



HAL
open science

Transcriptional regulation of jaw osteoblasts: development to pathology

Ali Nassif, Guilhem Lignon, Audrey Asselin, Charles Zadikian, Stephane Petit, Hongwei Sun, C. Klein, François Côme Ferré, Maria Mi Morasso, Ariane Berdal, et al.

► **To cite this version:**

Ali Nassif, Guilhem Lignon, Audrey Asselin, Charles Zadikian, Stephane Petit, et al.. Transcriptional regulation of jaw osteoblasts: development to pathology. 2022. hal-03557330

HAL Id: hal-03557330

<https://u-paris.hal.science/hal-03557330>

Preprint submitted on 4 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Dental Research

Transcriptional regulation of jaw osteoblasts: development to pathology

Journal:	<i>Journal of Dental Research</i>
Manuscript ID	JDR-21-0852.R1
Manuscript Type:	Research Reports
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Nassif, Ali; Université de Paris, UFR Odontology, Department of Oral Biology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Lignon, Guilhem; Université de Paris, UFR Odontology, Department of Biomaterials; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Asselin, Audrey; Université de Paris, UFR Odontology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Zadikian, Charles; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Petit , Stephane ; Université de Paris, UFR Odontology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Sun, Hong-Wei; NIAMS, Biodata Mining and Discovery Section , Office of Science and Technology</p> <p>Klein, Christophe; INSERM, Histology, cell Imaging and flow Cytometry Platform (CHIC)</p> <p>Ferré, Francois; Université de Paris, UFR Odontology, Department of Oral Biology; INSERM, Molecular oral Pathophysiology</p> <p>Morasso, Maria; NIAMS, Laboratory of Skin Biology</p> <p>Berdal, Ariane; Université de Paris, Department of Oral Biology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Fournier, Benjamin; Université de Paris, Oral biology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p> <p>Isaac, Juliane; Université de Paris, UFR Odontology, Department of Oral Biology; Centre de Recherche des Cordeliers, Molecular oral Pathophysiology</p>
Keywords:	Osteoblast(s), Transcription factor(s), Bioinformatics, Bone biology, Craniofacial anomalies, Gene expression
Abstract:	<p>Craniofacial and jaw bones have unique physiological specificities when compared to axial and appendicular bones. However, the molecular profile of the jaw osteoblast (OB) remains incomplete. The present study aimed to decipher the bone site-specific profiles of transcription factors (TFs) expressed in OBs in vivo. Using RNA-seq analysis, we mapped the transcriptome of confirmed OBs from two different skeletal sites: mandible (Md) and tibia (Tb). The OB transcriptome contains 709 TF genes: 608 are similarly expressed in Md-OB and Tb-OB, referred to as</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	<p>“OB-core”; 54 TF genes are upregulated in Md-OB, referred to as “Md-set”; and 18 TF genes are upregulated in Tb-OB, referred to as “Tb-set”. Notably, the expression of 29 additional TF genes depends on their RNA transcript variants. TF genes with no previously known role in OBs and bone were identified. Bioinformatics analysis combined with review of genetic disease databases and a comprehensive literature search showed a significant contribution of anatomical origin to the OB signatures. Md-set and Tb-set are enriched with site-specific TF genes associated with development and morphogenesis (neural crest versus mesoderm), and this developmental imprint persists during growth and homeostasis. Jaw and tibia site-specific OB signatures are associated with craniofacial and appendicular skeletal disorders as well as neurocristopathies, dental disorders and digit malformations. The present study demonstrates the feasibility of a new method to isolate pure OB populations and map their gene expression signature in the context of OB physiological environment, avoiding in vitro culture and its associated biases. Our results provide insights into the site-specific developmental pathways governing OBs and identify new major OB regulators of bone physiology. We also established the importance of the OB transcriptome as a prognostic tool for human rare bone diseases to explore the hidden pathophysiology of craniofacial malformations, among the most prevalent congenital defects in humans.</p>



Transcriptional regulation of jaw osteoblasts: development to pathology

A. Nassif^{1,2,3,4†}, G. Lignon^{1,2†}, A. Asselin^{1,2†}, C.C. Zadikian², S. Petit^{1,2}, H.W. Sun⁵, C. Klein⁶,
F.C. Ferré^{2,7}, M.I. Morasso⁸, A. Berdal^{1,2,3†}, B.P.J. Fournier^{1,2,3†}, J. Isaac^{1,2*}

- 1- Université de Paris, Dental Faculty Garancière, Department of Oral Biology, Paris, France
- 2- Centre de Recherche des Cordeliers, Université de Paris, Sorbonne Université, Inserm, Lab. of Molecular Oral Pathophysiology, Paris, France
- 3- AP-HP, Reference Center for Dental Rare Diseases, Rothschild Hospital (ORARES), Paris, France.
- 4- AP-HP, Pitié Salpêtrière, Service d'Orthopédie Dento-faciale, Paris, France.
- 5- Biodata Mining and Discovery Section, Office of Science and Technology, Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA.
- 6- Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, Histology, cell Imaging and flow Cytometry Platform (CHIC), Paris, France
- 7- AP-HP, Charles Foix-Pitié Salpêtrière Hospital, Dental Department, Ivry, France.
- 8- Laboratory of Skin Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA.

† These authors contributed equally to this work

*Author to whom correspondence should be addressed. juliane.isaac@u-paris.fr

Abstract

Craniofacial and jaw bones have unique physiological specificities when compared to axial and appendicular bones. However, the molecular profile of the jaw osteoblast (OB) remains incomplete. The present study aimed to decipher the bone site-specific profiles of transcription factors (TFs) expressed in OBs *in vivo*. Using RNA-seq analysis, we mapped the transcriptome of confirmed OBs from two different skeletal sites: mandible (Md) and tibia (Tb). The OB transcriptome contains 709 TF genes: 608 are similarly expressed in Md-OB and Tb-OB, referred to as “OB-core”; 54 TF genes are upregulated in Md-OB, referred to as “Md-set”; and 18 TF genes are upregulated in Tb-OB, referred to as “Tb-set”. Notably, the expression of 29 additional TF genes depends on their RNA transcript variants. TF genes with no previously known role in OBs and bone were identified. Bioinformatics analysis combined with review of genetic disease databases and a comprehensive literature search showed a significant contribution of anatomical origin to the OB signatures. Md-set and Tb-set are enriched with site-specific TF genes associated with development and morphogenesis (neural crest *versus* mesoderm), and this developmental imprint persists during growth and homeostasis. Jaw and tibia site-specific OB signatures are associated with craniofacial and appendicular skeletal disorders as well as neurocristopathies, dental disorders and digit malformations. The present study demonstrates the feasibility of a new method to isolate pure OB populations and map their gene expression signature in the context of OB physiological environment, avoiding *in vitro* culture and its associated biases. Our results provide insights into the site-specific developmental pathways governing OBs and identify new major OB regulators of bone physiology. We also established the importance of the OB transcriptome as a prognostic tool for human rare bone diseases to explore the hidden pathophysiology of craniofacial malformations, among the most prevalent congenital defects in humans.

Number of figures, tables: 4 figures, 1 table

Number of references: 40

Key Words: Osteoblast, Transcriptome, RNA-seq, Jawbone, Tibia, Transcription Factor, Neural Crest, Bone disease, Development

Introduction

Driven by a complex network of transcription factors (TFs), the osteoblast (OB) is traditionally considered to be a unique differentiated mesenchymal cell that synthesizes various peptides in charge of matrix deposition, biomineralization and remodeling during bone growth, homeostasis and healing (*reviewed in* (Lian et al. 2006)). However, craniofacial and jawbones exhibit a number of physiological specificities when compared to axial and appendicular bones (*reviewed in* (Wang et al. 2020)). Craniofacial bones and long bones primarily develop through different ossification processes: intramembranous ossification for most craniofacial bones and endochondral ossification for long bones. Long bone growth and homeostasis, which ensure skeletal mobility and load bearing, are mainly regulated by mechanical strains and hormones. Jawbone development additionally depends on teeth, as evidenced by its reduced volume in inherited tooth agenesis or after tooth extraction (Bertl et al. 2018; Nowwarote et al. 2019). Jawbones and long bones also respond differently to various agents. Estrogens and parathyroid hormone imbalances affect mineral density in long bones disproportionately (Liu et al. 2009; Coutel et al. 2019; Watanabe et al. 2020). Bisphosphonates and other antiresorptive drugs have few adverse effects on long bones but alter jawbone metabolism, sometimes culminating in jaw osteonecrosis (*reviewed in* (Katsarelis et al. 2015)). Finally, autografts taken from craniofacial bones offer better outcomes regardless of the acceptor site when compared with grafts from long bones (Akintoye et al. 2006; Leucht et al. 2008).

Unsurprisingly then, jawbones and long bones show major differences at the cellular level. Craniofacial bone progenitors show differential proliferation and osteogenic potential when compared to long bone progenitors (Matsubara et al. 2005; Leucht et al. 2008; Aghaloo et al. 2010; Kelder et al. 2020). Their cell lineages also differ: while the vast majority of craniofacial OBs derive from cranial neural crest, long bone OBs are of mesodermal origin (Le Douarin 1980; Gans and Northcutt 1983; Chai et al. 2000; Yoshida et al. 2008). Furthermore, *in vivo* and *in vitro* analyses suggest the existence of cell site-specific signatures between jawbones and long bones (Kasperk et al. 1995), notably for a number of TFs (Rawlinson et al. 2009; Kingsmill et al. 2013; Reichert et al. 2013; Lee et al. 2015). However, to date, the transcriptome of the jaw OB in its physiological environment has remained elusive, as the studies focusing on jaws were performed using either cell culture or whole-tissue extracts. Moreover, jawbone cells, including OBs but also osteocytes and bone marrow cells, have been omitted from large-scale omics studies on the tissue-specific landscape (Zhou et al. 2017; Tabula Muris Consortium et al. 2018; Youtlen et al. 2021). Here, we prove the feasibility of a new isolation method to obtain

1
2
3 pure populations of functional OBs within their physiological environment. This strategy
4 allowed us to define the complete repertoires of TF genes expressed in jaw- and tibia-OBs *in*
5 *vivo*. These data provide insights into the developmental and genetic pathways governing site-
6 specific OB differentiation and activity, and identify major OB regulators involved in jawbone
7 physiology and disorders.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Materials and Methods

Animal models

Col1a1*2,3-GFP mice (JAX013134) and Pax3Cre KI mice (JAX-005549) were used to identify functional OBs and neural crest (NC) derivatives, respectively. These mouse models were crossed with Rosa^{tomato} mice (JAX-008851) to obtain Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice. All experiments were presented to and approved by the Charles Darwin ethics committee (CEEACD/05) before the study began and conducted in accordance with the European legislation for the care and use of laboratory animals. This study adheres to the ARRIVE Guidelines for reporting animal. For ARRIVE guidelines checklist, mouse breeding, genotyping, histology, immunofluorescence (IF) analyses, bone microdissection, cell isolation, RNA extraction, RT-qPCR, bioinformatics resources and statistical analyses, see the **Appendix**.

RNA Sequencing (RNA-Seq)

Mandible and tibia bones from P9 Col1a1*2,3-GFP mice were dissected (**Appendix**). Three independent dissections, each pooling bone tissues from approximately fifty P9 mice, were carried out. Immediately after dissection and collagenase I digestion, GFP-positive cells were FACS-sorted (**Appendix**). Protocols of RNA extraction, cDNA library construction and data processing pipeline are detailed in **Appendix** and are publicly available in the Gene Expression Omnibus (GEO), together with the RNA-Seq raw and processed data files (accession# GSE186535). Criteria for significant gene selection included q -values ≤ 0.1 (for multi-test adjustment), $FC \geq 1.4$ or ≤ -1.4 , and mean RPKM ≥ 1 for expressed genes.

Results

OB expression profiles include TFs conserved among bone sites as well as site-specific TFs

The *in vivo* TF signatures expressed in functional OBs were defined by using a global gene profiling approach based on RNA-seq of flow-sorted GFP-positive cells from mandibular (Md) and tibia (Tb) bones of 9-day-old (P9) *Coll1a1*2,3-GFP* mice (**Fig. 1A**). Our RNA-seq results were cross-referenced with a list of 1388 TFs (**GEO accession# GSE186535**). Md-OB or/and Tb-OB expressed 709 TF genes (977 transcripts), i.e. more than 50% of the known TF genes (**Appendix Table 2**). The largest group, termed “OB-core”, comprises 608 TF genes (86%) similarly expressed in Md-OBs and Tb-OBs. The remaining TF genes were differentially expressed between the two skeletal sites: “Md-set” consists of 54 TF genes (8%) upregulated in Md-OBs, and “Tb-set” consists of 18 TF genes (2%) upregulated in Tb-OBs (**Fig. 1B**). Notably, for 29 TF genes (4%), the bone site-specificity depended on their RNA transcript variants (TVs) (**Appendix Fig. 1**). To decipher how the anatomical origin contributed to the OB signatures, the three OB groups were further investigated with an enrichment analysis based on the Biological Processes (BPs) of the functional GO annotation using DAVID and a literature review (**Fig. 1C-E and Appendix Fig. 2 and 3**).

OB and bone BPs: Analysis of the OB transcriptome showed that OB-core and Md-set were enriched with TF genes associated with OB and bone-related BPs (**Fig. 1C, Table 1**). In OB-core, 21% of the TF genes (135/637) were previously established as regulators of OB (*e.g.* *Atf3*, *Klf6*, *Runx2*, *Sp7*, *Tsc22d*) (**Appendix Fig. 2**). The highest expressed TF genes of OB-core (with mean (RPKM) > 100) were all previously known for their involvement in OB (*Atf3/4*, *Tsc22d1/3*, *Egr1*, *FosB*, *Hnrnpk*, *Klf6* and *Nr4a*). However, the percentage of TF genes with known roles in OB decreased along with the TF expression level, and between 75% and 80% of the TFs with mean (RPKM) < 25 have no previously known role in OB and bone (**Appendix Fig. 2**). To further investigate the expression of some of the TF genes of OB-core, qPCR and IF analyses were performed. In line with our RNAseq results, no significant difference in mRNA levels between Md- and Tb-bones isolated from P9 C57BL/6JR mice was detected for the eight investigated TF genes (*Arid5b*, *Atf3*, *Klf6*, *Meox2*, *Runx2*, *Sp7*, *Tsc22d1*, *Ybx3*) (**Appendix Fig. 2 and 4**). Moreover, *Runx2* and *Sp7* IF analyses in E15.5 and P9 *Coll1a1*2,3-GFP* mice showed similar protein expression pattern with almost all the functional OBs expressing both *Sp7* and *Runx2* independently of their localization (**Appendix Fig. 4**). In Md-

1
2
3 set, 62% of the TF genes (46/74) were previously known to regulate OB and bone (**Fig. 1D**).
4 Among them, 22 TF genes were expressed specifically in Md-OBs [including Alx genes
5 (Alx1/3/4), Dlx genes (Dlx1/2), Pax genes (Pax3/9) and Hand2, Msx2, Six2, Sox5] and 24 TF
6 genes showed lower expression in Tb-OBs [including Dlx genes (Dlx3/4/5), Sox genes
7 (Sox9/11) and Msx1, Osr2, Pbx1, Sp6] (**Fig. 1D**).
8
9

10
11 **Other connective/mesenchymal tissue BPs:** OB-core was significantly enriched with TF genes
12 associated with cartilage, skeletal muscle and fat BPs, whereas enrichment with TF genes
13 associated with dental BPs was low (**Fig. 1C and Appendix Fig. 2**). Although Md-set displayed
14 an enrichment profile similar to that of OB-core for bone and cartilage-related BPs, dental BPs
15 were overrepresented in Md-set, whereas skeletal and fat-related BPs were low and mostly
16 associated with negative regulation (**Fig. 1C and Appendix Fig. 3**). Gene-by-gene analysis of
17 Md-set confirmed the high contribution of dental-associated genes, as 26% of the TF genes
18 (19/74) have known roles in odontogenesis and odontoblast differentiation (*e.g.* Dlx4, Alx3,
19 Lhx8, Dlx1) (**Fig. 1D**).
20

21
22 In Tb-set, a gene-by-gene literature analysis revealed bone-related TFs known for their
23 regulatory roles in cartilage (*e.g.* Tbx18, Hoxa5/d8, Nfia), skeletal muscle (*e.g.* Heyl, Hoxa7,
24 Crem, Nfya) and fat (*e.g.* Creb3l3, Hoxa9/c5, Tbx15). Shox2 was the only TF known to have a
25 role in odontogenesis but also in bone, cartilage, muscle and fat (**Appendix Fig. 3**). Notably,
26 TFs known for their involvement in the regulation of mesenchymal cells and tissues other than
27 OBs and bone and never described in differentiated OBs were identified in the RNA-seq panel.
28 These novel OB TFs include Arid5a/5b, Ybx3, Meox2 and Creb5 in OB-core; Zfp90/110, Sp6,
29 Rxrb and Bhlhb9 in Md-set; and Sox7 and Irf7 in Tb-set (**Fig. 1E and Appendix Fig. 2 and**
30 **3**). mRNA expression of some of these TF genes was validated by qPCR in raw bones from P9
31 C57BL/6JR mice (**Appendix Fig. 2 and 3**). To further investigate one of the novel OB TFs
32 identified in Md-set, the expression of Sp6, a known odontogenesis regulator, was analyzed by
33 qPCR and IF from development (E10.5) to post-natal stages (P9). In line with our RNA-seq
34 results, Sp6 was detected at mRNA and protein levels in craniofacial tissues, including in Md-
35 bone and Md-OBs from E14.5 to P9 in C57BL/6JR mice, while no expression was detected in
36 appendicular skeleton (**Fig. 1F-I and Appendix Fig. 5**).
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

54 **2- Site-specific OB-omics: blueprints of embryonic signatures**

55 **Development and morphogenesis BPs:** To address the contribution of embryonic BPs to OB
56 signatures at post-natal stage P9, we first performed enrichment analysis of the 3 OB sets (**Fig.**
57 **2A**). The 3 sets were enriched in TF genes associated with the following BPs: skeletal system,
58
59
60

limb and appendage development, and morphogenesis. Only OB-core and Md-set displayed enrichment in TF genes associated with head and craniofacial developmental BPs, while “limb bud formation” and “embryonic digit morphogenesis” were enriched exclusively in Md-set. OB-core was enriched with TF genes associated with mesoderm development BPs, and Md-set with NC BPs, whereas Tb-set displayed no enrichment in any of these BPs. RNA-seq data were also analyzed in light of 278 TFs associated with NC, based on information generated from the literature (**Fig. 2B-C, Appendix Table 3**). 39% of Md-set TF genes (29/74) were associated with NC, and all these genes were directly associated either with cranial NC (CNC) (*e.g.* Alx1/3/4, Dlx1/2/3/5, Msx1/2, Sox5/9/11) or with organogenesis/patterning of craniofacial skeletal elements (Pax9) or NC differentiation (Nfatc4). Six Tb-set TF genes were associated with caudal NC and participated in positional identity in the anterior–posterior axis (Hoxa5/a7/c10/c5, Tbx18) and the epithelial–mesenchymal transition (Akna). **The expression of two TFs associated with CNC, Dlx3 and Sox9, was further investigated from development (E10.5) to post-natal stages (P9). Both Dlx3 and Sox9 showed high mRNA expression levels in embryonic and peri-natal skeletal tissues as previously described (Appendix Table 4) and, in line with our RNA-seq data, were significantly higher from E10.5 to P9 in head and craniofacial tissues, including Md-bone, when compared to appendicular skeleton (Fig. 2D-E, Appendix Fig. 5). *In situ* IF imaging of Sox9 confirmed the known Sox9 localization in cartilaginous tissues and further demonstrated its expression in Md-bone and Md-OBs at protein level (Fig. 2F-H).**

Long-term maintenance of in vivo and ex vivo OB embryonic positional identity: To determine whether Md-OB origin remains neuroectodermal with age, we examined the distribution of NC-derived cells in the growing oral bones at embryonic (E15.5), postnatal (P9) and homeostasis stages [2 months (2M), 4 months (4M) and 8 months (8M)], taking advantage of the individual cell visibility of fluorescent markers in Pax3-cre::Colla1*2,3-GFP::Rosa^{tomato} mice (**Fig. 3 and Appendix Fig. 5**). Microscopic observations (**Fig. 3A-C**) and cell quantification using co-localization of tomato, GFP and DAPI signals on FACS-isolated cells (**Fig. 3D**) and tissue sections (**Fig. 3E**) revealed that OBs from both jaws (upper and lower) were of NC origin from development to adulthood. Dlx3, Msx1, Msx2 and Sox9 mRNA levels were significantly higher in mandible bone when compared to tibia bone at P9, 2M and 8M, whereas Hoxa10 and Hoxc10 mRNA was only detected in tibia bone (**Fig. 3F**). To investigate this embryonic signature *ex vivo*, we cultured primary osteoprogenitors from mandible (Md-cells) and tibia (Tb-cells) of P9 Pax3-cre::Colla1*2,3-GFP::Rosa^{tomato} mice (**Fig. 3G-H**). While the Md-cells that differentiated *in vitro* were double positive (GFP+, Tom+), Tb-cells displayed

only GFP signal (Fig. 3G). RT-qPCR results showed differential expression of *Dlx3*, *Msx1*, *Msx2* and *Sox9* in Md-cells, and of *Hoxa10* and *Hoxc10* in Tb-cells persisted *in vitro* (Fig. 3H).

3—Site-specific OB-omics: a prognostic tool for rare diseases

Using the OMIM database and a comprehensive literature review, we identified 29 TFs associated with human genetic disorders (out of 101 differentially expressed genes) in our omics data (Fig. 4A, Appendix Table 5). Twenty-six of these TF genes are directly associated with mineralized tissue pathologies, whereas 3 are linked to deficiencies in hematopoiesis and/or immune function (*KLF1*, *IRF1* and *IRF7*) (Fig. 4B-C).

Twenty neurocristopathies are known to be caused by 9 TFs in Md-set (*ALX1/3/4*, *DLX3*, *GLI3*, *GSC*, *MSX2*, *PAX3*, *SOX9*). Except for *IRF1*, all the human disorders associated with Md-set genes display orofacial features. Thirteen TF genes are associated with jawbone phenotypes (11 in mandible, 5 in maxilla) as well as cranial bone pathologies (13/13), orofacial clefts (9/13), and dental abnormalities (6/13) (Fig. 4B). In humans, mutations of some Md-set TF genes are not associated with clinical jawbone features but instead with defects in cranial bones, palate and/or teeth (*ALX3/4*, *DLX4*, *MSX1/2*, *PAX9*, *PBX1*, *SALL4*, *SOX11*, *SP6*, *TBX3*) (Fig. 4C). Mutations of the Md-set genes *DLX3*, *DLX5*, *GLI3*, *GSC*, *MSX2*, *SALL4*, *SOX9* and *TBX3* cause not only craniofacial defects in humans, but also disorders in long bones and/or shoulder girdle/gluteal regions (Fig. 4C, Appendix Table 5).

The dysmorphological patterns of these human rare diseases are mirrored by the differential mRNA levels of TFs measured in mouse jaw and tibia OBs and bones by RNA-seq and qPCR (Fig. 1F, Fig. 2D-E, Fig. 3F, Fig. 4D) and the site-specific *in situ* localization of TFs identified by IF (Fig. 1G-I and Fig. 2F-H) and previously reported (Appendix Table 4). For example, mutations of TF genes highly expressed in Tb-OBs (*Dlx3/5*, *Sox9*) (Fig. 2D-E and Appendix Fig. 5) are associated with long-bone disorders in humans. Mutations of genes weakly expressed in both Md-OBs and Tb-OBs or with relatively low fold changes between the two OB populations affect both craniofacial bones and long bones (*e.g.* *GSC*, *SALL4*, *TBX3*) and/or are included in clinical synopses of short stature (*GLI3*, *GSC*, *PBX1*, *SOX11*, *MYCN*) (Appendix Table 5). This is also the case for *Nfia* and *TBX15*, which are more significantly expressed in Tb-set but at low levels and whose mutations in humans affect both long bones and craniofacial bones. Among the genes we found by RNA-Seq and which cause both oral and appendicular bone disorders in humans, are *Dlx5*, *Pbx1* and *Nfia*; these genes are characterized by both bone site-specific TVs as well as TVs commonly expressed in OB-core (Appendix Fig. 1). Finally, in line with our bioinformatics analyses showing enrichment in Md-set with

1
2
3 TF genes associated with limb bud formation and digit morphogenesis (**Fig. 2A**), mutations of
4 14 Md-set TF genes have been implicated in human digit malformations. Together, these data
5 suggest that the OB transcriptome may be used as a diagnostic tool to predict the affected bone
6 sites as well as other tissue disorders such as dental anomalies, neurocristopathies or digit
7 malformations associated with TF mutations.
8
9
10
11
12
13

14 **Discussion**

15
16
17
18 This study defined the complete repertoires of TF genes expressed in jaw and tibia OBs in their
19 native physiological environments. To do so, we established a new and easily reproducible
20 approach by isolating functional OBs directly from tissues and thus avoiding cell culture, which
21 may introduce biases in gene expression (Ryu et al. 2017). Notably, our results evidenced
22 differentially expressed transcript variants from a given gene between bone sites, including
23 coding and non-coding RNA sequences, of key OB regulators such as *Dlx5* and *Runx2*. These
24 results are in line with osteocyte transcriptome mapping and strongly suggest additional levels
25 of regulation of proteins synthesized by bone cells (Youtlen et al. 2021).
26
27
28
29
30
31

32 In these OB populations free of contaminating cell-types and -stages, we identified a large TF
33 set conserved between mandible and tibia OBs. This supports a model in which a central
34 transcriptional network, OB-core, operates in all OBs of the skeleton independently of their
35 localization. Other TF genes were site-specific, highlighting the importance of the experimental
36 models chosen when studying bone disorders, notably those of the jaws. In line with previous
37 studies (Aïoub et al. 2007; Duverger et al. 2013; Reichert et al. 2013; Nassif et al. 2014; Lee et
38 al. 2015), the preferential expression of *Msx/Dlx* genes in mandible and *Hox* genes in tibia is
39 supported here, and these genes are identified as being specifically related to functional OBs *in*
40 *vivo*. *Msx/Dlx/Hox* homeoproteins operate under different contexts: (1) in early stages they
41 drive rostro-caudal patterning and morphogenesis; and (2) thereafter, they trigger key OB
42 functions, including the secretion of a number of non-collagenous peptides (Molla et al. 2010;
43 Duverger et al. 2012). Overall, the maintained spatial expression of several TFs supports the
44 existence of developmental imprinting in differentiated OBs at postnatal and adult stages.
45 Furthermore, *in vivo* NC cell tracking demonstrated the exclusivity of CNC and the absence of
46 mesoderm participation in jaw OBs. In tibia, CNC participation was excluded. Thus, OBs
47 maintain both a spatial and embryologic identity from growth to advanced post-natal stages.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 This implies that OBs differentiate from resident cell niches during growth and homeostasis, as
4 fibroblasts do (Rinkevich et al. 2015; Mah et al. 2017).

5
6 In OB-core, although the cardinal bone TFs were found (*e.g.* Runx2, Sp7, Twist1), the vast
7 majority of TF genes (almost 80%) had not been previously described as being expressed in
8 differentiated OBs or associated with OB or bone. The same observation holds true for the Md
9 and Tb sets. The present exhaustive inter-site comparative approach therefore completes the
10 previously incomplete OB transcriptional profile and provides a better understanding of OBs
11 within their physiological environments. The newly identified genes in the OB-core and site-
12 specific TF sets will allow reevaluation of the bone symptoms of the associated genetic
13 diseases.
14

15
16 We confirmed the relevance of the jaw and tibia OB signatures by comparing the detected TFs
17 to genes implicated in monogenic skeletal diseases. Genes and rare diseases affecting other
18 organs were also analyzed. Such was the case for the teeth, which share an embryological origin
19 with orofacial bones. Jaw OBs display a unique mesenchymal profile highly enriched with
20 dental-related TF genes (*e.g.* Pax9, Msx2, Dlx2, Sp6), consistent with human tooth and bone
21 genes/phenotypes (de La Dure-Molla et al. 2019). Additionally, our bioinformatics analyses
22 showed enrichment of biological processes involved with limb bud formation and digit
23 morphogenesis in Md-set. Analysis of human mutations confirmed that 14 of these genes are
24 associated with digit malformations. Thus, in addition to a number of genes consistently
25 associated with neurocristopathies and craniofacial development, the present study also
26 suggests molecular similarities between craniofacial bones and digits.
27

28
29 Notably, mutations of some TF genes of Md-set have not yet been associated with jawbone
30 disorders in humans. These synopses appear to contradict not only the RNA-Seq, qPCR data
31 and IF analysis for OB and raw jaw/tibia bones reported here, but also the described mouse
32 jawbone phenotypes for various mutations, such as Alx3 and Alx4 (Beverdam et al. 2001), Dlx4
33 (Wu et al. 2015), Msx1 (Nassif et al. 2014) and Msx2 (Aïoub et al. 2007). These observations
34 support that jawbone abnormalities might also be present but not diagnosed in human disorders
35 and highlight the need for systematic exploration of the jawbones while investigating skeletal
36 and dental diseases.
37

38
39 Apart from rare diseases, this molecular signature may underlie known site-specific bone
40 behavior in pathophysiology. For instance, the exclusive CNC origin and Hox-negative status
41 of jaw OBs are shared with other oral mesenchymal cells such as gingival fibroblasts (Isaac et
42 al. 2018). This singular profile has been shown to provide these cells with increased stemness
43 and regenerative potential (Wang et al. 2009; Jiang et al. 2018). These qualities are crucial for
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 tissue regeneration and stem cell therapies, as consistently reflected in the superior efficiency
4 of jaw- versus long-bone grafts (Akintoye et al. 2006; Leucht et al. 2008).
5
6
7

8 In conclusion, the OB TF transcriptome provides insights into the pathways governing site-
9 specific OB development and homeostasis, and identifies novel OB regulators involved in bone
10 physiology and human rare diseases. Integrative analysis of the OB reveals that its TF
11 signatures appear to be developmental imprints of distinct embryonic origin (NC/mesoderm)
12 that may directly influence the behavior of bones and dictate their anatomical site specificities.
13
14
15
16
17 Finally, the present study also establishes the importance of the OB transcriptome as a
18 prognostic tool to predict affected bone sites as well as other tissue disorders such as dental
19 anomalies, neurocristopathies or digit malformations associated with TF mutations. Indeed,
20 deciphering these Md-OB and Tb-OB signatures will facilitate our understanding of the
21 pathophysiology of phenotypes associated to polymorphisms and/or identified by genome-wide
22 association studies (GWAS).
23
24
25
26
27
28
29
30

31 **Acknowledgements**

32 This work was supported by the Université de Paris and the French National Research Agency
33 (ANR) [IdEx Université de Paris (grant ANR-18-IDEX-0001) and the Osteodiversity-ANR
34 consortium (ANR-12-BSV1-0018)], grants from “Fondation des Gueules Cassées” and the
35 Federation Hospitalo-Universitaire FHU DDS-ParisNet, a grant from Assistance Publique-
36 Hôpitaux de Paris (AP-HP), Inserm and Université de Paris. We thank Dr. Benoit Robert and
37 Dr. Yvan Lallemand of the Institut Pasteur, the INSERM UMRS 1138 Histology, Cell Imaging
38 and Flow Cytometry Platform (CICH) and the animal core facilities of INSERM UMRS 1138
39 for their expert technical support.
40
41
42
43
44
45
46
47

48 **Author contributions**

49 A.N: Contributed to acquisition, and interpretation, drafted the manuscript critically and
50 revised the manuscript
51 G. L: Contributed to acquisition, and interpretation, drafted the manuscript critically and
52 revised the manuscript
53 A. A: Contributed to conception and design, to acquisition and interpretation, and critically
54 revised the manuscript
55 S. P: Contributed to acquisition and critically revised the manuscript
56 C.C. Z: Contributed to acquisition and critically revised the manuscript
57 H.W.S: Contributed to analysis and critically revised the manuscript
58 C. K: Contributed to acquisition and critically revised the manuscript
59
60

1
2
3 F.C.F: Contributed to analysis and critically revised the manuscript

4 M.I.M: Contributed to analysis and critically revised the manuscript

5 A. B: Contributed to analysis and interpretation and critically revised the manuscript

6 B.J.P.F: Contributed to analysis and interpretation and critically revised the manuscript

7 J. I: Contributed to conception and design, acquisition, analysis, and interpretation, performed
8 all statistical analyses, drafted the manuscript and critically revised the manuscript
9

10
11 All authors gave their final approval and agree to be accountable for all aspects of the work.
12
13

14
15 **Conflict of interest statement**
16

17 The authors declare no potential conflicts of interest with respect to the authorship and/or
18 publication of this article.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Aghaloo TL, Chaichanasakul T, Bezouglaia O, Kang B, Franco R, Dry SM, Atti E, Tetradis S. 2010. Osteogenic Potential of Mandibular vs. Long-bone Marrow Stromal Cells. *J Dent Res.* 89(11):1293–1298. doi:10.1177/0022034510378427.
- Aïoub M, Lézet F, Molla M, Castaneda B, Robert B, Goubin G, Néfussi JR, Berdal A. 2007. *Msx2* $-/-$ transgenic mice develop compound amelogenesis imperfecta, dentinogenesis imperfecta and periodontal osteopetrosis. *Bone.* 41(5):851–859. doi:10.1016/j.bone.2007.07.023.
- Akintoye SO, Lam T, Shi S, Brahim J, Collins MT, Robey PG. 2006. Skeletal site-specific characterization of orofacial and iliac crest human bone marrow stromal cells in same individuals. *Bone.* 38(6):758–768. doi:S8756-3282(05)00442-4 [pii] 10.1016/j.bone.2005.10.027.
- Bertl K, Bertl MH, Heimel P, Burt M, Gahleitner A, Stavropoulos A, Ulm C. 2018. Alveolar bone resorption after primary tooth loss has a negative impact on straightforward implant installation in patients with agenesis of the lower second premolar. *Clin Oral Implants Res.* 29(2):155–163. doi:10.1111/clr.13033.
- Beverdam A, Brouwer A, Reijnen M, Korving J, Meijlink F. 2001. Severe nasal clefting and abnormal embryonic apoptosis in *Alx3/Alx4* double mutant mice. *Development.* 128(20):3975–3986.
- Chai Y, Jiang X, Ito Y, Bringas P, Han J, Rowitch DH, Soriano P, McMahon AP, Sucov HM. 2000. Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development.* 127(8):1671–1679.
- Coutel X, Delattre J, Marchandise P, Falgayrac G, Béhal H, Kerckhofs G, Penel G, Olejnik C. 2019. Mandibular bone is protected against microarchitectural alterations and bone marrow adipose conversion in ovariectomized rats. *Bone.* 127:343–352. doi:10.1016/j.bone.2019.06.031.
- Duverger O, Isaac J, Zah A, Hwang J, Berdal A, Lian JB, Morasso MI. 2013. In vivo impact of *Dlx3* conditional inactivation in neural crest-derived craniofacial bones. *J Cell Physiol.* 228(3):654–664. doi:10.1002/jcp.24175.
- Duverger O, Zah A, Isaac J, Sun HW, Bartels AK, Lian JB, Berdal A, Hwang J, Morasso MI. 2012. Neural crest deletion of *Dlx3* leads to major dentin defects through down-regulation of *Dspp*. *J Biol Chem.* 287(15):12230–12240. doi:10.1074/jbc.M111.326900.
- Gans C, Northcutt RG. 1983. Neural crest and the origin of vertebrates: a new head. *Science.* 220(4594):268–273. doi:10.1126/science.220.4594.268.
- Isaac J, Erthal J, Gordon J, Duverger O, Sun HW, Lichtler AC, Stein GS, Lian JB, Morasso MI. 2014. *DLX3* regulates bone mass by targeting genes supporting osteoblast differentiation and mineral homeostasis in vivo. *Cell Death and Differentiation.* 21(9):1365–1376. doi:10.1038/cdd.2014.82.
- Isaac J, Nassif A, Asselin A, Taïhi I, Fohrer-Ting H, Klein C, Gogly B, Berdal A, Robert B, Fournier BP. 2018. Involvement of neural crest and paraxial mesoderm in oral mucosal development and healing. *Biomaterials.* 172:41–53. doi:10.1016/j.biomaterials.2018.04.036.
- Jiang D, Correa-Gallegos D, Christ S, Stefanska A, Liu J, Ramesh P, Rajendran V, De Santis MM, Wagner DE, Rinkevich Y. 2018. Two succeeding fibroblastic lineages drive dermal development and the transition from regeneration to scarring. *Nat Cell Biol.* 20(4):422–431. doi:10.1038/s41556-018-0073-8.

- 1
2
3 Kasperk C, Wergedal J, Strong D, Farley J, Wangerin K, Gropp H, Ziegler R, Baylink DJ. 1995.
4 Human bone cell phenotypes differ depending on their skeletal site of origin. *The Journal of clinical*
5 *endocrinology and metabolism*. 80(8):2511–2517. doi:10.1210/jcem.80.8.7629252.
6
7 Katsarelis H, Shah NP, Dhariwal DK, Pazianas M. 2015. Infection and medication-related
8 osteonecrosis of the jaw. *J Dent Res*. 94(4):534–539. doi:10.1177/0022034515572021.
9
10 Kelder C, Kleverlaan CJ, Gilijamse M, Bakker AD, de Vries TJ. 2020. Cells Derived from Human
11 Long Bone Appear More Differentiated and More Actively Stimulate Osteoclastogenesis Compared to
12 Alveolar Bone-Derived Cells. *Int J Mol Sci*. 21(14). doi:10.3390/ijms21145072.
13
14 Kingsmill VJ, McKay IJ, Ryan P, Ogden MR, Rawlinson SC. 2013. Gene expression profiles of
15 mandible reveal features of both calvarial and ulnar bones in the adult rat. *Journal of dentistry*.
16 41(3):258–264. doi:10.1016/j.jdent.2012.11.010.
17
18 de La Dure-Molla M, Fournier BP, Manzanares MC, Acevedo AC, Hennekam RC, Friedlander L,
19 Boy-Lefèvre M-L, Kerner S, Toupenay S, Garrec P, et al. 2019. Elements of morphology: Standard
20 terminology for the teeth and classifying genetic dental disorders. *Am J Med Genet A*. 179(10):1913–
21 1981. doi:10.1002/ajmg.a.61316.
22
23 Le Douarin NM. 1980. The ontogeny of the neural crest in avian embryo chimaeras. *Nature*.
24 286(5774):663–669. doi:10.1038/286663a0.
25
26 Lee J-T, Choi S-Y, Kim H-L, Kim J-Y, Lee H-J, Kwon T-G. 2015. Comparison of gene expression
27 between mandibular and iliac bone-derived cells. *Clin Oral Investig*. 19(6):1223–1233.
28 doi:10.1007/s00784-014-1353-8.
29
30 Leucht P, Kim JB, Amasha R, James AW, Girod S, Helms JA. 2008. Embryonic origin and Hox status
31 determine progenitor cell fate during adult bone regeneration. *Development*. 135(17):2845–2854.
32 doi:dev.023788 [pii] 10.1242/dev.023788.
33
34 Lian JB, Stein GS, Javed A, van Wijnen AJ, Stein JL, Montecino M, Hassan MQ, Gaur T, Lengner
35 CJ, Young DW. 2006. Networks and hubs for the transcriptional control of osteoblastogenesis. *Rev*
36 *Endocr Metab Disord*. 7(1–2):1–16. doi:10.1007/s11154-006-9001-5.
37
38 Liu H, Guo J, Wang L, Chen N, Karaplis A, Goltzman D, Miao D. 2009. Distinctive anabolic roles of
39 1,25-dihydroxyvitamin D(3) and parathyroid hormone in teeth and mandible versus long bones. *J*
40 *Endocrinol*. 203(2):203–213. doi:10.1677/JOE-09-0247.
41
42 Mah W, Jiang G, Olver D, Gallant-Behm C, Wiebe C, Hart DA, Koivisto L, Larjava H, Häkkinen L.
43 2017. Elevated CD26 Expression by Skin Fibroblasts Distinguishes a Profibrotic Phenotype Involved
44 in Scar Formation Compared to Gingival Fibroblasts. *Am J Pathol*. 187(8):1717–1735.
45 doi:10.1016/j.ajpath.2017.04.017.
46
47 Matsubara T, Suardita K, Ishii M, Sugiyama M, Igarashi A, Oda R, Nishimura M, Saito M, Nakagawa
48 K, Yamanaka K, et al. 2005. Alveolar bone marrow as a cell source for regenerative medicine:
49 differences between alveolar and iliac bone marrow stromal cells. *J Bone Miner Res*. 20(3):399–409.
50 doi:10.1359/JBMR.041117.
51
52 Molla M, Descroix V, Aïoub M, Simon S, Castañeda B, Hotton D, Bolaños A, Simon Y, Lezot F,
53 Goubin G, et al. 2010. Enamel protein regulation and dental and periodontal physiopathology in
54 MSX2 mutant mice. *Am J Pathol*. 177(5):2516–2526. doi:10.2353/ajpath.2010.091224.
55
56 Nassif A, Senussi I, Meary F, Loiodice S, Hotton D, Robert B, Bensidhoum M, Berdal A, Babajko S.
57 2014. Msx1 role in craniofacial bone morphogenesis. *Bone*. 66:96–104.
58 doi:10.1016/j.bone.2014.06.003.
59
60

- 1
2
3 Nowwarote N, Osathanon T, Kanjana K, Theerapanon T, Pornraveetus T, Shotelersuk V. 2019.
4 Decreased osteogenic activity and mineralization of alveolar bone cells from a patient with
5 amelogenesis imperfecta and FAM83H 1261G>T mutation. *Genes Dis.* 6(4):391–397.
6 doi:10.1016/j.gendis.2019.07.005.
7
- 8 Rawlinson SC, McKay IJ, Ghuman M, Wellmann C, Ryan P, Prajaneh S, Zaman G, Hughes FJ,
9 Kingsmill VJ. 2009. Adult rat bones maintain distinct regionalized expression of markers associated
10 with their development. *PLoS One.* 4(12):e8358. doi:10.1371/journal.pone.0008358.
11
- 12 Reichert JC, Gohlke J, Friis TE, Quent VM, Hutmacher DW. 2013. Mesodermal and neural crest
13 derived ovine tibial and mandibular osteoblasts display distinct molecular differences. *Gene.*
14 525(1):99–106. doi:10.1016/j.gene.2013.04.026.
15
- 16 Rinkevich Y, Walmsley GG, Hu MS, Maan ZN, Newman AM, Drukker M, Januszyk M, Krampitz
17 GW, Gurtner GC, Lorenz HP, et al. 2015. Skin fibrosis. Identification and isolation of a dermal
18 lineage with intrinsic fibrogenic potential. *Science.* 348(6232):aaa2151. doi:10.1126/science.aaa2151.
19
- 20 Ryu AH, Eckalbar WL, Kreimer A, Yosef N, Ahituv N. 2017. Use antibiotics in cell culture with
21 caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation.
22 *Sci Rep.* 7(1):7533. doi:10.1038/s41598-017-07757-w.
23
- 24 Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and
25 processing, Library preparation and sequencing, Computational data analysis, Cell type annotation,
26 Writing group, Supplemental text writing group, Principal investigators. 2018. Single-cell
27 transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* 562(7727):367–372.
28 doi:10.1038/s41586-018-0590-4.
29
- 30 Wang D, Gilbert JR, Zhang X, Zhao B, Ker DFE, Cooper GM. 2020. Calvarial Versus Long Bone:
31 Implications for Tailoring Skeletal Tissue Engineering. *Tissue Eng Part B Rev.* 26(1):46–63.
32 doi:10.1089/ten.TEB.2018.0353.
33
- 34 Wang KC, Helms JA, Chang HY. 2009. Regeneration, repair and remembering identity: the three Rs
35 of Hox gene expression. *Trends Cell Biol.* 19(6):268–275. doi:10.1016/j.tcb.2009.03.007.
36
- 37 Watanabe K, Lewis S, Guo X, Ni A, Lee BS, Deguchi T, Kim D-G. 2020. Regional variations of jaw
38 bone characteristics in an ovariectomized rat model. *J Mech Behav Biomed Mater.* 110:103952.
39 doi:10.1016/j.jmbbm.2020.103952.
40
- 41 Wu D, Mandal S, Choi A, Anderson A, Prochazkova M, Perry H, Gil-Da-Silva-Lopes VL, Lao R,
42 Wan E, Tang PL-F, et al. 2015. DLX4 is associated with orofacial clefting and abnormal jaw
43 development. *Hum Mol Genet.* 24(15):4340–4352. doi:10.1093/hmg/ddv167.
44
- 45 Yoshida T, Vivatbutsiri P, Morriss-Kay G, Saga Y, Iseki S. 2008. Cell lineage in mammalian
46 craniofacial mesenchyme. *Mech Dev.* 125(9–10):797–808. doi:10.1016/j.mod.2008.06.007.
47
- 48 Youlten SE, Kemp JP, Logan JG, Ghirardello EJ, Sergio CM, Dack MRG, Guilfoyle SE, Leitch VD,
49 Butterfield NC, Komla-Ebri D, et al. 2021. Osteocyte transcriptome mapping identifies a molecular
50 landscape controlling skeletal homeostasis and susceptibility to skeletal disease. *Nat Commun.*
51 12(1):2444. doi:10.1038/s41467-021-22517-1.
52
- 53 Zhou Q, Liu M, Xia X, Gong T, Feng J, Liu W, Liu Y, Zhen B, Wang Y, Ding C, et al. 2017. A mouse
54 tissue transcription factor atlas. *Nat Commun.* 8:15089. doi:10.1038/ncomms15089.
55
56
57
58
59
60

Figure and table legends

Figure 1. Osteoblasts display a bone site-specific transcriptional profile and mesenchymal identity in their physiological environment

(A) Workflow of the RNASeq transcriptional study of Md-OBs and Tb-OBs. (B) Total number of transcription factors (TFs) commonly expressed (OB-core), or overexpressed by Md-OBs (Md-Set) or by Tb-OBs (Tb-set). (C-E) Genes are classified according to their tissue affinities. (C) Gene ontology (GO) and enrichment analysis of TFs associated with mesenchymal tissues/cells in OB-core, Md-set and Tb-set (data are presented as $-\log_{10}$ (P value)). (D) RPKM expression heat map of TF genes in Md-set associated with mesenchymal tissues/cells (only RPKM values >1 are shown). (E) Genes known to be involved in different mesenchymal tissues but reported here for the first time as regulators of OBs and bone for the Md-set. (F) RT-qPCR analysis of *in vivo* gene expression of Sp6 in tissues isolated from C57BL/6JR mice at E10.5, E14.5, E16.5, E18.5, P1 and P9. mRNA levels are expressed in percentage of TF expression in P9 Md-bone. (G-I) Fluorescence imaging of functional OBs (GFP+) and Sp6-expressing cells (pink+) in the first branchial arch (BA1) at E10.5 (G) and in jawbone (JB) at E15.5 (H) and P9 (I) in *Coll1a1*2,3-GFP* mice (Scale bars: 200 μ m for the main images, 50 μ m for the inserts).

*BA1: first branchial arch, BP: biological process, GO: gene ontology, HL: Hindlimb, Inc: Incisor, JB: Jawbone, M: Meckel's cartilage, Md: Mandible, Mo: Molar, NT: neural tube, OB: osteoblast, OV: Otic vesicle, Tb: Tibia, TF: transcription factor, RPKM: Reads Per Kilobase of transcript per Million mapped reads, *site-specific genes (all the expressed transcripts of a given gene are increased in Md-OB), vXX: transcript variant identified by the last 2 numbers of the mRNA sequence NM number.*

Figure 2. The osteoblast transcriptome is characterized by site-specific developmental imprints

(A) GO and enrichment analysis of TF genes associated with developmental processes and morphogenesis in OB-core, Md-set and Tb-set. (B) Venn diagram of TF genes associated with neural crest (NC) cells in OB-core, Md-set and Tb-set. (C) RPKM expression heat map of TF genes in Md-set and Tb-set associated with NC cells (only RPKM values >1 are shown). (D-E) RT-qPCR analysis of *in vivo* gene expression of *Dlx3* (D) and *Sox9* (E) in tissues isolated from C57BL/6JR mice at E10.5, E14.5, E16.5, E18.5, P1 and P9. mRNA levels are expressed in percentage of TF expression in P9 Md-bone. (F-H) Fluorescence imaging of functional OBs (GFP+) and *Sox9*-expressing cells (pink+) in the first branchial arch (BA1) at E10.5 (F) and jawbone (JB) at E15.5 (G) and P9 (H) in *Coll1a1*2,3-GFP* mice (Scale bars: 200 μ m for the main images, 50 μ m for the inserts).

BA1: first branchial arch, BP: biological process, CNC: Cranial Neural Crest, GO: gene ontology, HL: Hindlimb, Inc: Incisor, JB: Jawbone, M: Meckel's cartilage, Md: Mandible, Mo: Molar, NT: neural tube, OB: osteoblast, OV: Otic vesicle, Tb: Tibia, TF: transcription factor, RPKM: Reads Per Kilobase of transcript

per Million mapped reads, *site-specific genes (all the expressed transcripts of a given gene are increased in Md-OB), vXX: transcript variant identified by the last 2 numbers of the mRNA sequence NM number.

Figure 3. Osteoblast embryonic positional identity is maintained *in vivo* and *ex vivo*

(A-D) Fluorescence imaging of functional OBs (GFP+) and NC-derived cells (Tomato+) in jawbone at E15.5 (A), P9 (B) and 2M (C) in Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice (Scale bars: 500µm for the main images, 50µm for the inserts). (D) Double labeling flow cytometry for GFP and Tomato of cells isolated from mandible and tibia of P9 Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice. (E) Quantification (cell counting) of NC-derived cells (tomato+) in mandible, maxillary and tibia from Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice at 4M. (F) RT-qPCR analysis of *in vivo* gene expression of Dlx3, Msx1, Msx2, Sox9, Hoxa10 and Hoxc10 in mandible and tibia bone tissues isolated from C57BL/6JR mice at P9, 2M and 8M. mRNA levels are expressed in percentage of TF expression in P9 Md-bone (for Dlx3, Msx1, Msx2, Sox9) or in P9 Tb-Bone (for Hoxa10 and Hoxc10). (G) Primary cell culture of osteoprogenitors isolated from mandible and tibia of P9 Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice. At confluency, cells were grown in osteogenic medium and observed at day 0 (D0), day 9 (D9) and day 18 (D18) (Scale bar: 400µm). (H) RT-qPCR analysis of *ex vivo* gene expression of Dlx3, Msx1, Msx2, Sox9, Hoxa10 and Hoxc10 in mandible and tibia osteoprogenitors. mRNA levels are expressed in percentage of TF expression in D0 mandible cells (for Dlx3, Msx1, Msx2, Sox9) or in D0 tibia cells (for Hoxa10 and Hoxc10).

Inc: Incisor, JB: Jawbone, M: Meckel's cartilage, Md: Mandible, Mo: Molar, NC: Neural Crest, OB: osteoblast, PH: phase, Tb: Tibia, TF: transcription factor, TG: Tooth Germ.

Figure 4: Human mutations associated with TF genes of jaw and tibia site-specific osteoblast signatures and their consequences in pathophysiology.

(A) RPKM expression heat map of TF genes in Md-set and Tb-set associated with human genetic disorders (only RPKM values >1 are shown). (B-C) OMIM and Orphanet based gene-by-gene analysis of genetic disorders and conditions caused by human mutations of Md-set and Tb-set according to tissue type and anatomical region. Pathological growth modification patterns are indicated with yellow arrows. (D) RTq-PCR analysis of *in vivo* gene expression of Alx3, Pax3 and Pax9 in mandible and tibia bone tissues isolated from C57BL/6JR mice at P9, 2M and 8M. mRNA levels are expressed in percentage of TF expression in P9 Md-bone.

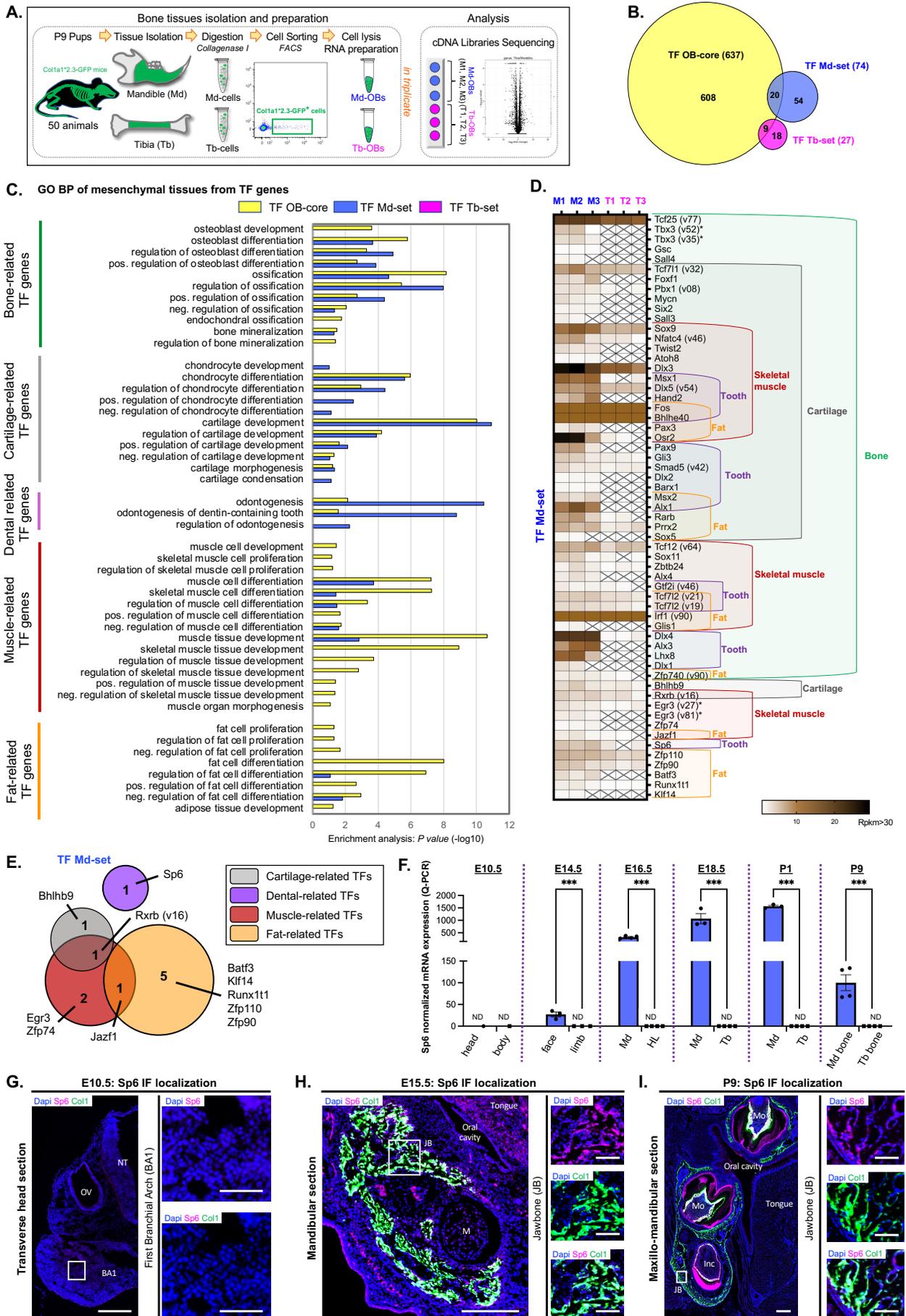
TF: transcription factor, OB: osteoblast, Md: Mandible, Tb: Tibia, RPKM: Reads Per Kilobase of transcript per Million mapped reads, *site-specific genes (all the expressed transcripts of a given gene are increased in Md-OBs), vXX: transcript variant identified by the last 2 numbers of the mRNA sequence NM number.

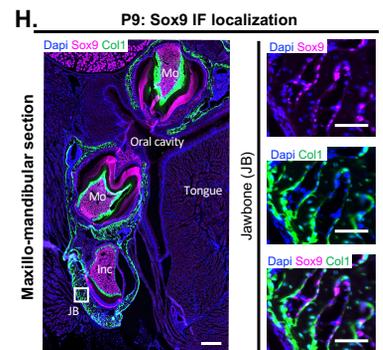
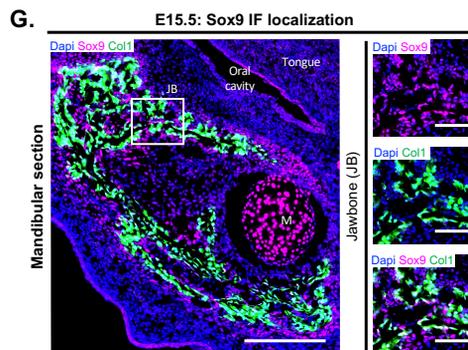
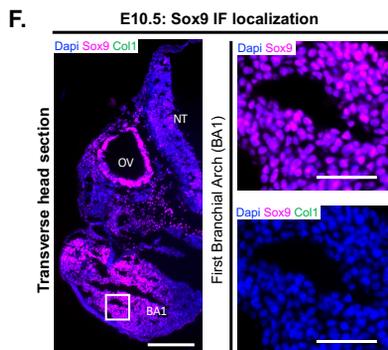
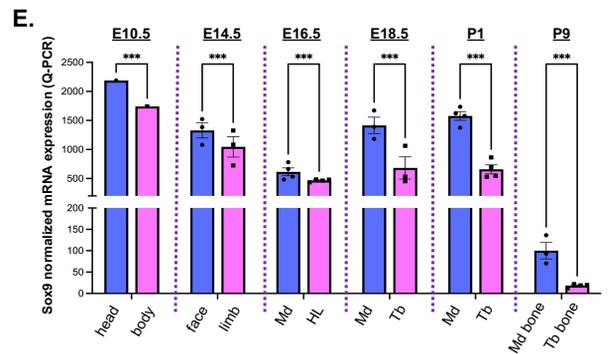
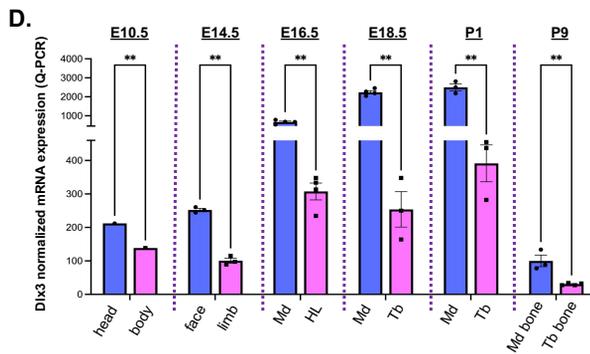
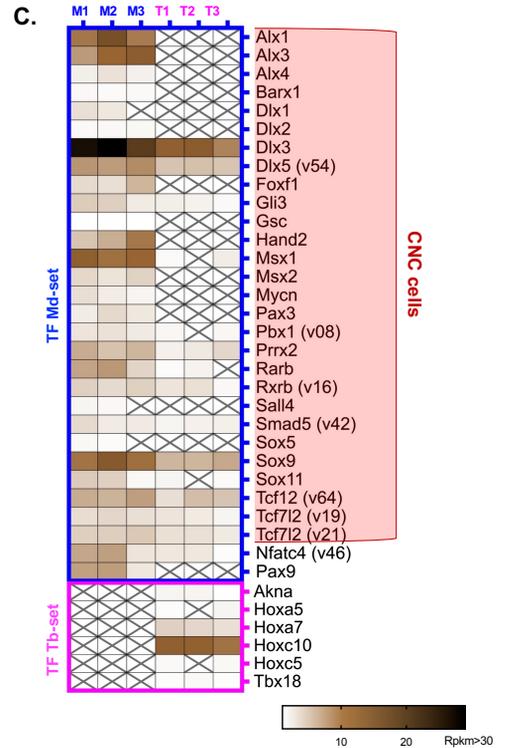
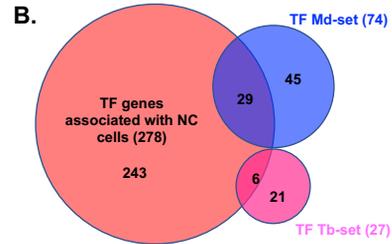
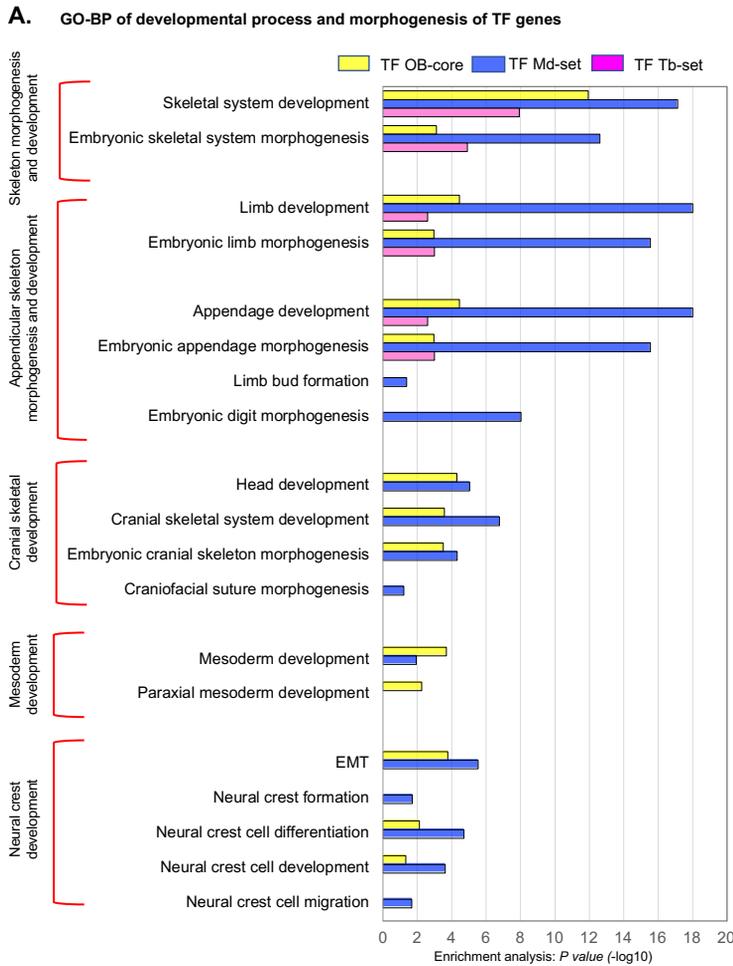
Abbreviations of human genetic disorders: AI: Amelogenesis imperfecta, AML: Acute myelogenous leukemia, AMS: Ablepharon-macrostomia syndrome, BRMUTD: Brain malformations with or without

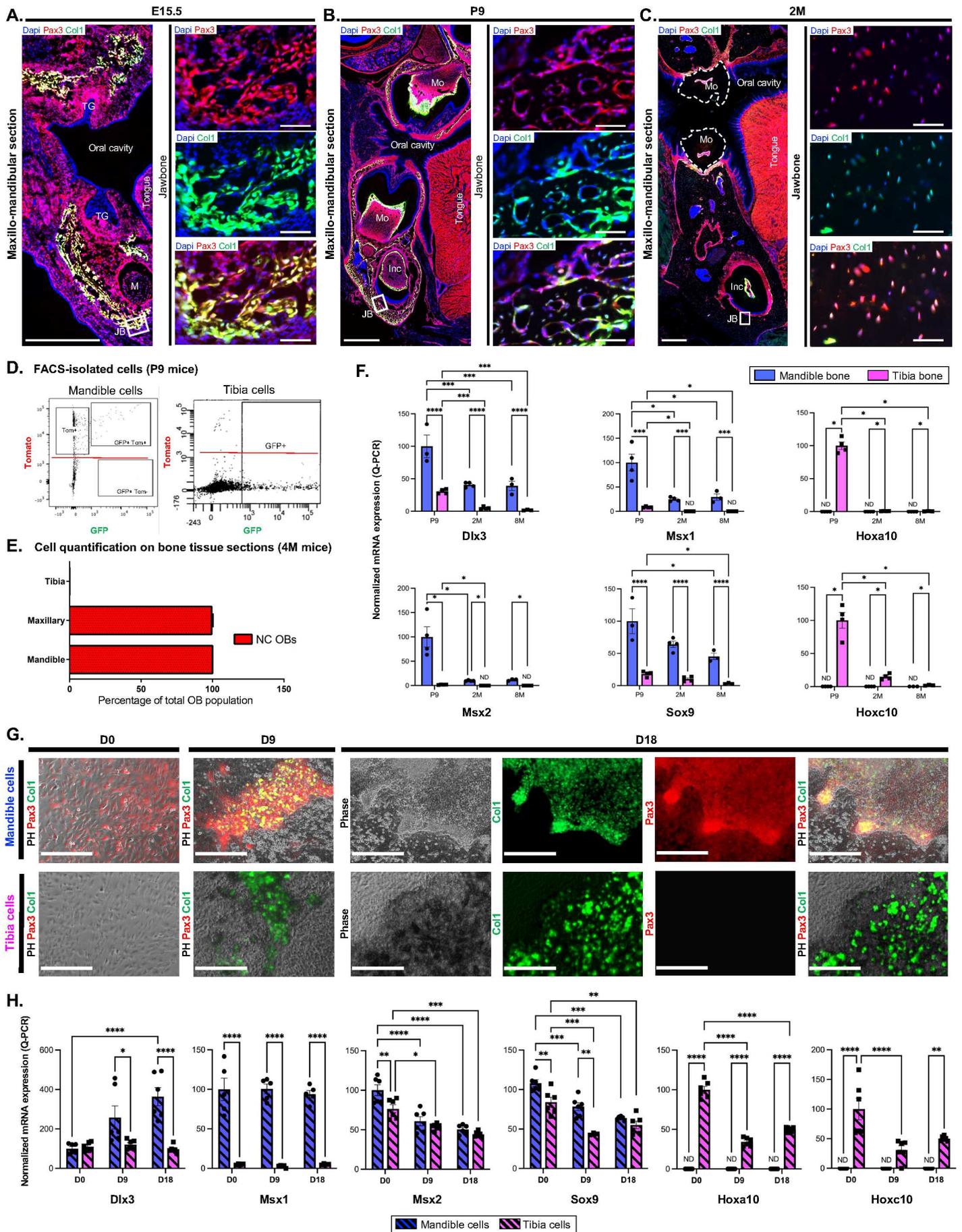
1
2
3 urinary tract defects/1p31p32 microdeletion syndrome, *CAKUTHEd*: Congenital anomalies of kidney and
4 urinary tract syndrome with or without hearing loss, abnormal ears, or developmental delay, *CDAN*:
5 Congenital dyserythropoietic anemia (type IV), *CDHS*: Craniofacial-deafness-hand syndrome, *CRS*:
6 Craniosynostosis (type 3 and 5), *DRRS*: Duane-radial ray syndrome (Okhiro syndrome), *FFDD3*: Focal
7 facial dermal dysplasia, *FND*: frontonasal dysplasia (type 1, 2 and 3), *GCPS*: Greig cephalopolysyndactyly
8 syndrome, *ICF2*: Immunodeficiency-centromeric instability-facial anomalies syndrome 2, *INLU*: Blood
9 group-Lutheran inhibitor, *IVIC*: Instituto Venezolano de Investigaciones Cientificas, *MDS*: Myelodysplastic
10 syndrome, *OFC*: Orofacial cleft (type 5), *PAP*: Polydactyly, postaxial (type A1 and B), *PFM*: Parietal
11 foramina 1, *PFMCCD*: Parietal foramina with cleidocranial dysplasia, *PHS*: Pallister-Hall syndrome, *PPD*:
12 Preaxial postdactyly (type 4), *PFM*: Parietal foramina, *RMS*: Rhabdomyosarcoma 2, alveolar, *SAMS*: Short
13 stature, auditory canal atresia, mandibular hypoplasia, skeletal abnormalities, *SHFM*: split hand/foot
14 malformation, *STHA*: selective tooth agenesis (type 3), *TDO*: trichodontoosseous syndrome, *WS*:
15 Waardenburg syndrome, type 1
16
17
18
19
20
21
22
23
24

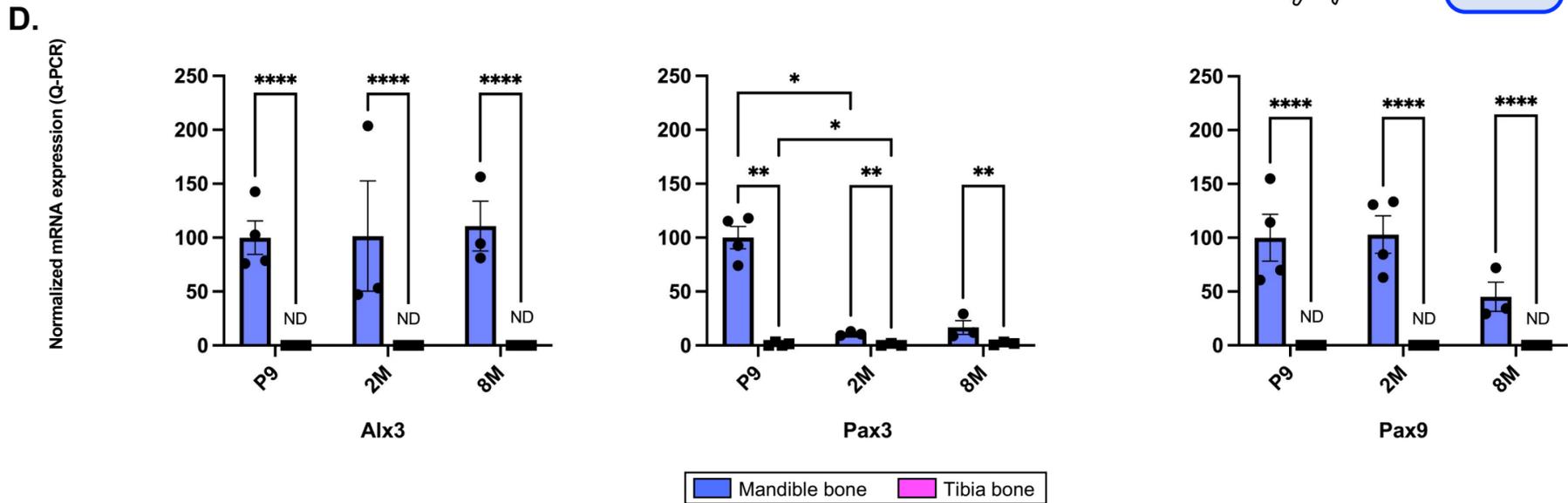
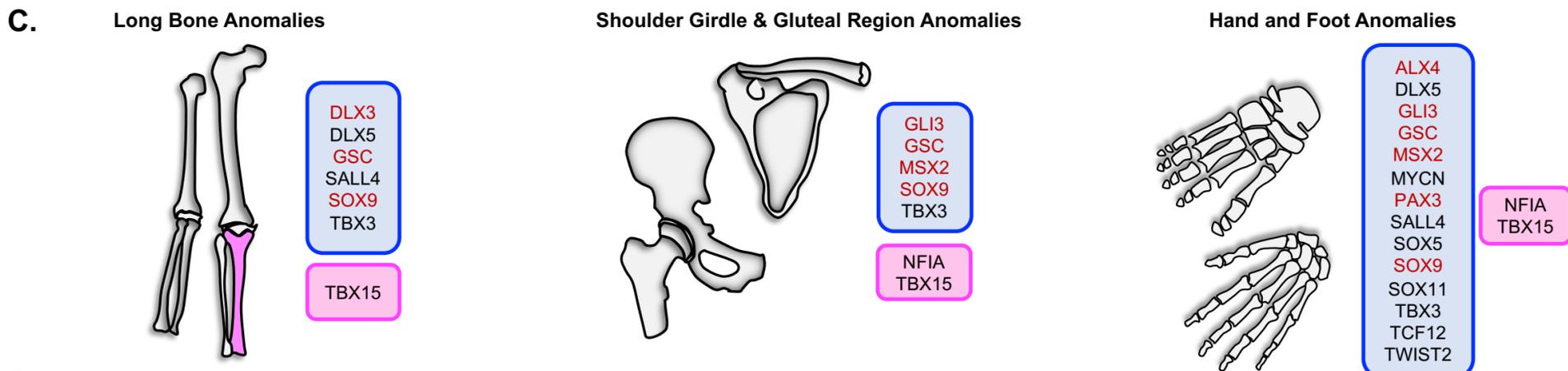
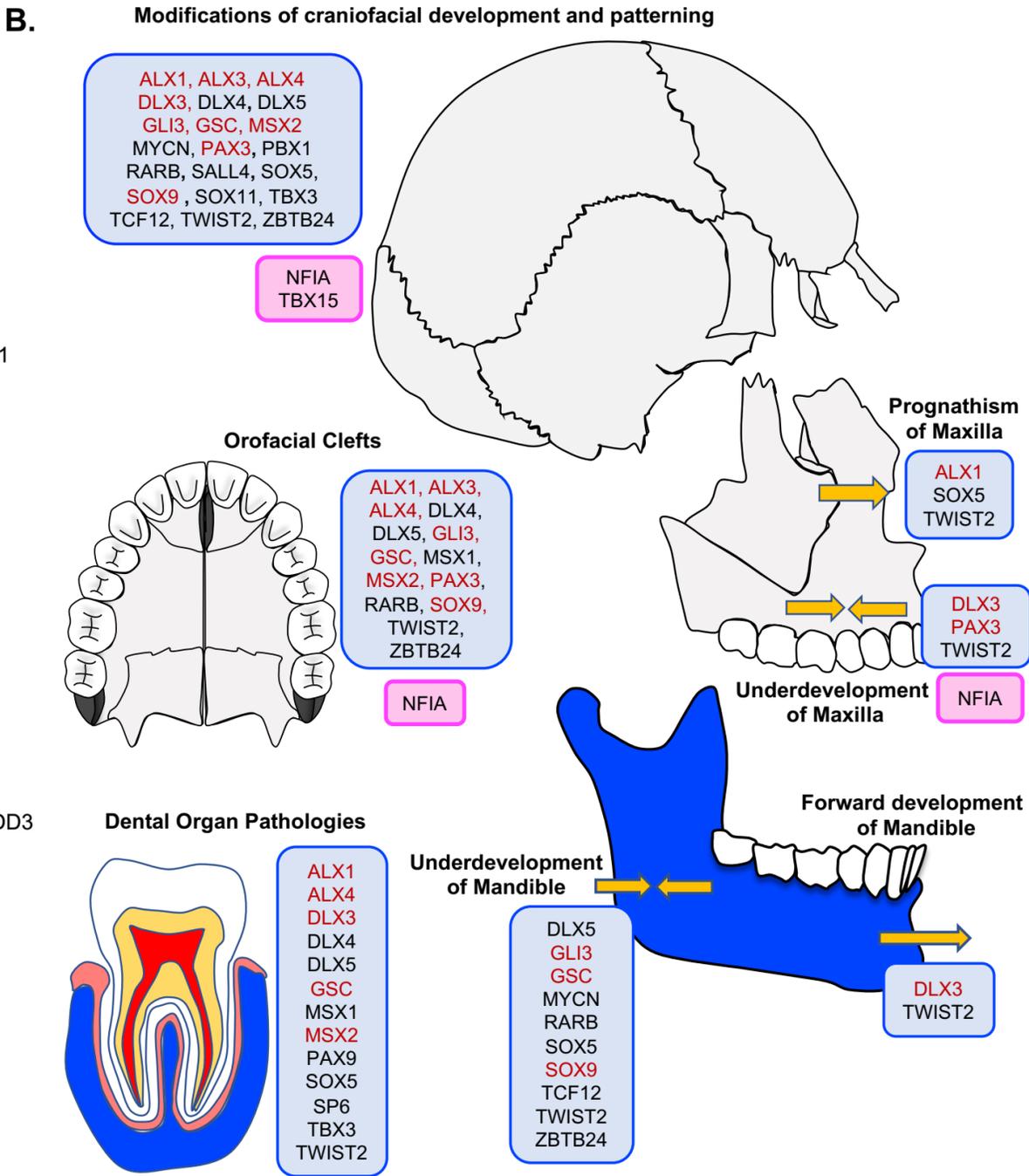
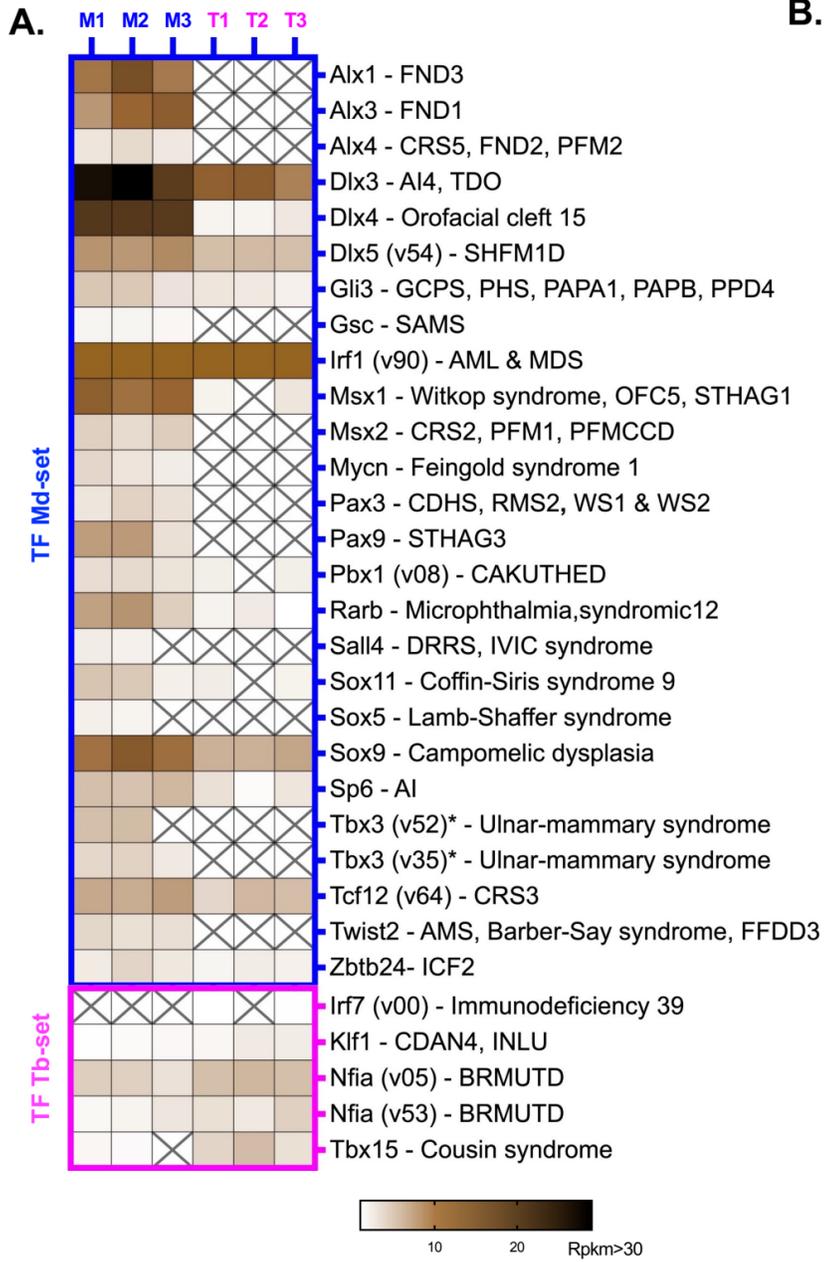
25 **Table 1: Functional enrichment analysis of GO terms in the OB-core and Md-set of TF genes**
26 **associated with biological processes categorized into “osteoblast development and**
27 **differentiation,” and “ossification and mineralization”.**
28
29

30 *TF*: transcription factor, *OB*: osteoblast, *Md*: Mandible, *Tb*: Tibia, *GO*: gene ontology, *BP*: biological process
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60









Bone site		Biological Process (BP) GO term	Category	# Genes	Gene Name	Fold Enrichment	p value	
OB-core	OB development & differentiation	osteoblast differentiation	GO:0001649	24	Cbfb, Cebpb, Cebpd, Creb3l1, Dlx5, Gli1, Gli2, Hey1, Junb, Jund, Lef1, Mef2c, Mef2d, Nfatc1, Phb, Runx2, Satb2, Smad1, Smad3, Smad5, Snai1, Snai2, Sp7, Twist1	3,2	1,6E-06	
		osteoblast development	GO:0002076	6	Gli2, Hey1, Jund, Runx2, Satb2, Smad3	10,0	2,4E-04	
		regulation of osteoblast differentiation	GO:0045667	14	Cebpb, Cebpd, Dlx5, Gli1, Hey1, Jund, Mef2c, Nfatc1, Runx2, Smad1, Smad3, Smad5, Snai2, Twist1	3,1	5,2E-04	
		positive regulation of osteoblast differentiation	GO:0045669	9	Cebpb, Cebpd, Dlx5, Hey1, Jund, Mef2c, Runx2, Smad1, Smad5	4,0	1,9E-03	
		osteoblast fate commitment	GO:0002051	3	Runx2, Smad1, Smad5	17,9	1,0E-02	
	Ossification & Mineralization		ossification	GO:0001503	39	Cbfb, Cebpb, Cebpd, Creb3l1, Dlx5, Foxc1, Gata1, Gli1, Gli2, Hey1, Hif1a, Junb, Jund, Klf10, Lef1, Mef2c, Mef2d, Nfatc1, Nfe2, Osr1, Phb, Rbpj, Runx1, Runx2, Runx3, Satb2, Scx, Smad1, Smad3, Smad5, Smad6, Snai1, Snai2, Sp1, Sp3, Sp7, Tcf7l2, Thra, Twist1	2,9	7,0E-09
			regulation of ossification	GO:0030278	23	Cebpb, Cebpd, Creb3l1, Dlx5, Egr2, Gata1, Gli1, Hey1, Hif1a, Jund, Mef2c, Nfatc1, Nfe2, Osr1, Pbx1, Rbpj, Runx2, Smad1, Smad3, Smad5, Smad6, Snai2, Twist1	3,2	3,9E-06
			positive regulation of ossification	GO:0045778	11	Cebpb, Cebpd, Creb3l1, Dlx5, Egr2, Gata1, Gli1, Hey1, Hif1a, Jund, Mef2c, Nfatc1, Nfe2, Osr1, Pbx1, Rbpj, Runx2, Smad1, Smad3, Smad5, Smad6, Snai2, Twist1	3,3	1,9E-03
			negative regulation of ossification	GO:0030279	9	Gata1, Hif1a, Mef2c, Nfatc1, Nfe2, Rbpj, Smad3, Smad6, Twist1	3,1	8,6E-03
			replacement ossification	GO:0036075	5	Dlx5, Mef2c, Mef2d, Runx2, Scx	5,0	1,7E-02
			enchondral ossification	GO:0001958	5	Dlx5, Mef2c, Mef2d, Runx2, Scx	5,0	1,7E-02
			bone mineralization	GO:0030282	8	Gata1, Hif1a, Klf10, Mef2c, Nfe2, Osr1, Smad3, Tcf7l2	2,6	3,4E-02
			regulation of bone mineralization	GO:0030500	7	Gata1, Hif1a, Mef2c, Nfe2, Osr1, Smad3, Twist1	2,8	3,9E-02
			Mid-Set	OB development & differentiation	regulation of osteoblast differentiation	GO:0045667	7	Dlx5, Gli3, Hand2, Msx2, Smad5, Sox11, Twist2
positive regulation of osteoblast differentiation	GO:0045669	5			Dlx5, Gli3, Msx2, Smad5, Sox11	18,7	1,4E-04	
osteoblast differentiation	GO:0001649	7			Dlx5, Gli3, Hand2, Msx2, Smad5, Sox11, Twist2	8,0	2,2E-04	
Ossification & Mineralization		regulation of ossification		GO:0030278	11	Dlx5, Gli3, Hand2, Msx2, Osr2, Pbx1, Six2, Smad5, Sox11, Sox9, Twist2	12,9	1,1E-08
		ossification		GO:0001503	10	Dlx5, Gli3, Hand2, Msx2, Osr2, Smad5, Sox11, Sox9, Tcf7l2, Twist2	6,4	2,2E-05
		positive regulation of ossification		GO:0045778	6	Dlx5, Gli3, Msx2, Osr2, Smad5, Sox11	15,3	4,3E-05
		negative regulation of ossification		GO:0030279	3	Hand2, Sox9, Twist2	8,8	4,5E-02
		bone mineralization		GO:0030282	3	Osr2, Sox9, Tcf7l2	8,3	4,9E-02

Transcriptional regulation of jaw osteoblasts: development to pathology

A. Nassif, G. Lignon, A. Asselin, C.C. Zadikian, S. Petit, H.W. Sun, C. Klein, F.C. Ferré, M.I. Morasso, A. Berdal, B.P.J. Fournier, J. Isaac

Appendix

Appendix Materials and Methods

Animal models and study design

To identify functional OBs, Col1a1*2,3-GFP mice that express GFP under the control of a 2.3-kb rat procollagen, type 1, alpha 1 promoter (JAX013134) were used (Kalajzic et al. 2002). Neural crests (NC) cells and their derivatives were investigated using a Pax3-cre mouse strain expressing Cre recombinase from the endogenous Pax3 locus (JAX-005549)(Lang et al. 2005). These mouse models were crossed with transgenic mice carrying a tomato flox/flox reporter, Rosa^{tomato} mice (JAX-008851) to obtain Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice. C57BL/6JR mice were used as an inbred strain and for a control. Mice were genotyped as previously described (Isaac et al. 2018).

Histology, immunofluorescence (IF) imaging and cell counting analyses

Heads and hindlimbs (HL) from transgenic mice at embryonic day 10.5 (E10.5), 14.5 (E14.5), 15.5 (E15.5), 16.5 (E16.5), 18.5 (E18.5), post-natal day 1 (P1), post-natal day 9 (P9), 2 months (2M), 4 months (4M) and 8 months (8M) were harvested, fixed, decalcified with EDTA (Sigma-Aldrich)

1
2
3 (for 2M, 4M and 8M), embedded in OCT (VWR), sectioned with cryostat and scanned using a
4
5 ZEISS Axio Scan.Z1 slide scanner (Zeiss) as previously described (Isaac et al. 2018). For IF
6
7 analyses, frozen sections were warmed at room temperature for 20 minutes and blocked with
8
9 10 mg/mL bovine serum albumin (BSA) (Sigma-Aldrich) and 0.1% Triton (Thermo Fischer
10
11 Scientific)-X100 in PBS (Thermo Fischer Scientific) for two hours. After two washes in 1X PBS,
12
13 sections were incubated overnight at 4 °C with rabbit monoclonal Runx2 antibody at 1:500
14
15 (Abcam, ab192256), rabbit monoclonal Sox9 antibody at 1:500 (Abcam, ab185966), rabbit
16
17 monoclonal Sp6 antibody at 1:500 (Abcam, ab246952) or rabbit monoclonal Sp7 antibody at
18
19 1:1000 (Abcam, ab209484). Sections were then washed and incubated for 2 hours at room
20
21 temperature with Alexa Fluor® 647 dye (Cy5) conjugated to donkey anti-rabbit IgG (1:500,
22
23 Thermo Fischer Scientific). Sections were mounted using DAPI Fluoromount-G® (Clinisciences).
24
25 For cell counting, regions of interest (ROI) were manually selected on tissue sections and cell
26
27 counts using Cyanine-5 (Cy5), Tomato (Tom), GFP and DAPI signals were determined by using
28
29 Oncotopix software (Visiopharm) and QuPath v0.3.0 (Bankhead et al. 2017). The number of GFP-
30
31 positive/Tom-positive and GFP-positive/Tom-negative cells or the number of GFP-positive/Cy5-
32
33 positive and GFP-positive/Cy5-negative cells was then divided by the total number of bone-
34
35 forming osteoblastic cells (GFP-positive) for each tissue section to obtain percentages. To limit
36
37 artifacts, only DAPI-positive cells were counted.
38
39
40
41
42
43
44
45
46
47
48
49

50 **Bone microdissection**

51
52
53 For mouse E10.5 embryos, heads were meticulously dissociated from the body. At E14.5, the
54
55 frontal part of the heads (referred as “Face”) and the limbs were isolated. Of note, both forelimbs
56
57
58
59
60

1
2 (FL) and hindlimbs (HL) were pooled and referred as “Limb”. At E16.5, E18.5 and P1, the
3
4 mandibles (with developing teeth) were meticulously isolated and referred as “Mandible (Md)”.
5
6 At E16.5, whole hindlimbs (HL) were isolated. At E18.5 and P1, tibias (Tb) were individualized.
7
8 For mandibles at P9, 2M and 8M, the ascending branch, muscle tissue, molars and incisors were
9
10 meticulously removed to preserve only basal and alveolar bones. For tibias, cartilage caps and
11
12 epiphyses were removed, bone marrow cells were flushed, and the remaining cortical bone was
13
14 harvested. All dissections were performed under a microscope (Zeiss).
15
16 The obtained tissue fragments were then rinsed in 1X PBS and processed for cell isolation (FACS
17
18 and primary cell culture) and for RNA extraction and analyses (RNA-seq and RT-qPCR).
19
20
21
22
23
24
25
26
27

28 **Cell isolation and primary cell culture**

29
30 The aseptically dissected mandibles and tibias of Col1a1*2,3-GFP and Pax3-cre::Col1a1*2,3-
31
32 GFP::Rosa^{tomato} P9 mice were enzymatically digested with 0.25% collagenase type I (Sigma-
33
34 Aldrich) in 1X PBS for 2 hours at 37 °C (two successive one-hour incubations). Dissociated cells
35
36 were filtered using a 70-micron filter (Falcon, Thermo Fisher Scientific) and plated at a density of
37
38 8.4×10^4 cells/cm² in proliferative medium composed of DMEM High Glucose (Gibco), 10% fetal
39
40 bovine serum (FBS) (Invitrogen) and 1% penicillin/streptomycin (Invitrogen). At confluency
41
42 (D0), cells were grown in osteogenic medium composed of DMEM (Gibco) supplemented with
43
44 10% FBS, 50 µg/mL ascorbic acid (Sigma), 10 mM β-glycerophosphate (Sigma), and
45
46 penicillin/streptomycin 1%. The cells were maintained at 37 °C in a fully humidified atmosphere
47
48 in air with 5% CO₂. Culture medium was changed at 48-h intervals. Cell RNA was isolated at Day
49
50
51
52
53
54
55
56
57
58
59
60 0, D9 and D18.

RNA extraction and RT-qPCR

Dissected bone samples were put in TRI-Reagent (Euromedex) and homogenized with a tissue homogenizer (Omni International) and cellular layers were scraped in TRI-reagent. RNA extraction, reverse transcription and RT-qPCR analysis were performed as previously described (Jacques et al. 2014). Target genes were run in triplicate for each sample and normalized against the geometric mean of the mRNA levels of the housekeeping genes Hprt and Tbp. The specific primers used are listed in **Appendix Table 1**.

Flow cytometry cell sorting (FACS)

To analyze the cellular distribution in bone tissues and isolate cells for transcriptomics, FACS analysis was performed on the enzymatically dissociated and filtered cells from bone fragments of Col1a1*2,3-GFP and Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice using a BD LSRII instrument (BD) to collect cells according to their osteoblastic phenotype and embryonic origin.

RNA Sequencing (RNA-Seq)

Mandible and tibia bones from P9 Col1a1*2,3-GFP mice were dissected as described in the “**Bone microdissection**” section. Three independent dissections, each pooling bone tissues from approximately fifty P9 mice, were carried out. Immediately after dissection and collagenase I digestion, GFP-positive cells were FACS-sorted as described above, RNA was extracted using a single-cell RNA purification kit (Norgen) and RNA quantification and integrity were performed with a Bioanalyzer RNA (Agilent). Based on RNA quantification and RIN number, 3 experiments with RNA quantity ranging from 2 to 8 ng were selected and RNAseq analysis was performed by Oxford Gene Technology (OGT). Briefly, generation of cDNA was performed using SMART-Seq

1
2 v4 Ultra Low Input RNA Kit (Takara) and processed with an Illumina TruSeq RNA Sample
3 Preparation kit (Illumina). Data were generated with an Illumina HiSeq 2000 sequencing system.
4 Raw sequencing data were processed as previously described (Isaac et al. 2014) with CASAVA
5 1.8.2 to generate FASTQ files. Reads of 100 bases were mapped to the mouse transcriptome and
6 genome mm10 using TopHat version 2.1.0. Gene expression Reads Per Kilobase of transcript per
7 Million mapped reads (RPKM) values and ANOVA scores were calculated using Partek GS 6.6.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Protocols of RNA extraction, cDNA library construction and data processing pipeline are publicly available in the Gene Expression Omnibus (GEO), together with the RNA-Seq raw and processed data files (accession# GSE186535)

Bioinformatics resources

The RNA-seq database was cross-referenced with a list of 1388 transcription factors (TF) generated using three databases [TcoF-DB v2 (Schmeier et al. 2017), TFcheckpoint (Chawla et al. 2013) and Universal Protein Resource (The UniProt Consortium 2021)]. This TF list has been deposited in the Gene Expression Omnibus (GEO) and is publicly available (accession# GSE186535). RNA-Seq-based transcriptome annotation and analysis and gene-ontology (GO) enrichment were performed using the DAVID Functional Annotation Clustering tool 6.8 (Huang et al. 2009a; Huang et al. 2009b), AmiGO2 (Carbon et al. 2009) and Metascape (Zhou et al. 2019).

Statistical analyses

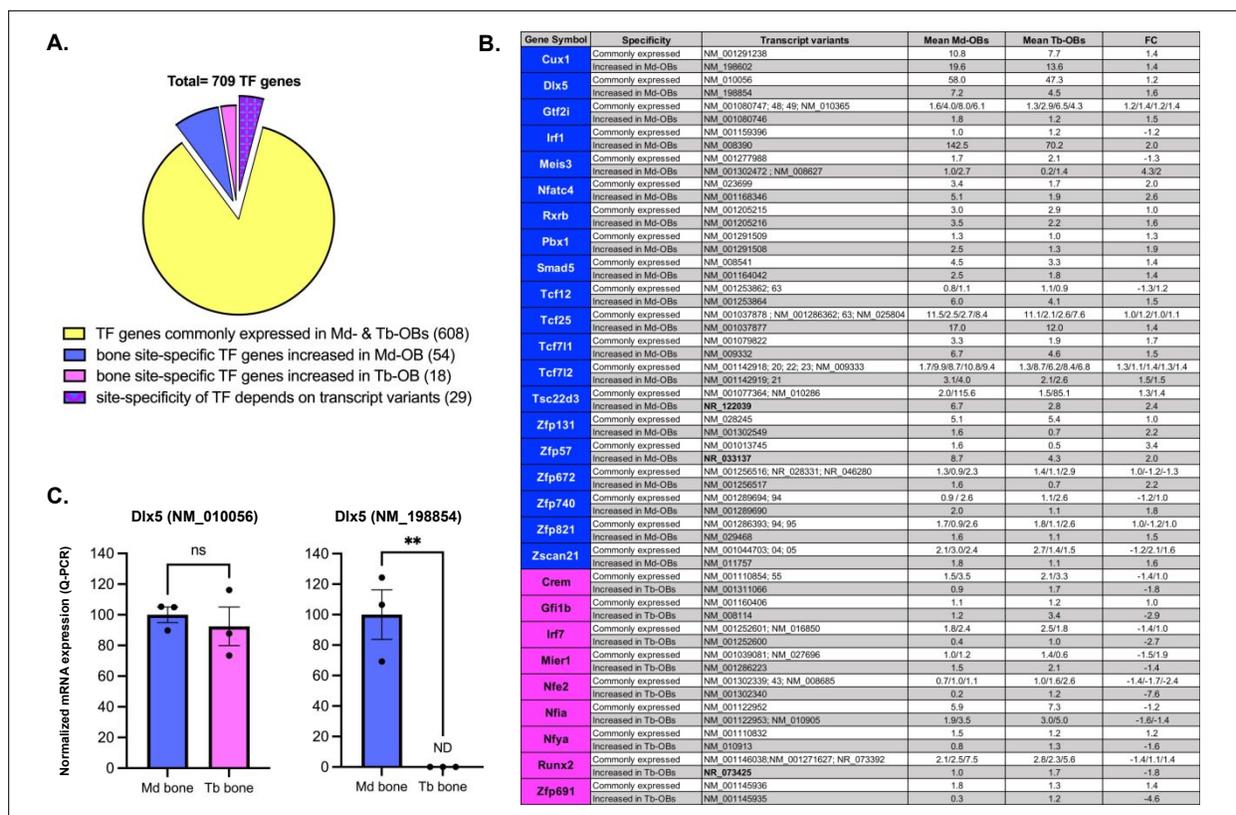
All quantitative measurements were performed on at least three different animals per stage and per tissue. For cell counting, between 2 and 5 slides per animal were scanned and analyzed. Three independent primary cell cultures were performed, and one representative experiment is presented.

Data presentation and statistical analyses were performed using Prism 9 software (GraphPad Software). After the normality was tested using the Shapiro-Wilk test, statistical analyses were

1
2 performed according to the number of parameters using either an unpaired t-test or two-way
3
4 ANOVA followed by Tukey's multiple comparisons test. Data are expressed as mean \pm standard
5
6 error of mean (SEM) and significance values are shown in the figures. Non detected (ND), $p > 0.5$
7
8 (ns), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), $p \leq 0.0001$ (****). Area-proportional Venn diagrams
9
10
11 were designed using BioVenn (Hulsen et al. 2008).
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Appendix figures and figure legends



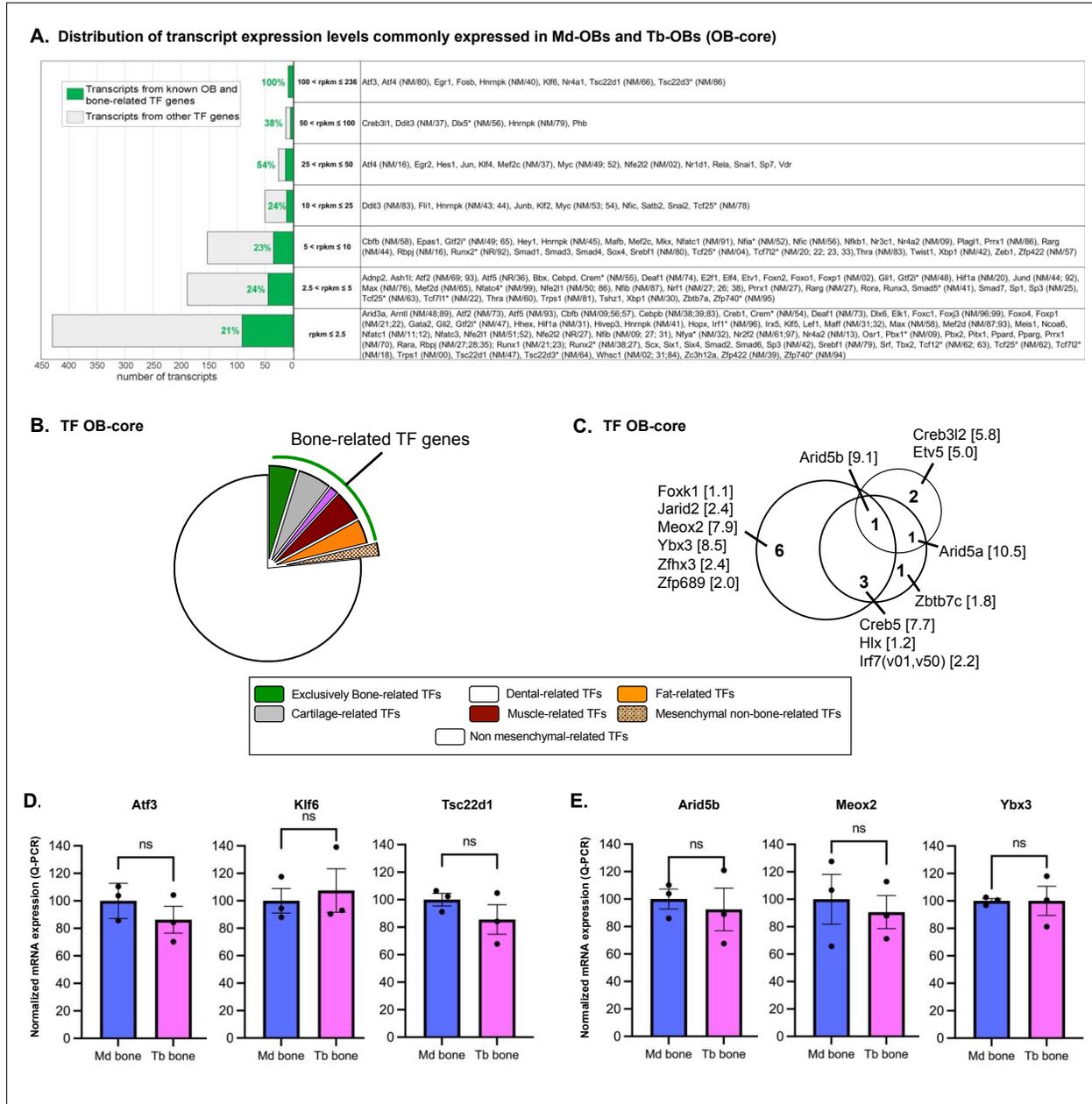
Appendix Figure 1: Differential expression of mRNA transcript variants encoded by TF genes.

(A) Pie-charts showing TF gene distribution based on bone site-specificity: 608 TF genes are commonly expressed in Md-OB and Tb-OBs, 54 TF genes are increased in Md-OB, 18 TF genes are increased in Tb-OB and the bone site-specificity of 29 genes depended on their transcript variants (TVs). (B) RNA-seq results of the TF genes whose bone site-specificity depends on their TVs. The non-coding RNA sequences (NR) are indicated in bold when they are the only differentially expressed transcripts from a given gene. For Tsc22d3, Zfp57 and Runx2, the only site-specific TVs are non-coding RNA sequences (NR_122039, NR_033137 and NR_073425, respectively). Because NR sequences are not expected to be translated into functional proteins, the 3 corresponding TFs are considered as commonly expressed in Md-OBs and Tb-OBs in this study. (C) Validation of differential expression of bone site-specific expression of RNA-seq-predicted Dlx5 TV in Md and Tb raw bone tissues from P9 C57BL/6J mice. Relative expression of two Dlx5 TVs was assessed through RT-qPCR (n = 3 mice per tissue). NM_010056 shows no significant difference between Md- and Tb-bones, whereas NM_198854 is significantly increased in Md-bone when compared to Tb-bone. mRNA levels are expressed in percentage of TF expression in P9 Md-bone. *TF*:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

transcription factor, OB: osteoblast, Md: Mandible, Tb: Tibia, TV: transcript variant, NM: mRNA RefSeq accession prefixes for validated protein-coding transcripts, NR: RNA RefSeq accession prefixes for non-protein-coding transcripts, FC: Fold Change.

For Peer Review

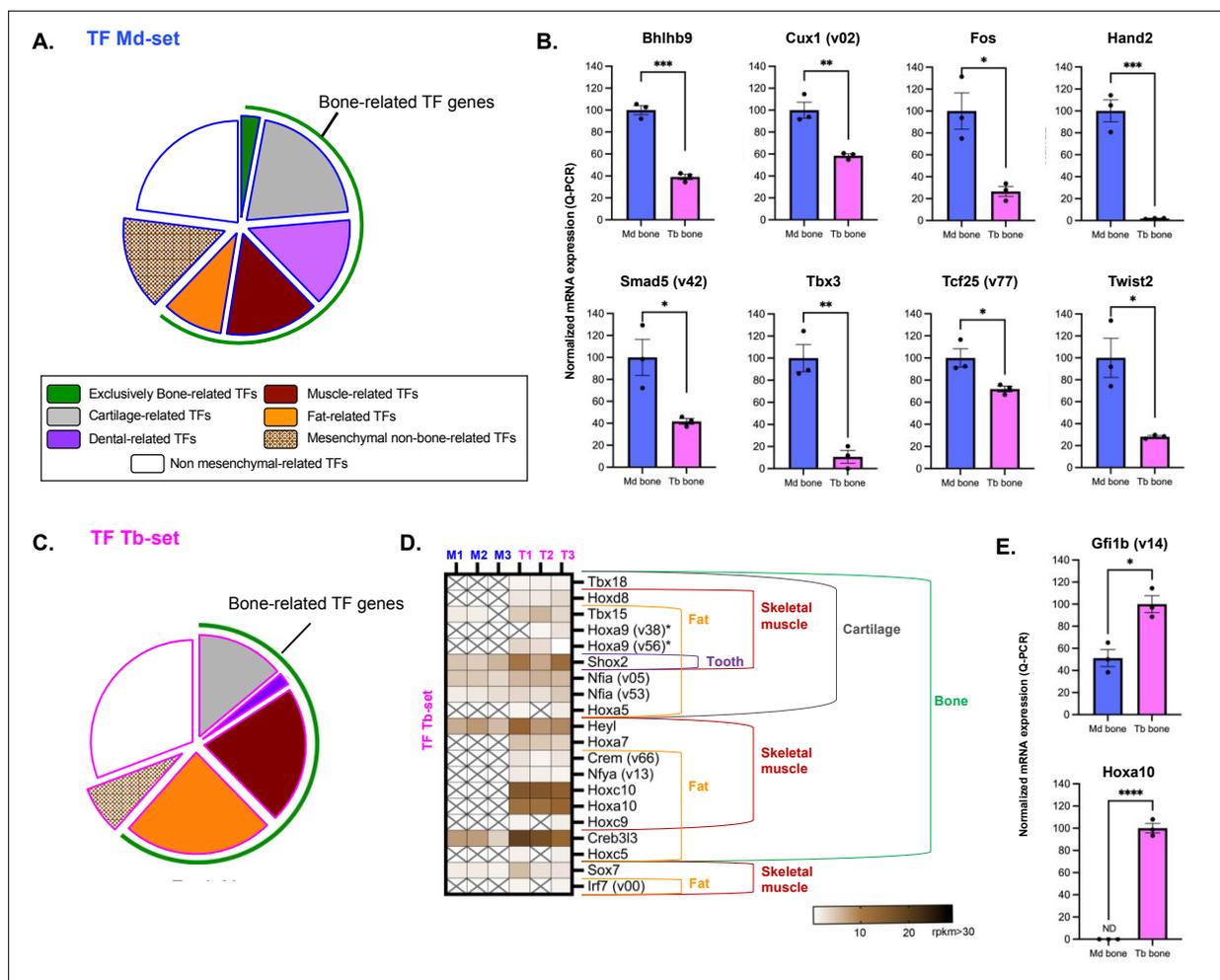


Appendix Figure 2: TF genes in the OB-core associated with OB and bone and other mesenchymal tissues/cells

Genes are classified according to their tissue affinities. (A) Distribution of OB-core TF gene expression based on their level of expression and their previously identified roles in OB and bone regulation. (B) Distribution of bone-related TF genes out of the total TF genes from OB-core. (C) Genes known to be involved in different mesenchymal tissues but reported here for the first time as regulators of OBs and bone

1
2 for OB-core. (D-E) RTq-PCR analysis of *in vivo* gene expression of 3 TF genes from OB-core previously
3 known for OB and bone regulation (Atf3, Klf6, Runx2 and Tsc22d1) and 3 TF genes from OB-core newly
4 described in differentiated OB (Arib5b, Meox2 and Ybx3) in mandible and tibia bone tissues isolated from
5 C57 BL/6JR mice at P9. mRNA levels are expressed in percentage of TF expression in P9 Md-bone. No
6 difference in expression is observed for these 6 genes between bone and tibia. *TF: transcription factor, OB:*
7 *osteoblast, Md: Mandible, Tb: Tibia, RpkM: Reads Per Kilobase of transcript per Million mapped reads,*
8 **site-specific genes (all the expressed transcripts of a given gene are commonly expressed in Md-OBs and*
9 *Tb-OBs), vXX: transcript variant identified by the last 2 numbers of the mRNA sequence NM number.*
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review



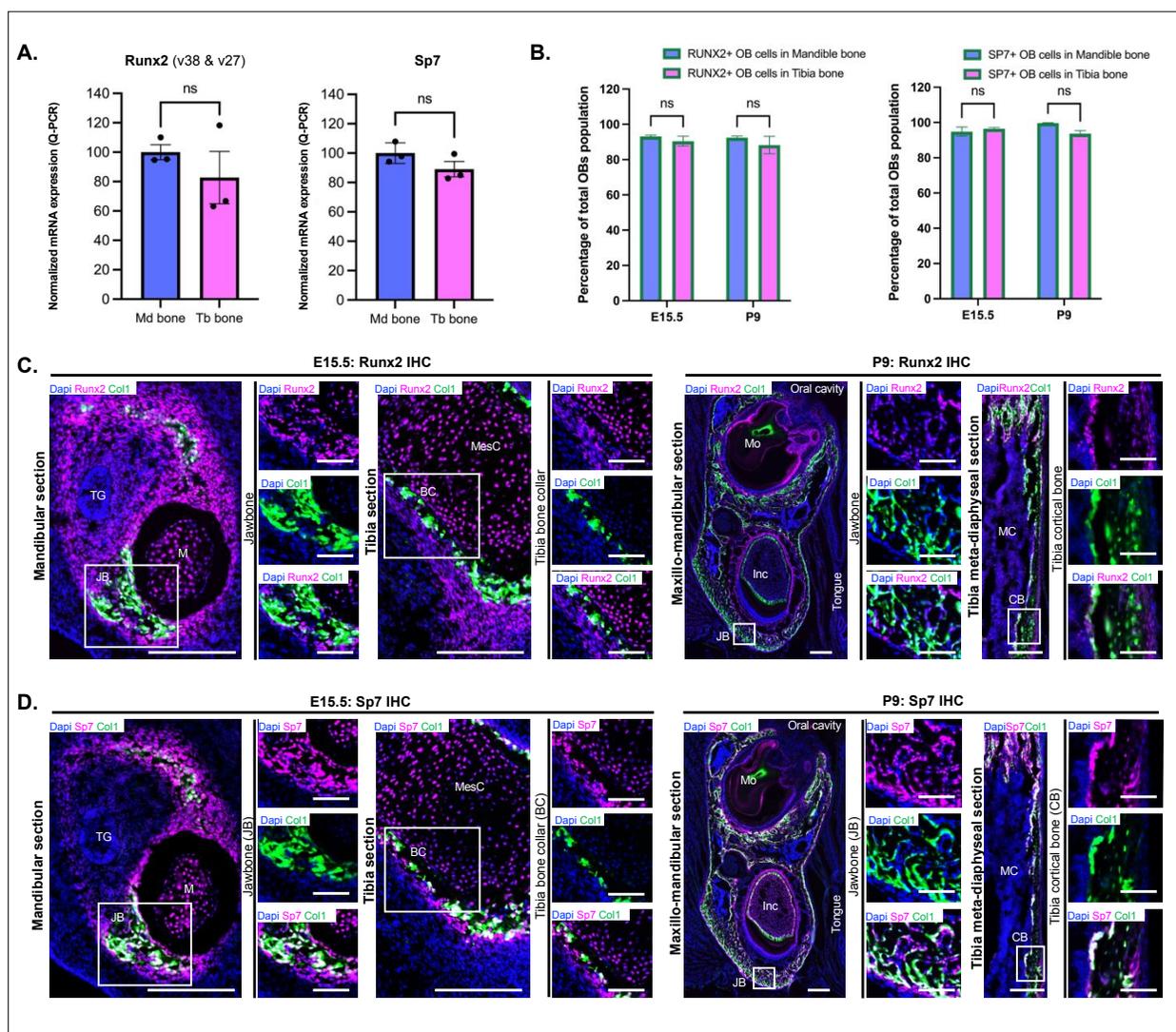
Appendix Figure 3: TF genes in the Md-set and Tb-set

Genes are classified according to their tissue affinities. (A) Distribution of bone-related TF genes out of the total TF genes from Md-set. (B) RT-qPCR analysis of *in vivo* gene expression of 8 TF genes from Md-set (Bhlhb9, Cux1, Fos, Hand2, Smad5, Tbx3, Tcf25, Twist2) in mandible and tibia bone tissues isolated from C57 BL/6JR mice at P9. (C) Distribution of bone-related TF genes from Tb-set. (D) Expression heat map of TF genes in the Tb-set associated with mesenchymal tissues/cells (rpkm values >1 are shown). (E) RT-qPCR analysis of *in vivo* gene expression of 2 TF genes from Tb-set (Gfi1b, Hoxa10) in mandible and tibia bone tissues isolated from C57 BL/6JR mice at P9. mRNA levels are expressed in percentage of TF expression in P9 Md-bone (B) or in P9 Tb-bone (E).

TF: transcription factor, OB: osteoblast, Md: Mandible, Tb: Tibia, Rpkms: Reads Per Kilobase of transcript per Million mapped reads, *site-specific genes (all the expressed transcripts of a given gene are increased in Tb-OB), vXX: transcript variant identified by the last 2 numbers of the mRNA sequence NM number.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review



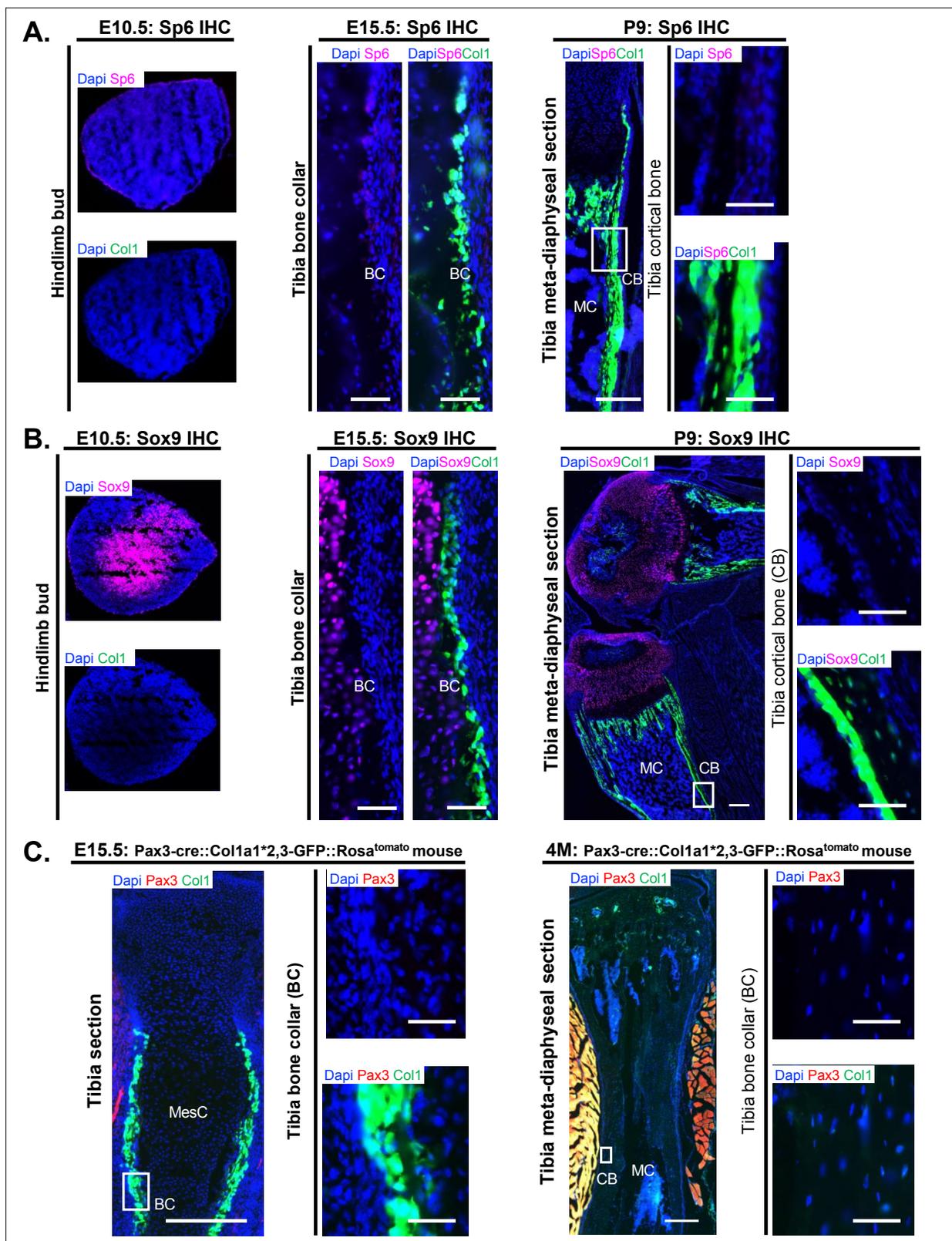
Appendix Figure 4: *In vivo* expression of Runx2 and Sp7, two TFs commonly expressed in Md-OBs and Tb-OBs, using RT-qPCR, IF imaging and cell counting analyses.

(A) RT-qPCR analysis of *in vivo* gene expression of Runx2 and Sp7 in mandible and tibia bone tissues isolated from C57BL/6JR mice at P9. mRNA levels are in percentage of TF expression in P9 Md-bone. (B) Histological quantification of Runx2-expressing or Sp7-expressing (pink+) functional OBs (GFP+) in mandible and tibia bone tissues from Col1a1*2,3-GFP mice at E15.5 and P9 using Runx2 and Sp7 antibodies, respectively. (C) Fluorescence imaging of functional OBs (GFP+) and Runx2-expressing cells (pink+) in jawbone and in tibia bone at E15.5 and P9 in Col1a1*2,3-GFP mice, using Runx2 antibody. (D) Fluorescence imaging of functional OBs (GFP+) and Sp7-expressing cells (pink+) in jawbone and in tibia bone at E15.5 and P9 in Col1a1*2,3-GFP mice, using Sp7 antibody (Scale bars: 200µm for the main images, 50µm for the inserts).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BC: Bone Collar, CB: Cortical Bone, Inc: Incisor, JB: JawBone, M: Meckel's cartilage, Md: Mandible, MC: Medullary Cavity; MesC: Mesenchymal Condensation, Mo: Molar, OB: osteoblast, Tb: Tibia, TG: Tooth Germ.

For Peer Review



1
2
3 **Appendix Figure 5: *In vivo* expression in appendicular bone tissues of selected TFs and analysis**
4 **of embryonic origin**

5
6 (A) Fluorescence imaging of functional OBs (GFP+) and Sp6-expressing cells (pink+) in tibia bone from
7 E10.5, E15.5 and P9 Col1a1*2,3-GFP mice, using Sp6 antibody. (B) Fluorescence imaging of functional
8 OBs (GFP+) and Sox9-expressing cells (pink+) in tibia bone from E10.5, E15.5 and P9 Col1a1*2,3-GFP
9 mice, using Sox9 antibody. (C) Fluorescence imaging of functional OBs (GFP+) and NC-derived cells
10 (Tomato+) in tibia bone from E15.5 and 4M Pax3-cre::Col1a1*2,3-GFP::Rosa^{tomato} mice. (Scale bars:
11 200µm for the main images, 50µm for the inserts).

12
13 *BC: Bone Collar, CB: Cortical Bone, MC: Medullary Cavity; MesC: Mesenchymal Condensation.*
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix Tables

Strain name	Transcript variant for primer design	Forward Primers (5'-3')	Reverse Primers (5'-3')	Product size (pb)
mAlx3	NM_007441	TGG-CCA-AGA-CCA-AGA-GCA-AG	TTC-TGG-AAC-CAG-ACC-TGT-ACC-C	172
mArid5b	NM_023598.2	CCA-CCT-TGA-AGG-CAA-ACC-AAG	AAT-GCG-CCC-ATT-TGA-CAA-GG	261
mAtf3	NM_007498	GGA-AGA-GCT-GAG-ATT-CGC-CA	CTC-ATC-TTC-TTC-AGG-GGC-CG	136
mBhlhb9	NM_001098222	TGG-CCA-GGA-TCC-ATT-TTC-TCA	AAG-GCA-GCA-GAA-CAC-AAA-GC	158
mCux1	NM_198602	CTG-GAG-GAA-GCT-GAG-CAC-AA	GCA-CCT-CTA-TGG-CCA-CAT-CT	274
mDlx3	NM_010055	ATT-ACA-GCG-CTC-CTC-AGC-AT	CTT-CCG-GCT-CCT-CTT-TCA-C	223
mDlx5	NM_198854.2	CAC-CAG-CCA-GCC-AGC-TTT-C	CTG-GTG-ACT-GTG-GCG-AGT-TA	241
mDlx5	NM_010056 (common)	TTC-CAA-GCT-CCG-TTC-CCG-A	CTG-GCT-CCG-CCA-CTT-CTT-TC	328
mFos	NM_010234	GGC-TTT-CCC-CAA-ACT-TCG-AC	CTG-CGC-AAA-AGT-CCT-GTG-TG	175
mGfi1b	NM_008114	CAA-GGT-GTT-CTC-CAC-CCC-TC	GAA-GCT-TCG-CTC-CTG-TGA-GT	154
mHand2	NM_010402.4	AGA-AGA-GGA-AGA-AAG-AGC-TGA-ATG	ACC-GGA-TGA-TCC-AAC-TGC-AA	228
mHoxa10	NM_001122950	ACT-AAG-AGC-AGC-ACG-GTA-CG	GCT-GCA-TTT-TCG-CCT-TTG-GA	212
mHoxc10	NM_010462	CCA-GTC-CAG-ACA-CCT-CGG-AT	ATA-GGG-GCA-CCT-CTT-CTT-CCT	107
mHPRT	NM_013556.2	GTT-GGA-TAT-GCC-CTT-GAC-TAT-AAT-GA	CAA-CAT-CAA-CAG-GAC-TCC-TCG-TAT-T	141
mKlf6	NM_011803	TTG-AAA-GCA-CAT-CAG-CGC-AC	AGC-CTA-CAG-GAT-TCG-TCC-CT	232
mMeox2	NM_008584	AGC-GAC-AGC-TCA-GAT-TCC-CA	TTT-GAC-CCG-CTT-CCA-CTT-CAT	234
mMsx1	NM_010835	TCC-TCA-AGC-TGC-CAG-AAG-AT	TAC-TGC-TTC-TGG-CGG-AAC-TT	220
mMsx2	NM_0103601	CCT-GAG-GAA-ACA-CAA-GAC-CA	AGT-TGA-TAG-GGA-AGG-GCA-GA	278
mPax3	NM_011041	CAC-TGT-GCC-CTC-AGT-GAG-TT	TGA-AGG-TGG-TTC-TGC-TCC-TG	242
mPax9	NM_011041	CCA-CCT-CTT-TTG-GGG-TGT-GT	GAA-GGC-TGG-CTC-CAT-TGC-T	300
mRunx2	NM_001146038 / NM_001271627	GGG-AAC-CAA-GAA-GGC-ACA-GA	TGG-AGT-GGA-TGG-ATG-GGG-AT	279
mSmad5	NM_001164042	GGA-ACC-CTA-AGC-TCT-GGG-AA	CTT-AGT-GCA-AGT-CCT-CGA-CCA	121
mSox9	NM_011448	CAC-AAG-AAA-GAC-CACC-CCG-A	GGA-CCC-TGA-GAT-TGC-CCA-GA	209
mSp6	NM_031183	CAC-TGC-TTC-TGC-CAG-CTC-T	CCA-CAG-ACA-GCG-GTT-AGC-AT	111
mSp7	NM_001348205 / NM_130458	CCT-AGG-TTA-TCT-CCT-TGC-ATG-TCT	CTT-GGG-AAG-CAG-AAA-GAT-TAG-ATG	208
mTBP	NM_013684.3	AGC-TCT-GGA-ATT-GTA-CCG-CA	AAT-CAA-CGC-AGT-TGT-CCG-TG	168
mTbx3	NM_198052 / NM_011535	TGA-GGT-GCT-CTG-GAC-TGG-AT	TTG-GAC-ATC-CAT-TGC-TCC-CC	187
mTcf25	NM_001037877	TGG-AAC-TTG-CAG-GCT-GGA-TT	GTT-CCG-ATG-AGC-CTC-CCA-TT	250
mTsc22d1	NM_009366	ACG-CTT-CCG-TGA-GAC-TTG-AC	CTG-CTC-CAG-CTG-GGA-GTT-TT	176
mTwist2	NM_007855	GGC-GCT-ACA-GCA-AGA-AAT-CG	GAT-GTG-CAG-GTG-GGT-CCT-G	441
mYbx3	NM_011733	CCC-CGT-AAT-GCT-GGT-GAG-AT	GGA-ACG-GCG-CCT-ATA-GTT-GT	237

Appendix Table 1: List of specific primers used for RT-qPCR.

Each primer pair was designed with primer-BLAST to span an exon-exon junction.

	# Nucleotide sequences from TF genes	# RefSeq Category of TF nucleotide sequences	# TF genes
<u>Detected</u> (Raw RNA-seq Data)	2,166	2,113 mRNAs (NM) 53 non-coding RNAs (NR)	1,272
<u>Expressed</u> [Mean (RPKM Md-OB) ≥ 1 or Mean (RPKM Tb-OB) ≥ 1]	977	958 mRNAs (NM) 19 non-coding RNAs (NR)	709
Commonly expressed in Md-OBs and Tb-OBs (All expressed transcripts from a given TF gene are commonly expressed)	817	804 mRNAs (NM) 13 non-coding RNAs (NR)	608
Site-specifically increased in Md-OBs [$p \leq 0.1$ AND $ FC \geq 1.4$] (All expressed transcripts from a given TF gene are increased)	56	56 mRNAs (NM)	54
Site-specifically increased in Tb-OBs [$p \leq 0.1$ AND $ FC \leq -1.4$] (All expressed transcripts from a given TF gene are increased)	19	19 mRNAs (NM)	18
Site-specificity depends on transcript variant (Some expressed transcripts from a given TF gene are increased)	85	79 mRNAs (NM) 6 non-coding RNAs (NR)	29

Appendix Table 2: RNA-seq data summary of the bone site-specific TF profile of functional OB isolated from mouse mandible and tibia using RNA-seq. Filtering steps included: q-values ≤ 0.1 (for multi-test adjustment), $FC \geq 1.4$ or ≤ -1.4 , and mean RPKM ≥ 1 for expressed genes.

TF: transcription factor, OB: osteoblast, Md: Mandible, Tb: Tibia, RPKM: Reads Per Kilobase of transcript per Million mapped reads, NM: mRNA RefSeq accession prefixes for validated protein-coding transcripts, NR: RNA RefSeq accession prefixes for non-protein-coding transcripts

Gene	Gene ID	Gene	Gene ID	Gene	Gene ID	Gene	Gene ID	Gene	Gene ID
ADNP	23394	FOXJ3	22887	IRX1	79192	NRL	4901	SMAD9	4093
AKNA	80709	FOXN2	3344	IRX3	79191	OLIG2	10215	SNAI1	6615
ALX1	8092	FOXN3	1112	IRX5	10265	OLIG3	167826	SNAI2	6591
ALX3	257	FOXN4	121643	JUN	3725	OSR1	130497	SOX1	6656
ALX4	60529	FOXO1	2308	KCNIP3	30818	OTP	23440	SOX10	6663
ARX	170302	FOXO2	2309	KLF15	28999	OTX2	5015	SOX11	6664
ATF4	468	FOXP2	93986	ISL1	3670	OVOL2	58495	SOX17	64321
ASCL1	429	FOXS1	2307	ISL2	64843	PAX2	5076	SOX2	6657
ATF5	22809	GATA2	2624	LBX1	10660	PAX3	5077	SOX3	6658
ATOX7	220202	GATA3	2625	LBX2	85474	PAX6	5080	SOX4	6659
BARX1	56033	GATA4	2626	LEF1	51176	PAX7	5081	SOX5	6660
BCL11A	53335	GATA6	2627	LHX1	3975	PAX9	5083	SOX6	55553
BCL11B	64919	GBX1	2636	LHX2	9355	PBX1	5087	SOX8	30812
CAMTA2	23125	GBX2	2637	LHX3	8022	PBX4	80714	SOX9	6662
CREB1	1385	GLI1	2735	LHX4	89884	PHOX2A	401	SP1	6667
DBX1	120237	GLI2	2736	LHX5	64211	PHOX2B	8929	SP3	6670
DEAF1	10522	GLI3	2737	LHX9	56956	PITX2	5308	SP4	6671
DLX1	1745	GRHL2	79977	LYAR	55646	PITX3	5309	SRF	6722
DLX2	1746	GRHL3	57822	MAF1	84232	PKNOX1	5316	STAT1	6772
DLX3	1747	GSC	145258	MAFB	9935	POU3F1	5453	STAT3	6774
DLX5	1749	GSC2	2928	MAX	4149	POU3F2	5454	TAL1	6886
DLX6	1750	GSX1	219409	MECOM	2122	POU3F3	5455	TBX1	6899
DMRT3	58524	GSX2	170825	MEF2A	4205	POU4F1	5457	TBX18	9096
DMRTA2	63950	HAND1	9421	MEF2C	4208	POU4F2	5458	TBX20	57057
EBF1	1879	HAND2	9464	MEF2D	4209	POU4F3	5459	TBX21	30009
EGR2	1959	HES1	3280	MEIS2	4212	POU5F1	5460	TCF12	6938
ELK1	2002	HES3	390992	MITF	4286	PRDM1	639	TCF3	6929
ELK4	2005	HES5	388585	MNT	4335	PRDM2	7799	TCF4	6925
EMX1	2016	HEY2	23493	MNX1	3110	PROX1	5629	TCF7	6932
EMX2	2018	HIF1A	3091	MSX1	4487	PRRX1	5396	TCF7L2	6934
EN1	2019	HLX	3142	MSX2	4488	PRRX2	51450	TEAD3	7005
ERG	2078	HMX1	3166	MYB	4602	PTF1A	256297	TFAP2A	7020
ETS1	2113	HMX3	340784	MYBL2	4605	PURA	5813	TFAP2B	7021
ETS2	2114	HOXA1	3198	MYC	4609	RARA	5914	TGIF1	7050
ETV1	2115	HOXA2	3199	MYCN	4613	RARB	5915	THRB	7068
ETV4	2118	HOXA3	3200	NCOA1	8648	RARG	5916	TLX1	3195
ETV5	2119	HOXA4	3201	NEUROD1	4760	RELA	5970	TLX2	3196
EVX1	2128	HOXA5	3202	NEUROD4	58158	RERE	473	TLX3	30012
FERD3L	222894	HOXA7	3204	NEUROG1	4762	RORA	6095	TSC22D1	8848
FEZF2	55079	HOXB1	3211	NEUROG2	63973	RORB	6096	TSHZ3	57616
FLI1	2313	HOXB2	3212	NFATC3	4775	RUNX1	861	TWIST1	7291
FOSL1	8061	HOXB3	3213	NFATC4	4776	RUNX2	860	USF1	7391
FOXA1	3169	HOXB4	3214	NFE2L1	4779	RXRA	6256	USF2	7392
FOXB1	27023	HOXB5	3215	NFIB	4781	RXRB	6257	VAX1	11023
FOXC1	2296	HOXB6	3216	NFYB	4801	RXRG	6258	VSX1	30813
FOXC2	2303	HOXB7	3217	NKX2-1	7080	SALL2	6297	ZEB1	6935
FOXD1	2297	HOXB8	3218	NKX2-2	4821	SALL4	57167	ZEB2	9839
FOXD2	2306	HOXB9	3219	NKX2-5	1482	SFPQ	6421	ZFP704	619279
FOXD3	27022	HOXC10	3226	NKX6-1	4825	SIX1	6495	ZHX2	22882
FOXE1	2304	HOXC4	3221	NKX6-2	84504	SIX3	6496	ZIC1	7545
FOXF1	2294	HOXC5	3222	NR1D1	9572	SMAD2	4087	ZIC2	7546
FOXF2	2295	HOXD1	3231	NR2E1	7101	SMAD3	4088	ZIC3	7547
FOXG1	2290	HOXD10	3236	NR2F1	7025	SMAD4	4089	ZIC4	84107
FOXI1	2299	HOXD3	3232	NR2F2	7026	SMAD5	4090	ZIC5	85416
FOXI2	399823	HOXD4	3233	NR5A1	2516	SMAD6	4091		
FOXI3	344167	INSM1	3642	NR5A2	2494	SMAD7	4092		

Appendix Table 3: List of TF genes expressed in neural crest

List of TF genes generated from published studies (Sauka-Spengler and Bronner-Fraser 2008; Nelms and Labosky 2010; Simões-Costa et al. 2014; Simões-Costa and Bronner 2015; Wang et al. 2019; Odashima et al. 2020) and bioinformatic DAVID database. This list includes all genes reported in all species. TF genes

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

highlighted in red are related to cranial neural crest (CNC). Corresponding human gene symbols and genes ID were established using ToppGene Suite (Chen et al. 2009).

For Peer Review

FT / Gene / protein	Study PMID	Detection	Species	Stage / Tissue (cells)
Alx3	J.M. Mitchell, 2021 PMID: 33741714	ISH (RNA)	Zebrafish	24 hpf embryos: NC cells of frontonasal bourgeon and anterior neurocranium skeletal progenitors
Dlx3	K. Sumiyama, 2003 PMID: 12642674	TG Model	Mouse	E10.5: Full-length construct (Dlx3-lacZ-79kb) transgenic lacZ-stained embryo: 1 st and 2 nd branchial arches and AER Truncated Dlx3-lacZ-19-kb transgenic lacZ: only at AER of both forelimbs and hindlimbs
	M.Q. Hassan, 2004 PMID: 15456894	ISH (RNA)	Mouse	E15: developing limbs (perichondrium and mature chondrocytes) and vertebrae E16: mature limbs (periosteal cells, cells surrounding primary spongiosa trabeculae in metaphysis and in endosteal OBs of diaphyseal cortical bone)
	S. Ghoul-Mazgar, 2005 PMID: 16172034	IHC (protein) ISH (RNA)	Mouse human	Mouse: E 15.5: Protein and RNA: Sites of mesenchymal condensation and in bone trabeculae E16.5: Protein: endochondral bone ossification centers, growth plate cartilage, proliferative zone, pre-hypertrophic chondrocytes P1: Transcript detected in OBs lining alveolar bone trabeculae. Protein: OBs and newly differentiated osteocytes Human: 9 weeks embryo: Palatal membranous ossification (OBs progenitor, OBs and newly differentiated osteocytes)
	H. Li, 2008 PMID: 18280462	ISH (RNA)	Chick Mouse	Chick: Stage 41 (E15): periosteum covering the newly formed bone (osteogenic cells), calvaria bone formation (osteogenic cells lining bone marrow spaces) Dlx3 mRNA was found in total RNA of chick long bone, calvaria as well as in RNA from chick calvarial OB cultures at days 13 and 20 Dlx3 signal was also present in condensing pre-osteogenic mesenchyme and periosteal osteogenic cells of mandible sections from 8- to 10-day-old chick embryos Mouse: P18: sagittal section of mandible (OBs, Pre-osteocytes and osteocytes)
	O. Duverger, 2013 PMID: 22886599	TG Model	Mouse	E16.5: Developing bones in craniofacial region (frontal and parietal bones) and especially in the mandible P3: Frontal and parietal bones and calvaria bone
	J. Isaac, 2014 PMID: 24948010	IHC (protein) TG Model	Mouse	LacZ TG mouse E14.5: Forelimbs and radius in bone collar E16.5: Humeri (cartilage matrix) P1.5 Calvaria, ribs, manus and tibia Wild type (ISH) E16.5: Growth plate (hypertrophic chondrocytes), perichondrium (osteoblastic cells), primary spongiosa and cortical bone 5 weeks: hypertrophic chondrocytes in the metaphysis, trabecular bone (OBs), cortical bone (osteocytes)
HoxA10	M.Q. Hassan, 2007 PMID: 17325044	IHC (Protein)	Mouse	P1: Periosteum, hypertrophic chondrocytes zone of growth plate, and OBs on all bone surfaces. Weak expressions in flattened cells (prehypertrophic chondrocyte)
Hoxc10	S.L. Hostikka, 2009 PMID: 19623272	TG model	Mouse	E13.5: Proximal hindlimb, perichondrium surrounding cartilage condensations of femur and tibia
Msx1	K.M. Catron, 1996 PMID: 8861098	ISH (RNA)	Mouse	E11.5: Limb buds, dorsal midline, and craniofacial primordia (maxillary and mandibular primordia)
	F. Lezot, 2000 PMID: 10750557	TG Model	Mouse	E10.5: Maxillary and mandibles areas
	S.M. Orestes-Cardoso, 2001 PMID: 11357189	TG Model	Mouse	Early development : Mandible and maxillary: Progenitor NC derived cells of alveolar bone and basal bone (entire dentition, mandible and maxillary) birth-M1 : Alveolar bone of mandibular incisor and mandibular basal bone P14-P21: Skull : Periosteal surface of developing skull bone, appendicular skeleton (sites of endochondral bone formation (limbs, vertebrae, sternum and ribs) M3 : Distal segments of the skull (nasal bone and cartilage) M6-M15 : Restricted areas of periosteal surface of maxillary and mandibular bone
	T. Yamashiro, 2003 PMID: 12598544	ISH (RNA)	Mouse	P14: Alveolar bone
	A. Berdal, 2009 PMID: 18769023	TG Model	Mouse	PW 2-3: mandible (forming basal bone, OBs and osteoclasts)
	F. Wehrhan, 2011 PMID: 21366807	IHC (protein)	Human	Jaw OBs
	A. Nassif, 2014 PMID: 24929242	ISH (RNA)	Mouse	P1: Mandible
B. Xuan, 2017 PMID: 28153754	IHC (protein)	Rat	PW 8: Periodontal membrane, endosteal cells of maxilla and mandible, endosteal cells of the parietal bone and ilium	
Msx2	M. Mina, 1995 PMID: 7734736	ISH (RNA)	Chick	Stage 20-25 (E.H.70-E4.5): Medio-labial regions of the mandibular arch
	K.M. Catron, 1996 PMID: 8861098	ISH (RNA)	Mouse	E11.5: Msx2 is expressed asymmetrically at higher levels in the anterior margin of the limb bud. E11.5: Msx2 is expressed in the anterior primordium of 1 st branchial arch. E12.5 to 16.5: Developing bones (osteoid and Pre-OB regions of the sutural margins and in differentiating cranial cartilages) E13: Expression of Msx-2 is relatively higher in the interdigital zones during the early phases of digit condensation
	X. Nie, 2006 PMID: 17090159	ISH (RNA)	Mouse	E13: Mesenchyme underlying the cranial base E14: Strongly in mesenchyme underlying the cranial base and in sphenoid.
	K. Kishimoto, 2006 PMID: 16283360	IHC (protein)	Mouse	PW 8: Proximal tibia chondrocytes (tangential layer of the articular cartilage)
	A. Berdal, 2009 PMID: 18769023	TG Model	Mouse	PW 2-3: Growth plate cartilage (ribs, limbs, vertebrae), mandible (bone areas adjoining growing teeth)
Pax9	H. Peters, 1995 PMID: 7647370	ISH (RNA)	Chick	E7: Intervertebral disks located between the 2nd, 3rd and 4th vertebrae and in appendicular skeleton (mesenchymal tissue of the median metatarsus proximal to metatarsal I and metatarsal II)
	R. Kriangkrai, 2006 PMID: 16416307	ISH (RNA)	Rat	E12 and 13: Underlying ectomesenchyme of the prospective site of tooth formation and medial nasal process (future maxillary)
	E.E. LeClair, 1999 PMID: 10030590	ISH (RNA)	Chick	Stage 20-22: Pax-9 signal in both the axial sclerotome and the pharyngeal pouches but not in the limb buds Stage 23 (E3.5-4): Anterior margin of the leg buds (between leg base and AER) Stage 28-29: Leg (junction of the footplate and the zeugopod, immediately anterior to the developing ankle and digit I) Stage 29-30: expression of Pax-9at early limb bud stages continuing through stage 29/30
	A. Neubüser, 1995 PMID: 7649395	ISH (RNA)	Mouse	E10.5: Nasal process E11: Anterior proximal of the hand and foot plates, nasal mesenchyme, mandibular and maxillary components of 1st branchial arch E11.5: Limb buds E14: hindlimb, between tarsal and metatarsal E12.5: anterior proximal corner of paddle-shaped hand and foot plates E14.5: the condensations of mesenchymal tissues in the intervertebral discs developing hand
	L. Sonnesen, 2008 PMID: 18594453	IHC (protein)	Mouse	E13.5: between corporal bodies in body axis and in craniofacial region but not around the basilar part of the occipital bone E13.5: expression of PAX9 at the cervical part of the vertebral column at the intervertebral area
Sox9	H. Dupuis, 2019 PMID: 31635173	IHC (protein)	Rat	P16: Tibia growth plate chondrocytes
	X. He, 2017 PMID: 28627474	TG Model	Mouse	Adult: Femora (articular and growth plate cartilage, primary spongiosa, periosteum and endosteum)
	S. Shibata, 2006 PMID: 16441561	ISH (RNA)	Mouse	E14: Condylar anlage E15: Newly formed cartilage, slightly reduced in the bone collar E16: Condylar cartilage in coronal plane: polymorphic cell zone to the flattened cell zone but reduced in an entire hypertrophic cell zone and in bone collar
	S. Shibata, 2008 PMID: 18624832	ISH (RNA)	Mouse	E11: Mandibular anlage E11.5: Meckel's anlagen, but not in the mandibular anlage E12 to 13: Meckel's cartilage, but not in the mandibular anlage. E14: Condylar anlage, but not in the anterior position of the mandibular anlage

1
2
3
4
5
6
7
8 **Appendix Table 4: Non exhaustive compilation of *in situ* localization in bone tissues of selected**
9 **TFs previously reported in the literature using immunohistochemistry, immunofluorescence, *in situ***
10 **hybridization and transgenic animals with reporter genes.**

11
12 *AER: Apical Ectodermal Ridge, E: Embryonic Day, EH: Embryonic Hour, HPF: Hours Post Fertilization, ISH:*
13 *in situ hybridization, M: Month, NC: neural crest, P: Post Natal Day, PW: Post Natal Weeks, OB: osteoblast,*
14 *TG: Transgenic,*
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Gene	NCP association	Associated disease	MIM	Jaw bones	Palate	Tooth	Face & cranium	Long bones	Shoulder Girdle - Gluteal region	Digits	Axial skeleton	Short stature	Hematopoiesis/ Immune system	AER/ZPA/IDM	
ALX1	X	Frontonasal dysplasia 3	613456	X	X	X	X				X				
ALX3	X	Frontonasal dysplasia 1	136760		X		X								
ALX4	X	{Craniosynostosis 5, susceptibility to}	615529				X							X	
		Frontonasal dysplasia 2	613451				X								
		Parietal foramina (PFM) 2	609597		X	X	X			X					
DLX3	X	Amelogenesis imperfecta, type IV	104510			X								X	
		Trichodontoosseous syndrome (TDO)	190320	X		X	X	X							
DLX4		Orofacial cleft 15	616788		X	X	X								
DLX5		SHFM 1 with sensorineural hearing loss	220600				X	X		X	X			X	
		Split-hand/foot malformation 1	183600	X	X	X	X			X					
GLI3	X	Greig cephalopolysyndactyly syndrome	175700				X		X	X		X		X	
		Pallister-Hall syndrome	146510	X	X		X		X	X		X			
		Polydactyly, postaxial, types A1 and B	174200							X					
		Polydactyly, preaxial, type IV	174700							X					
GSC	X	SAMS	602471	X	X	X	X	X	X	X	X	X			
IRF1		Gastric cancer, somatic	613659												
		Myelodysplastic syndrome, preleukemic	-										X		
		Myelogenous leukemia, acute	-											X	
		Nonsmall cell lung cancer, somatic	211980												
MSX1		Ectodermal dysplasia 3, Witkop type	189500			X								X	
		Orofacial cleft 5	608874		X										
		Tooth agenesis, selective, 1, with or without orofacial cleft	106600			X									
MSX2	X	Craniosynostosis 2	604757				X			X				X	
		Parietal foramina 1	168500		X	X	X				X				
		Parietal foramina with cleidocranial dysplasia	168550				X		X						
MYCN		Feingold syndrome 1	164280	X		X			X	X	X		X		
PAX3	X	Craniofacial-deafness-hand syndrome	122880	X	X		X			X				X	
		Rhabdomyosarcoma 2, alveolar (RMS)	268220												
		Waardenburg syndrome, type 1	193500				X				X				
		Waardenburg syndrome, type 3	148820				X			X					
PAX9		Tooth agenesis, selective (STHA), type 3	604625			X									
PBX1		CAKUTHEd	617641				X			X	X		X		
RARB		Microphthalmia, syndromic 12	615524	X	X		X								
SALL4		Duane-radial ray syndrome (Okihiro syndrome)	607323					X		X	X			X	
		IVIC syndrome	147750				X	X		X					
SOX5		Lamb-Shaffer syndrome	616803	X		X	X			X	X		X		
SOX9	X	Campomelic dysplasia	114290	X	X		X	X	X	X	X		X		
SOX11		Coffin-Siris syndrome 9	615866				X			X		X	X		
SP6		Hypoplastic Amelogenesis imperfecta	-			X								X	
TBX3		Ulnar-mammary syndrome	181450			X	X	X	X	X	X			X	
TCF12		Craniosynostosis 3	615314	X			X			X					
TWIST2		Ablepharon-macrostomia syndrome	200110	X		X	X			X				X	
		Barber-Say syndrome	209885	X	X	X	X			X					
		Focal facial dermal dysplasia 3, Setleis type	227260			X	X								
ZBTB24		ICF2	614069	X	X		X						X		
IRF7		Immunodeficiency 39 (IMD39)	616345										X		
KLF1		[Hereditary persistence of fetal hemoglobin]	613566											X	
		Blood group--Lutheran inhibitor	111150											X	
		Dyserythropoietic anemia, congenital, type IV	613673											X	
NFIA		Brain malformations with or without urinary tract defects	613735	X	X		X		X	X					
TBX15		Cousin syndrome	260660				X	X	X	X				X	

Appendix Table 5: Compilation of human clinical synopses associated with Md-set and TB-set TF gene mutations with Mendelian association (OMIM and Orphanet)

AER: apical ectodermal ridge, IDM: interdigital mesenchyme, NCP: neurocristopathies (Vega-Lopez et al. 2018), ZPA: zone of polarizing activity.

Abbreviations of human genetic disorders: CAKUTHEd: Congenital anomalies of kidney and urinary tract syndrome with or without hearing loss, abnormal ears, or developmental delay, ICF2: Immunodeficiency-centromeric instability-facial anomalies syndrome 2, , IVIC: Instituto Venezolano de Investigaciones Cientificas, PFM: Parietal foramina 1, , RMS: Rhabdomyosarcoma 2, alveolar, SAMS: Short stature, auditory canal atresia, mandibular hypoplasia, skeletal abnormalities, SHFM: split hand/foot malformation), TDO: trichodontoosseous syndrome

Appendix References

Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, et al. 2017. QuPath: Open source software for digital pathology image analysis. *Sci Rep.* 7(1):16878.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 25(2):288–289.

Chawla K, Tripathi S, Thommesen L, Lægreid A, Kuiper M. 2013. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics.* 29(19):2519–2520.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 37(Web Server issue):W305–311.

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37(1):1–13.

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.

Hulsen T, de Vlieg J, Alkema W. 2008. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics.* 9:488.

Isaac J, Erthal J, Gordon J, Duverger O, Sun HW, Lichtler AC, Stein GS, Lian JB, Morasso MI. 2014. DLX3 regulates bone mass by targeting genes supporting osteoblast differentiation and mineral homeostasis in vivo. *Cell Death and Differentiation.* 21(9):1365–76.

Isaac J, Nassif A, Asselin A, Taihi I, Fohrer-Ting H, Klein C, Gogly B, Berdal A, Robert B, Fournier BP. 2018. Involvement of neural crest and paraxial mesoderm in oral mucosal development and healing. *Biomaterials.* 172:41–53.

Jacques J, Hotton D, De la Dure-Molla M, Petit S, Asselin A, Kulkarni AB, Gibson CW, Brookes SJ, Berdal A, Isaac J. 2014. Tracking endogenous amelogenin and ameloblastin in vivo. *PLoS One.* 9(6):e99626.

Kalajzic I, Kalajzic Z, Kaliterna M, Gronowicz G, Clark SH, Lichtler AC, Rowe D. 2002. Use of type I collagen green fluorescent protein transgenes to identify subpopulations of cells at different stages of the osteoblast lineage. *J Bone Miner Res.* 17(1):15–25.

Lang D, Lu MM, Huang L, Engleka KA, Zhang M, Chu EY, Lipner S, Skoultchi A, Millar SE, Epstein JA. 2005. Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature.* 433(7028):884–7.

1
2 Nelms BL, Labosky PA. 2010. *Transcriptional Control of Neural Crest Development*. San Rafael
3 (CA): Morgan & Claypool Life Sciences (Developmental Biology). [accessed 2021 Jun 30].
4 <http://www.ncbi.nlm.nih.gov/books/NBK53145/>.

5
6 Odashima A, Onodera S, Saito A, Ogihara Y, Ichinohe T, Azuma T. 2020. Stage-dependent
7 differential gene expression profiles of cranial neural crest-like cells derived from mouse-
8 induced pluripotent stem cells. *Med Mol Morphol*. 53(1):28–41.

9
10
11 Sauka-Spengler T, Bronner-Fraser M. 2008. A gene regulatory network orchestrates neural crest
12 formation. *Nat Rev Mol Cell Biol*. 9(7):557–568.

13
14 Schmeier S, Alam T, Essack M, Bajic VB. 2017. TcoF-DB v2: update of the database of human
15 and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res*.
16 45(D1):D145–D150.

17
18
19 Simões-Costa M, Bronner ME. 2015. Establishing neural crest identity: a gene regulatory recipe.
20 *Development*. 142(2):242–257.

21
22 Simões-Costa M, Tan-Cabugao J, Antoshechkin I, Sauka-Spengler T, Bronner ME. 2014.
23 Transcriptome analysis reveals novel players in the cranial neural crest gene regulatory network.
24 *Genome Res*. 24(2):281–290.

25
26
27 The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic*
28 *Acids Research*. 49(D1):D480–D489.

29
30 Vega-Lopez GA, Cerrizuela S, Tribulo C, Aybar MJ. 2018. Neurocristopathies: New insights
31 150 years after the neural crest discovery. *Dev Biol*. 444 Suppl 1:S110–S143.

32
33 Wang H, Holland PWH, Takahashi T. 2019. Gene profiling of head mesoderm in early zebrafish
34 development: insights into the evolution of cranial mesoderm. *Evodevo*. 10:14.

35
36
37 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK.
38 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.
39 *Nat Commun*. 10(1):1523.



The ARRIVE guidelines 2.0: author checklist

The ARRIVE Essential 10

These items are the basic minimum to include in a manuscript. Without this information, readers and reviewers cannot assess the reliability of the findings.

Item	Recommendation	Section/line number, or reason for not reporting	
Study design	1 For each experiment, provide brief details of study design including: <ol style="list-style-type: none"> The groups being compared, including control groups. If no control group has been used, the rationale should be stated. The experimental unit (e.g. a single animal, litter, or cage of animals). 	M&M para 1, Appendix M&M M&M para 1,2, Ap.M&M para 1,8	
	Sample size	2 <ol style="list-style-type: none"> Specify the exact number of experimental units allocated to each group, and the total number in each experiment. Also indicate the total number of animals used. Explain how the sample size was decided. Provide details of any <i>a priori</i> sample size calculation, if done. 	M&M para 1, Appendix M&M M&M para 1,2, Ap.M&M para 1,8
Inclusion and exclusion criteria	3 <ol style="list-style-type: none"> Describe any criteria used for including and excluding animals (or experimental units) during the experiment, and data points during the analysis. Specify if these criteria were established <i>a priori</i>. If no criteria were set, state this explicitly. For each experimental group, report any animals, experimental units or data points not included in the analysis and explain why. If there were no exclusions, state so. For each analysis, report the exact value of <i>n</i> in each experimental group. 	M&M para 1 Ap.M&M para 8 NA: no experiment on living mice M&M para 2, Ap.M&M para 8	
	Randomisation	4 <ol style="list-style-type: none"> State whether randomisation was used to allocate experimental units to control and treatment groups. If done, provide the method used to generate the randomisation sequence. Describe the strategy used to minimise potential confounders such as the order of treatments and measurements, or animal/cage location. If confounders were not controlled, state this explicitly. 	NA: no experiment on living mice NA: no experiment on living mice
		Blinding	5 Describe who was aware of the group allocation at the different stages of the experiment (during the allocation, the conduct of the experiment, the outcome assessment, and the data analysis).
Outcome measures	6 <ol style="list-style-type: none"> Clearly define all outcome measures assessed (e.g. cell death, molecular markers, or behavioural changes). For hypothesis-testing studies, specify the primary outcome measure, i.e. the outcome measure that was used to determine the sample size. 	no exp. on living mice no exp. on living mice	
	Statistical methods	7 <ol style="list-style-type: none"> Provide details of the statistical methods used for each analysis, including software used. Describe any methods used to assess whether the data met the assumptions of the statistical approach, and what was done if the assumptions were not met. 	Ap.M&M para 8 Ap.M&M para 8
Experimental animals		8 <ol style="list-style-type: none"> Provide species-appropriate details of the animals used, including species, strain and substrain, sex, age or developmental stage, and, if relevant, weight. Provide further relevant information on the provenance of animals, health/immune status, genetic modification status, genotype, and any previous procedures. 	M&M para 1 Ap.M&M para 1 M&M para 1 Ap.M&M para 1
	Experimental procedures	9 For each experimental group, including controls, describe the procedures in enough detail to allow others to replicate them, including: <ol style="list-style-type: none"> What was done, how it was done and what was used. When and how often. Where (including detail of any acclimatisation periods). Why (provide rationale for procedures). 	M&M para 1,2 Ap.M&M para 1-8 no exp.on living mice NA: no experiment on living mice M&M, results section
Results		10 For each experiment conducted, including independent replications, report: <ol style="list-style-type: none"> Summary/descriptive statistics for each experimental group, with a measure of variability where applicable (e.g. mean and SD, or median and range). If applicable, the effect size with a confidence interval. 	Ap.M&M para 8 fig. legends Ap.M&M para 8 fig. legends

The Recommended Set

These items complement the Essential 10 and add important context to the study. Reporting the items in both sets represents best practice.

Item	Recommendation	Section/line number, or reason for not reporting
Abstract	11 Provide an accurate summary of the research objectives, animal species, strain and sex, key methods, principal findings, and study conclusions.	Abstract
Background	12 a. Include sufficient scientific background to understand the rationale and context for the study, and explain the experimental approach. b. Explain how the animal species and model used address the scientific objectives and, where appropriate, the relevance to human biology.	Throughout M&M and Ap. M&M, results 3
Objectives	13 Clearly describe the research question, research objectives and, where appropriate, specific hypotheses being tested.	Intro, Results
Ethical statement	14 Provide the name of the ethical review committee or equivalent that has approved the use of animals in this study, and any relevant licence or protocol numbers (if applicable). If ethical approval was not sought or granted, provide a justification.	M&M-para 1
Housing and husbandry	15 Provide details of housing and husbandry conditions, including any environmental enrichment.	M&M-para 1
Animal care and monitoring	16 a. Describe any interventions or steps taken in the experimental protocols to reduce pain, suffering and distress. b. Report any expected or unexpected adverse events. c. Describe the humane endpoints established for the study, the signs that were monitored and the frequency of monitoring. If the study did not have humane endpoints, state this.	no exp. on living mice no exp. on living mice no exp. on living mice
Interpretation/ scientific implications	17 a. Interpret the results, taking into account the study objectives and hypotheses, current theory and other relevant studies in the literature. b. Comment on the study limitations including potential sources of bias, limitations of the animal model, and imprecision associated with the results.	results - part 1,2,3 Intro, Discussion
Generalisability/ translation	18 Comment on whether, and how, the findings of this study are likely to generalise to other species or experimental conditions, including any relevance to human biology (where appropriate).	results -part 3 Discussion
Protocol registration	19 Provide a statement indicating whether a protocol (including the research question, key design features, and analysis plan) was prepared before the study, and if and where this protocol was registered.	M&M-para 1 Ap.M&M para 1
Data access	20 Provide a statement describing if and where study data are available.	fig, tables
Declaration of interests	21 a. Declare any potential conflicts of interest, including financial and non-financial. If none exist, this should be stated. b. List all funding sources (including grant identifier) and the role of the funder(s) in the design, analysis and reporting of the study.	Acknowledgements Acknowledgements