2015;163:68−83.

[69] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485:376−80.

[70] Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 2015;161:1012−25.

[71] Taylor SE, Li YH, Wong WH, Bhutani N. Genome-wide mapping of DNA hydroxymethylation in osteoarthritic chondrocytes. Arthritis Rheumatol 2015;67:2129−40.

[72] Guo Y, Dong SS, Chen XF, Jing YA, Yang M, Yan H, et al. Integrating epigenomic elements and GWASs identifies BDNF gene affecting bone mineral density and osteoporotic fracture risk. Sci Rep 2016;6:30558.

[73] Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173−86.

[74] Kemp JP, Medina-Gomez C, Estrada K, St Pourcain B, Heppe DH, Warrington NM, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. PLoS Genet 2014;10: e1004423.

[75] Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337:1190−5.

[76] Kang J, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, et al. Modulation of tissue repair by regeneration enhancer elements. Nature 2016;532:201−6.

[77] Guenther CA, Wang Z, Li E, Tran MC, Logan CY, Nusse R, et al. A distinct regulatory region of the Bmp5 locus activates gene expression following adult bone fracture or soft tissue injury. Bone 2015;77:31−41.

[78] Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 2015;33:510−7.

[79] Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods 2015;12:401−3.