Unraveling the Molecular Landscape of Congenital Pseudoarthrosis of the Tibia: Insights from a Comprehensive Analysis of 162 Probands

Guanghui Zhu[1,2,3#], Nan Li[4#], Yu Zheng[1,3,5#], Shunyao Wang[4], Ge Yang[1,2,3], Yaoxi Liu [1,2,3], Zongren Xu[4], Hui Huang[4*], Huanhuan Peng[4*], Haibo Mei[1,2,3*]

[1]Department of Pediatric Orthopedics, Hunan Children's Hospital, Hunan 410007, PR China

[2]Hunan Provincial Key Laboratory of Pediatric Orthopedics, Hunan 410007, PR China

[3]The school of pediatrics, University of South China, Hunan 410007, PR China

[4]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

[5] Hunan Children's Research Institute (HCRI), Hunan Children's Hospital, Hunan 410007, PR China

[#] Guanghui Zhu, Nan Li and Yu Zheng contributed equally to this article.

[*] Corresponding authors:

Hui Huang, BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. E-mail address: huanghui@genomics.cn

Huanhuan Peng, BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. E-mail address: penghh@genomics.cn;

Haibo Mei, Hunan Key Laboratory of Pediatric Orthopedics; Department of Pediatric Orthopedics, Hunan Children's Hospital; The school of pediatrics, University of South China Hunan 410000, PR China E-mail address: meihaibo@sohu.com;

Fundings: Opening fundings of Hunan Provincial Key Laboratory of Pediatric Orthopedics

Abstract

Congenital pseudarthrosis of the tibia (CPT, HP:0009736), commonly known as bowing of the tibia, is a rare congenital tibia malformation characterized by spontaneous tibial fractures and the difficulty of reunion after tibial fractures during early childhood, with a very low prevalence between 1/250,000~1/140,000. While 80%–84% of CPT cases present with neurofibromatosis type 1, caused by the mutations in *NF1*, the underlying cause of CPT is still unclear. Considering the congenital nature and the low prevalence of CPT, we hypothesized that the rare genomic mutations may contribute to CPT. In this study, we conducted whole exome sequencing on 159 patients with CPT and full-length transcriptome sequencing on an additional 3 patients with CPT. The data analysis showed there were 179 significantly up-regulated genes which were enriched in 40 biological processes among which 21 biological processes hold their loss of function (LoF) excesses between 159 cases against 208 controls from 1000 Genomes Project. From those 21 biological processes with LoF excesses, there were 259 LoF-carried genes among which 40 genes with 56 LoF variations in 63 patients were enriched in osteoclast differentiation pathway (hsa04380) with its 3 directly regulated pathways including

MAPK signaling pathway (hsa04010), calcium signaling pathway (hsa04020) and PI3K-Akt signaling pathway (hsa04151), as well as fluid shear stress and atherosclerosis pathway (hsa05418) while 12 patients carried 9 LoF variations in the *NF1* gene. The rare LoF variations in these pathways accounted for ~39.6% of this CPT cohort. These findings shed light on the novel genetic mutations and molecular pathways involved in CPT, providing a new framework for understanding how the genetic variations regulate the biological processes in the pathology of CPT and indicating potential next directions to further elucidate the pathogenesis of CPT.

Introduction

Congenital pseudarthrosis of the tibia (CPT) is a rare orthopedic disease with an incidence between 1 in 140,000 and 1 in 250,000 live births [1]. It is characterized by the pseudarthrosis or pathological fractures of the anterolateral part of the tibia in early life, resulting in bowing, narrowing of the medullary canal, or the presence of a cyst. This condition poses a significant surgical challenge due to recurrent fractures and the inability to achieve bone union [2]. Congenital anterolateral bowing of the tibia is generally considered a precursor of CPT and is commonly

associated with neurofibromatosis type 1(NF1), a common autosomal dominant genetic disorder [3]. Previous research has reported that 80%–84% of CPT cases present with NF1(NF1-CPT) in the epidemiology and the genomic mutations in the *NF1* gene have been linked to CPT in the genomics [4-7]. The double inactivation model has been proposed as a hypothesis to explain these associations [8-11] . However, other researchers have suggested that the molecular pathogenesis of NF1-CPT may not be entirely explained by second mutations or loss of heterozygosity of the *NF1* gene [7, 12, 13].

Recent investigations have shed light on the molecular basis of CPT, for example, *NF1* gene has been reported to be involved in bone remodeling and mineralization [14]. Inhibited bone formation and stimulated bone resorption have also been reported in patients with CPT who harbor *NF1* mutations [15]. Furthermore, the involvement of FGFR3 signaling in cartilage-to-bone transformation for bone repair [16] and the contribution of vasoactive intestinal peptide through ERK and NF-κB signal pathways have been implicated in CPT [17]. Differentially expressed proteins (DEPs) found in serum-derived exosomes of CPT patients have been shown to inhibit bone formation and stimulate bone resorption[18] while DEPs from in the tibia

periosteum tissues are mainly involved in cell matrix assembly, cell adhesion, AKT-PI3K signal pathway activation, and vascular agglutination of CPT patients [19]. These conclusions indicate the intricate biological processes and molecular regulation pathways associated with bone homeostasis potentially contribute to CPT [20-23] . However, a comprehensive understanding of the etiology of CPT is still elusive due to the rarity of CPT and limited samples in previous studies.

To address this gap, we employed an approach used to investigate the severe pediatric developmental disorders with a low prevalence and the complex genetic architecture, aiming to identify the low-frequency gene mutations in 159 sporadic CPT patients. This approach have been successful in elucidating the genetic basis of other neurodevelopmental disorders, including epilepsy [24-28], craniosynostosis [29-31], autism [32-40] and congenital heart disease [41-46]. Specially the rare protein-damaging variations (PDV), particularly loss of function (LoF) mutations, have been identified a main contributor to congenital developmental diseases characterized by low prevalence and early onset. In light of this, we hypothesized that CPT can also be explained, at least partially, by the rare LoF variants with incomplete penetrance. To test this hypothesis, we conducted

whole exome sequencing (WES) on a substantial cohort of 159 individuals affected by idiopathic CPT. Additionally, we performed a full-length transcriptome sequencing on 3 CPT patients to validate the reported biological processes and detect the new ones associated with CPT.

In this study, we present our comprehensive genetic analysis, aiming to uncover the complex genetic architecture and molecular mechanism underlying CPT. Our findings have the potential to enhance the understanding of CPT pathogenesis and provide valuable guidance for the development of novel therapeutic approaches targeting the identified genetic factors.

Results

To comprehensively explore the genetic basis and biological processes underlying CPT, a cohort of 159 along with an additional 3 Chinese Han CPT patients, with an average age of 42 months, were enrolled for WES and full-length transcriptome sequencing respectively. In the WES, targeted bases in each sample were sequenced a mean of 136 times by independent reads, with 98.73% reading 20 or more times. Following variation calling, a totally of 286,945 single nucleotide polymorphisms (SNPs) and 27,807 insertions/deletions (InDels) were detected.

For the transcriptome sequencing, an average of 4,816,579 reads with a length of 787 bp were obtained for each sample.

40 bone homeostasis related biological processes were up-regulated in CPT patients

To get insights into the CPT-related pathways, we performed nanopore sequencing on the periosteum tissues from three CPT patients against two healthy individuals. The differential expression analysis of transcriptome identified 179 genes up-regulated and 134 down-regulated genes excluding *NF1* gene ($|logFC| > 2$, $FDR < 0.01$) in patients with CPT (Fig1 a and b). Functional enrichment analysis of those up-regulated genes uncovered 40 significantly enriched biological processes (threshold P=0.01). These enriched biological processes encompassed skeletal system development (logP=-11.96）, tissue remodeling including bone resorption, cell matrix assembly (logP=-10.37) [47, 48], regulation of cell adhesion (logP=-8.44), regulation of MAPK cascade [49] (logP=-5.22) and autophagy in bone metabolism [50, 51], which have been associated with bone formation and bone remodeling [19, 52-54] . Additional enriched biological progress included ossification (e.g., bone mineralization and endochondral ossification) [55], myeloid

leukocyte differentiation (e.g., osteoclast differentiation and osteoclast development) [56, 57], epithelial cell proliferation (e.g., angiogenesis) [58-61], inorganic ion homeostasis [62, 63], regulation of hormone levels [64, 65], negative regulation of response to external stimulus [57, 66], Vitamin D receptor pathway [67-69], cell migration and chemotaxis [70] (Fig.1.c and supplementary table 1, threshold P= 0.05/100=5e-4 by Bonferroni correction). Some of these processes have been reported in the DEPs of CPT patients [18, 19, 53] and also known to play crucial important role in the metabolism of bone homeostasis. The enriched biological processes identified in the transcriptome analysis provide a comprehensive reflection of the etiology of CPT.

21 of 40 biological processes beared the LoF excesses in CPT patients

Considering the rare prevalence and early onset of CPT, we postulate that some or all of the 40 biological processes may bear a significant burden of rare LoF mutations if they contribute to CPT pathogenesis. To test this hypothesis, we analyzed the genetic burden of LoF variations from WES data. After variation calling and filtering, 1,174 confidant LoF single nucleotide variants (SNVs)

in 1,037 genes (an average of 7.38 mutations per case with a mutation rate of 4.34e-9) were selected for the burden test between 159 CPT cases and 208 healthy Chinese Han controls from the 1000 Genomes Project. Among all the 40 significantly enriched and up-regulated biological processes in transcriptome, 21 processes showed LoF excesses under the threshold P <0.01 and OR > 10 (Fig2, Table 1), indicating their potential relevance to CPT. Additionally, the burden test of rare LoF mutations showed no significant excess of LoF (Bonferroni corrected threshold P-value 0.05/1,217=4.1e-5) for each gene except for *NF1* (p= 5.61357e-05) which was reported in the previous studies [4, 7]. This result suggests that the 21 CPT-related candidate biological processes, which exhibit a significant rare LoF excess, cannot be detected by the burden test for single gene. Furthermore, the other 19 biological processes significantly up-regulated in the transcriptome and enriched in the Gene Ontology (GO) database did not show a rare LoF excess, indicating that rare LoF mutations alone cannot entirely explain the pathogenicity and the complex genetic architecture of CPT.

To assess the association between the 21 biological processes and the risk of CPT, the odd ratio (OR) is used to quantify the probability LoF mutations contributing to CPT. We defined the

gene set including 5,173 genes from the 21 biological processes as the transcriptome high expression (THE). The gene set of biological process (GO:0008150) in GO includes 19,462 genes except THE, consisting of 14,254 genes, was referred to as the non-THE (NTHE). Comparing the OR of NTHE (1.73) between CPT cases and controls, a substantial increase to 18.29 in THE was displayed. This substantial rise in odds ratio demonstrates that the rare LoF excesses from THE significantly correlated with an increased risk of CPT risk. Ultimately, the specific genes contributing to CPT were identified as the those overlapping between the LoF-carrying genes and the THE. This resulted in the discovery of 151 CPT-contributing genes.

Osteoclast differentiation with its regulated pathways contributes to CPT pathogenesis

To get insight into the molecular mechanism of these 151 CPT-contributing genes, they were queried in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for the enrichment that highlighted the pathway osteoclast differentiation (hsa04380, logP=-4.83) likely plays a potential role in CPT, along with its 3 regulated pathways including MAPK signaling (hsa04010, logP=-2.99), calcium signaling (hsa04020, logP=-2.95) and PI3K-Akt

signaling (hsa04151, logP=-2.45). it was worth noting that tuberculosis pathway(hsa05152, logP=-4.59), Chemical carcinogenesis - DNA adducts pathway (hsa05204, logP=-4.56), as well as fluid shear stress and atherosclerosis pathway (hsa05418, logP=-4.57), were also identified(Fig3 a). To validate these findings, we examined the LoF excesses within these pathways. The results demonstrated significant LoF excesses in the MAPK signaling pathway (OR = 26, P = 1.45E-9), Tuberculosis pathway (OR = 12, P=1.37e-4), Chemical carcinogenesis-DNA adducts pathway (OR=10, P=0.001), PI3K-Akt signaling pathway (OR = 9, P = 1.48e-5), calcium signaling pathway (OR = 7, P = 2.66e-4), and osteoclast differentiation pathway (OR = 5.5, P = 0.001) in CPT cases (Fig3 b and Table 2).

Consistent with previous findings, the observed involvement in the osteoclast differentiation pathway and its 3 regulated pathways (MAPK signaling, calcium signaling and PI3K-Akt signaling) support the notion that CPT patients exhibit an imbalance in the low reparative potential of bone tissue due to the osteoclast activation prevailing over differentiation of osteoblasts, progenitor and mesenchymal cells [71]. The regulation of osteoclast differentiation and activation could become a

256 promising strategy for preventing bone erosion in patients with
257 CPT [17].

258 *NF1,* a reliable CPT-causing gene reported in previous studies,
259 is believed to involve disruptions in bone remodeling and vascular
260 abnormalities. Osteoclasts in *NF1* display enhanced resorption
261 capacity and aberrant morphology [72]. The disorders of the
262 Ras/MAPK pathway have an overlapping skeletal phenotype, such
263 as scoliosis and osteopenia. The Ras proteins regulate cell
264 proliferation and differentiation. Studies have shown that
265 individuals carrying *NF1* mutations have osteoclast hyperactivity
266 and increased bone resorption in the studies[73, 74]. In our study,
267 *NF1* gene exhibits 9 LoF variations only in 12 patients (7.54%)
268 with a heterozygous genotype. Alongside *AKT2*, *CD14*, *IL1RAP*,
269 *MAP3K4*, *MAP3K11*, *RPS6KA1*, *TGFA* and *RASGRP2, NF1* was
270 enriched in the MAPK signaling pathway (logP=-2.99). This
271 suggests a potential framework for understanding how *NF1*
272 contribute to the etiology of CPT.

273 Furthermore, *LILRB4*, *LILRB2*, *LILRB1* and *LILRA2* carrying 2
274 LoFs respectively are enriched in the osteoclast differentiation
275 pathway along with *AKT2*, *CALCR*, *FCGR3A* and *IFNG*. The complex
276 of *FCGR3A* and *LILRB1/LILRB2/LILRB4/LILRA2,* known as the
277 osteoclast-associated receptor (OSCAR), is a regulator of

osteoclast differentiation and dendritic cell maturation. It has been reported to cause skeleton-related phenotypes, such as abnormal osteoclast differentiation, reduced osteoarthritis manifestation, decreased chondrocyte apoptosis and abnormal bone morphology in knockout mice [75]. In osteoclasts, OSCAR is mediated by the immunoreceptor tyrosine-based activation motif adaptor protein FcRγ and co-stimulates osteoclasts cultured in an extracellular matrix or collagen [76]. *CALCR* is involved in regulating osteoclast-mediated bone resorption and has been reported to be associated with osteoporosis and bone mineral density quantitative trait locus 15 (https://www.genecards.org/). *IFNG* suppresses osteoclastogenesis by suppressing TRAF6-mediated signaling downstream of RANK, the receptor for RANKL [77]. That is, osteoclastogenesis would increase when the expression of *IFNG* is down-regulated due to the LoF variations.

Additionally, *COL6A5* carrying 2 LoF mutations, accompanied by *AKT2*, *JAK3*, *NOS3*, *TGFA*, *TLR2*, *SGK2*, *PKN3*, *TNN*, *KRT16* and *OPRM1*, is enriched in PI3K-Akt signaling pathway (logP = -2.45). This pathway has been reported to be involved in CPT by quantitative proteomics analysis [19]. The knockdown of *AKT2* was reported to suppress osteoclastogenesis through the PI3K/Akt/GSK3 signaling pathway [78]. Furthermore, alongside

HRC, *MST1R*, *NOS3*, *P2RX1*, *TACR2*, *PPIF* and *TRDN*, *PLCZ1* carrying 2 LoF mutations is enriched in the calcium signaling pathway (logP = -2.95). Calcium signaling plays a crucial role in osteocyte mechanotransduction, as the release of calcium serves as a powerful second messenger. Calcium provides essential information for mechanical responses and participates in downstream regulatory processes [66].

Finally, *GSTM4* carrying 2 LoF mutations, along with *AKT2*, *IFNG*, *NOS3*, *PLAT*, *SELE*, *ARHGEF2*, *GSTO1*, *AKR1C4*, *AKR1C3*, *PPIF* and *SLC26A6*, is enriched in fluid shear stress and atherosclerosis (has05418). Previous studies have reported the involvement of osteoclasts and osteoblasts in activating different mechanical transduction pathways such as fluid shear stress and reported changings in their differentiation, formation, and functional mechanism induced by the application of different types of mechanical stress to bone tissue[79-81].

In summary, the osteoclast differentiation pathway, along with its 3 directly regulated pathways, not only exhibited rare LoF excesses but also showed shared transcriptome patterns between CPT cases and controls. These pathways serve as a reliable set of potential candidate pathways contributing to the etiology of CPT. Overall, there were 63 patients carrying 40 CPT-contributing

genes with 56 LoF variations in the osteoclast differentiation pathways and their 3 directly regulated pathways (Table 3).

Hub genes enriched in hedgehog signaling pathway and likely regulate the molecular pathways of CPT

Interestingly, the Protein-protein interaction network analysis in the 151 CPT-contributing genes indicated that there were 8 hub genes (*BOC*, *SMO*, *PTCH2*, *ULK4*, *CORIN*, *USH2A*, *F12* and *MAGI3*) which showed a significant enrichment in hedgehog signaling pathway (hsa04340, logP= -6.5) and 3 genes *TIRAP*, *CD14* and *TLR2* in Lipid and atherosclerosis (hsa05417, logP=-6.5). As the members of Hedgehog signaling pathway, *BOC*, *SMO* and *PTCH2* respectively carried 1 LoF mutation, and was indicated to be regulated by the other 5 hub genes in the PPI network. Hedgehog signaling pathway had been reported play a crucial role in regulating bone formation and homeostasis[82, 83] and the premature SHH signaling resulting from disruption of *GPR161* causes defects in limb and skeletal morphogenesis[84].

Discussion

CPT is a rare bone disorder primarily affecting newborns and characterized by challenging complications such as refracture,

342 tibial non-union, and failed surgery [85]. The underlying
343 pathophysiology involves excessive osteoclast activity caused by
344 the fibrous hamartoma surrounding the bone and impaired
345 osteogenesis and bone morphogenetic protein function, resulting
346 in frequent tibial fractures, bone atrophy, and compromised bone
347 remodeling [86]. Neurofibromatosis type 1, with the prevalence
348 of 1/3,000 [87], is linked to CPT in many cases and contribute to
349 anterolateral bowing of the tibia, leading to pathologic
350 fracture. Previous study has claimed the gene *NF1* encodes
351 neurofibromin, a ubiquitous protein expressed in osteoblasts,
352 osteoclasts, chondrocytes, fibroblasts and endothelial cells, and
353 neurofibromin negatively regulates the activity of Ras which
354 involved in cellular proliferation and differentiation, leading to
355 the physically preventing bony bridging and endochondral bone
356 repairing by the recurrent invasion and increased proliferation of
357 fibrous lesion cells in the cellular deficiencies, suggesting the role
358 of *NF1* in bone development, homeostatic regulation, and repair
359 in Neurofibromatosis type 1 [88]. However, the complete
360 understanding of CPT cannot be fully attributed to the *NF1* gene
361 within the framework of neurofibromatosis type 1. And other
362 factors, including intrauterine trauma and generalized metabolic
363 disturbances at the time of birth, are believed to contribute [89].

Recent studies have identified several biological processes, such as skeletal system development, tissue morphogenesis and cell differentiation/proliferation, are associated to CPT[18, 19, 53], these findings suggest the involvement of intricate molecular processes in which other genes, in cooperation with *NF1*, may contribute to CPT. However, the precise mechanisms underlying these effects have not yet been elucidated.

In this study, we sequenced the transcriptome of periosteal tissues from three patients with CPT and two healthy individuals, and identified 180 up-regulated genes in CPT patients, which were enriched in 40 biological processes related to bone homeostasis. These biological processes encompassed skeletal system development, tissue remodeling (e.g., bone resorption), cell matrix assembly, regulation of cell adhesion, regulation of the MAPK cascade and autophagy in bone metabolism. These findings are consistent with previous studies and provide an overview of the biological processes likely associated with CPT, despite the limited number of patient samples included in this analysis. Furthermore, to gain insight into the effects of genomic LoF variations on CPT, we performed WES on 159 patients and observed significant excesses of rare LoF variations in 21 out of the 40 biological processes. This indicates the rare LoF variations

play an important role in CPT-related biological processes. Most importantly, through gene-pathway enrichment analysis using the KEGG database, we identified 63 patients carrying 40 genes with 56 LoF variations in the osteoclast differentiation pathway with its 3 directly regulated pathways (MAPK signaling pathway, Calcium signaling pathway and PI3K-Akt signaling pathway), as well as the fluid shear stress and atherosclerosis pathway. This provides a comprehensive illustration of how rare genomic LoF variations in the osteoclast differentiation pathways contribute to the pathological process of CPT.

Additionally, we also observed that 19 out of the 40 biological processes showed significant gene enrichment in the transcriptome but did not exhibit significant LoF excesses. This suggests the involvement of other types of genomic mutations contributing to CPT. For instance, among the CPT patients, apart from MAPK signaling pathway(P=3.67e-08), calcium signaling pathway (P=1.58e-10) and PI3K-Akt signaling pathway (P=4.15e-10), we found missense mutations exhibited the excesses specifically in apoptosis pathways (P=6.58e-05) and mTOR signaling pathway (P=3.15e-05) which have been confirmed as the crucial pathways orchestrating the bone resorption in bone metabolic diseases [90, 91].

Compared with the fact that up to 80%-84% of patients with CPT present with neurofibromatosis type 1, we found NF1 carried 9 LoF mutations in 12 patients, with a prevalence 19% in this study. This paradox in prevalence suggests some of our 56 LoF mutations may not directly cause CPT but instead participated in the regulation of the imbalanced bone homeostasis resulted from CPT. Furtherly considering the extreme low prevalence of CPT, de novo mutations would serve as more accurate genomic markers for the genetic diagnosis of CPT under our current pathway framework.

In summary, our findings highlight that rare genomic LoF variations in genes involved in osteoclast differentiation and regulation pathways play crucial role in the pathological process of CPT. This provides a new framework for understanding how genetic variations regulate biological processes in the pathology of CPT.

Method

Patient Recruitment and study subjects

A consecutive cohort of 165 CPT patients admitted into the department of Orthopedics of Hunan Children's Hospital from

2016 to 2020 was enrolled in this study. For WES, a total of 159 CPT patients, of which 75 cases had been reported in 2019[7], provided peripheral blood samples for DNA extraction and sequencing. For the transcriptome sequencing, two normal periosteal tissues were obtained from healthy children undergoing surgery for tibial fracture, as well as three diseased periosteum tissues were obtained from the patients with CPT during surgery at Hunan Children's Hospital. This study had been approved by the Ethics Committee of Hunan Children's Hospital, and informed consent had been obtained from all patients.

Transcriptome sequencing by nanopore platform and data analysis

RNA Extraction from Periosteal Tissue

During the surgical procedure, periosteal tissue devoid of bone needles and encompassing cortical bone was collected and immediately dissected into small fragments measuring approximately 200 mg with a diameter of approximately 5 mm. These fragments were then placed in centrifuge tubes filled with phosphate-buffered saline (PBS) and thoroughly washed. Subsequently, they were transferred to cryovials and snap-frozen in liquid nitrogen for storage. Total RNA extraction from the

periosteal tissue was performed using Takara TRIzol reagent according to the manufacturer's standard protocol. The extracted RNA was subjected to quality assessment using the IMPLEN Nano Photometer for spectrophotometric analysis, Nanodrop 2000 for concentration measurement of RNA samples, Agilent 2100 for integrity assessment, and agarose gel electrophoresis. One microgram of purified total RNA with a minimum RNA integrity number (RIN) value of 6 was reverse transcribed into complementary DNA (cDNA) using the cDNA-PCR sequencing kit (SQK-PCS109).

Full-Length Transcriptome Sequencing

The experimental workflow followed the standard procedure provided by Oxford Nanopore Technologies (ONT), including sample quality check, library construction, library quality assessment, and library sequencing. The total RNA extracted from periosteal tissue was required to be at least 1.5 μg, suitable for three library preparations, with a concentration of ≥40 ng/μl, a volume of ≥10 μl, an OD260/280 ratio between 1.7 and 2.5, an OD260/230 ratio between 0.5 and 2.5, and a normal peak at 260 nm absorption. Samples with an RIN value of 6 or higher were selected for full-length transcriptome sequencing using ONT. The

library construction process involved the following steps: First, oligo dT primers were utilized to anneal with the polyA tail of mRNA, followed by the addition of reverse transcriptase (RT) to synthesize the first cDNA strand. Next, specific PCR adapters were added to both ends of the first cDNA strand, and a 14-cycle cDNA PCR was performed using LongAmp Tag DNA polymerase (New England Biolabs, Ipswich) with an extension time of 8 minutes. This step amplified the double-stranded cDNA using amplification primers. Subsequently, DNA damage repair and end repair were conducted using the NEBNext FFPE DNA Repair Mix and NEBNext Ultra II End Repair/dA-Tailing Module to repair and end-repair the nucleic acid fragments, followed by A-tailing. The sequencing adapters were then ligated using the SQK-LSK110 kit (ONT), and purification was carried out using Agencourt XP beads (Beckman Coulter, USA). The prepared cDNA library was sequenced using the Oxford Nanopore PromethION 48 platform (Biomarker, Beijing).

After transcriptome sequencing, we trim the adaptors by software porechop v0.2.4(https://github.com/rrwick/Porechop), aligned the reads to reference hg19 and quantified gene expression using the FLAIR software [92]. We identified 179 significantly up-regulated genes and 134 significantly down-

493 regulated genes (|logFC| > 2 and FDR < 0.01) between CPT cases
494 and controls using the differential expression analysis module of
495 the Trinity software [93]. Furthermore, we obtained the
496 significantly enriched pathways (supplementary table 1) of these
497 179 up-regulated genes using the Metascape software [94]
498 (accessed May 2022).

499 WES, variation calling and filtering

500 Genomic DNA from peripheral blood was extracted using the
501 standard phenol-chloroform method. The DNA of all 159 CPT
502 patients was fragmented and the exome was captured using the
503 Agilent SureSelect Human All Exon V6 kit. The captured DNA
504 was sequenced with 100 bp pair-end reads using the
505 BGISEQ2000 platform, following the manufacturer's
506 instructions. Each sample yielded an average of 13.3 Gb of raw
507 data with an average depth of 132 and a coverage ratio of
508 99.84% of the captured target region of 60.46 M. Over 98%
509 (average ~92.9%) of the bases had a phred quality score > 30.
510 For the variation calling, the raw reads were filtered by
511 SOAPnuke [95] and then mapped to the human genome
512 reference (UCSC/GRCh37/hg19) using the Burrows-Wheeler
513 Aligner (BWA-MEM, version 0.7.10) [96]. Variants were called

using the Genome Analysis Tool Kit (GATK, version 4.1) [97]. To validate the accuracy of the variation calling, 208 Chinese Han individuals from the 1000 Genomes Project were selected as a control to perform principal component analysis (PCA). This allowed us to remove the false positive WES variations by setting genotype quality > 30, the minimum read depth of 20 for each genotype, and an additional setting of the allele balance of 0.3 for each heterozygous genotype. Furthermore, all indels with allele lengths greater than 10 bp were removed. The Variant Effect Predictor (VEP) [98] was used to annotate and classify all the variants. Subsequently, the variants were filtered based on their frequency in public database gnomAD (https://gnomad.broadinstitute.org/) and the bi-allelic LoF variants (frameshift_variant, splice_acceptor_variant, splice_donor_variant, start_lost, stop_gained and stop_lost) with MAF < 0.001 were retained.

LoF excess test for pathway gene sets

After obtaining the VCF file of LoF variations, the software rvtest[99] was used to test the LoF excess for each gene set from the biological processes or KEGG pathways. The burden test was performed using the CMC model, with a control population

consisting of 208 Chinese Han individuals selected from the 1000 Genomes Project as the control population against the 159 patients with CPT. The VCF files containing the variations information of both the 159 patients and the 208 control individuals were at the service for request. The genes corresponding to the biological processes were obtained from http://amigo.geneontology.org/amigo/term/GO:0008150, with the organism selected as Homo sapiens. The genes of KEGG pathways were downloaded from https://www.genome.jp/kegg/.

Tables

Table 1. LoF excess in 21 biological processes enriched in the GO database

Go Term	Num.LoF	Num.Case	Num.Ctrl	OR	P.fisher	P.rvtest
regulation_of_MAPK_cascade_GO:0043408	34	35	1	35	4.13E-13	2.81E-12
monoatomic_ion_homeostasis_GO:0050801	34	33	0	33	2.74E-12	5.69E-12
regulation_of_hormone_levels_GO:0010817	27	29	0	29	1.13E-10	1.38E-10
inflammatory_response_GO:0006954	33	28	0	28	2.81E-10	1.38E-10
positive_regulation_of_cytokine_production_GO:0001819	27	24	0	24	1.03E-08	3.14E-09
ossification_GO:0001503	18	23	0	23	2.50E-08	1.46E-08

regulation_of_monoatomic_ion_transport_GO:0043269	42	35	2	17.5	4.86E-12	3.20E-11
myeloid_leukocyte_differentiation_GO:0002573	13	17	0	17	4.45E-06	1.37E-06
regulation_of_bone_resorption_GO:0045124	11	16	0	16	1.03E-05	2.89E-06
regulation_of_cell_growth_GO:0001558	16	16	0	16	1.03E-05	2.89E-06
response_to_lipopolysaccharide_GO:0032496	15	16	1	16	1.03E-05	1.51E-05
ameboidal-type_cell_migration_GO:0001667	15	14	1	14	5.41E-05	6.58E-05
regulation_of_binding_GO:0051098	14	14	0	14	5.41E-05	1.28E-05
negative_regulation_of_locomotion_GO:0040013	27	27	2	13.5	6.33E-09	1.73E-08
negative_regulation_of_transport_GO:0051051	36	36	3	12	1.59E-11	2.94E-11
cell_morphogenesis_GO:0000902	23	23	2	11.5	1.94E-07	3.62E-07
skeletal_system_development_GO:0001501	26	23	2	11.5	1.94E-07	3.62E-07
regulation_of_T_cell_proliferation_GO:0042129	10	11	1	11	0.000612	0.00059
response_to_nutrient_GO:0007584	9	11	0	11	0.000612	0.000117
negative_regulation_of_response_to_external_stimulus_GO:0032102	24	21	2	10.5	1.03E-06	7.66E-07
B_cell_mediated_immunity_GO:0019724	10	10	0	10	0.001348	0.000117
extracellular_matrix_organization_GO:0030198	36	36	4	9	1.24E-10	2.77E-10
multicellular_organismal_homeostasis_GO:0048871	32	35	4	8.75	2.85E-10	6.09E-10
autophagy_GO:0006914	16	17	2	8.5	4.03E-05	3.06E-05
positive_regulation_of_cold-induced_thermogenesis_GO:0120162	6	8	0	8	0.011904	0.001072
positive_regulation_of_hydrolase_activity_GO:0051345	38	40	5	8	1.75E-11	4.54E-11
positive_regulation_of_cell_development_GO:0010720	22	19	4	4.75	0.0001	8.60E-05
cell_junction_organization_GO:0034330	18	18	4	4.5	0.000217	0.000171

heterotypic_cell-cell_adhesion_GO:0034113	5	4	1	4	0.170759	0.095634
endoderm_development_GO:0007492	3	3	0	3	0.320244	0.04668
tissue_remodeling_GO:0048771	3	3	1	3	0.320244	0.198617
regulation_of_homotypic_cell-cell_adhesion_GO:0034110	2	2	0	2	0.581098	0.104816
regulation_of_p38MAPK_cascade_GO:1900744	2	2	0	2	0.581098	0.104816
regulation_of_SMAD_protein_signal_transduction_GO:0060390	1	2	0	2	0.581098	0.104816
developmental_pigmentation_GO:0048066	3	3	2	1.5	0.656041	0.448644
innervation_GO:0060384	1	1	0	1	1	0.252078

549

550 Table 2. KEGG pathway enrichment of THE suffering from LoF

551 variations

Category	Term	Description	LogP	Log(q-value)	InTerm_
KEGG Pathway	hsa04380	Osteoclast differentiation	-4.83279	-2.616	8/128
KEGG Pathway	hsa05152	Tuberculosis	-4.58891	-2.616	9/180
KEGG Pathway	hsa05418	Fluid shear stress and atherosclerosis	-4.5734	-2.616	8/139
KEGG Pathway	hsa05204	Chemical carcinogenesis - DNA adducts	-4.56047	-2.616	6/69
KEGG Pathway	hsa04010	MAPK signaling pathway	-2.99947	-1.453	9/294
KEGG Pathway	hsa04020	Calcium signaling pathway	-2.95648	-1.451	8/240
KEGG Pathway	hsa04151	PI3K-Akt signaling pathway	-2.45406	-1.112	9/354
KEGG Pathway	hsa00280	Valine, leucine and isoleucine degradation	-2.09919	-0.836	3/48
KEGG Pathway	hsa04750	Inflammatory mediator regulation of TRP channels	-2.00287	-0.799	4/98

552

553 Table 3. The list of CPT-causing LoF variations

CHR	POS	REF	ALT	Consequence	Symbol	Biotype	gnomAD_EAS_AF	HGVSc	HGVSp
1	26857234	A	G	splice_acceptor_variant&NMD_transcript_variant	RPS6KA1	nonsense_mediated_decay	0	ENST00000525525.1:c.64-2A>G	.

1	110201481	C	T	stop_gained	GSTM4	protein_coding	0	ENST00000326729.5:c.406C>T	ENSP00000316471.5:p.Gln136Ter
1	110201494	A	AAG	frameshift_variant	GSTM4	protein_coding	5.44E-05	ENST00000326729.5:c.419_420insAG	ENSP00000316471.5:p.Phe141GlyfsTer17
1	155942054	G	A	stop_gained	ARHGEF2	protein_coding	0.0002787	ENST00000368315.4:c.58C>T	ENSP00000357298.3:p.Arg20Ter
1	161514550	CT	C	frameshift_variant	FCGR3A	protein_coding	0.0002175	ENST00000367967.3:c.517del	ENSP00000356944.3:p.Arg173GlyfsTer14
1	169699572	C	T	splice_donor_variant	SELE	protein_coding	0.0003343	ENST00000333360.7:c.715+1G>A	.
1	175096197	C	G	stop_gained	TNN	protein_coding	5.44E-05	ENST00000239462.4:c.3021C>G	ENSP00000239462.4:p.Tyr1007Ter
2	70780315	C	T	splice_donor_variant	TGFA	protein_coding	0.0001906	ENST00000444975.1:c.58+1G>A	.
3	48669375	G	A	stop_gained	SLC26A6	protein_coding	5.57E-05	ENST00000358747.6:c.625C>T	ENSP00000351597.6:p.Gln209Ter
3	49939857	G	A	stop_gained	MST1R	protein_coding	0.0001631	ENST00000296474.3:c.1186C>T	ENSP00000296474.3:p.Arg396Ter
3	130120699	G	T	splice_donor_variant	COL6A5	protein_coding	0.0002753	ENST00000265379.6:c.4227+1G>T	.
3	130189813	T	C	splice_donor_variant	COL6A5	protein_coding	0.0001126	ENST00000265379.6:c.7574+2T>C	.
3	190333145	T	C	splice_donor_variant	IL1RAP	protein_coding	0.0005515	ENST00000412080.1:c.47+2T>C	.
4	154625398	C	T	stop_gained	TLR2	protein_coding	0.0004349	ENST00000260010.6:c.1339C>T	ENSP00000260010.6:p.Arg447Ter
5	140012655	A	G	start_lost&splice_region_variant	CD14	protein_coding	0	ENST00000302014.6:c.2T>C	ENSP00000304236.6:p.Met1?

6	123702514	CT	C	frameshift_variant&splice_region_variant	TRDN	protein_coding	0.000244	ENST00000334268.4:c.1186del	ENSP00000333984.4:p.Lys396AsnfsTer27
6	154567863	C	T	stop_gained	OPRM1	protein_coding	0.000271	ENST00000337049.4:c.1201C>T	ENSP00000338381.4:p.Arg401Ter
6	161507701	T	C	splice_donor_variant	MAP3K4	protein_coding	0.000167	ENST00000348824.7:c.2556+2T>C	.
7	93125241	C	T	start_lost	CALCR	protein_coding	0.000784	ENST00000359558.2:c.3G>A	ENSP00000352561.2:p.Met1?
7	150708074	GGTAA	G	splice_donor_variant&intron_variant	NOS3	protein_coding	0	ENST00000297494.3:c.2984+3_2984+6del	.
8	42037806	G	T	stop_gained	PLAT	protein_coding	5.44E-05	ENST00000220809.4:c.1176C>A	ENSP00000220809.4:p.Tyr392Ter
9	131467786	C	T	stop_gained	PKN3	protein_coding	0	ENST00000291906.4:c.229C>T	ENSP00000291906.4:p.Arg77Ter
10	51447444	C	T	stop_gained	AKR1C3	protein_coding	0	ENST00000380554.3:c.748C>T	ENSP00000369927.3:p.Arg250Ter
10	52550013	AG	A	frameshift_variant	AKR1C4	protein_coding	5.44E-05	ENST00000263126.1:c.738del	ENSP00000263126.1:p.Lys247AsnfsTer8
10	71175865	AT	A	frameshift_variant	TACR2	protein_coding	0.000326	ENST00000373306.4:c.214del	ENSP00000362403.4:p.Ile72SerfsTer53
10	81112241	A	T	splice_acceptor_variant	PPIF	protein_coding	0.000381	ENST00000225174.3:c.316-2A>T	.
10	106025900	AC	A	frameshift_variant	GSTO1	protein_coding	5.44E-05	ENST00000369710.4:c.427del	ENSP00000358724.4:p.Leu143SerfsTer12
11	64507923	T	C	splice_acceptor_variant&NMD_transcript_variant	RASGRP2	nonsense_mediated_decay	0.000185	ENST00000445445.1:c.372-2A>G	.

11	65374790	G	A	stop_gained&NMD_transcript_variant	MAP3K11	nonsense_mediated_decay	0.0002229	ENST00000524848.1:c.556C>T	ENSP00000431506.1:p.Gln186Ter
12	18854625	C	T	splice_donor_variant	PLCZ1	protein_coding	0.0002179	ENST00000266505.7:c.949+1G>A	.
12	18890316	C	G	splice_donor_variant&non_coding_transcript_variant	PLCZ1	processed_transcript	0.0002721	ENST00000541109.1:n.235+1G>C	.
12	68549150	G	A	stop_gained	IFNG	protein_coding	0	ENST00000229135.3:c.484C>T	ENSP00000229135.3:p.Arg162Ter
17	38069112	G	A	stop_gained	P2RX1	protein_coding	0	ENST00000225538.3:c.538C>T	ENSP00000225538.3:p.Arg180Ter
17	29496923	CTG	CTT	frameshift_variant	NF1	protein_coding	0	ENST00000356175.3:c.499_502del	ENSP00000348498.3:p.Cys167GlnfsTer10
17	29497003	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.574C>T	ENSP00000348498.3:p.Arg192Ter
17	29533315	C	T	stop_gained	NF1	protein_coding	5.44E-05	ENST00000356175.3:c.1318C>T	ENSP00000348498.3:p.Arg440Ter
17	29553492	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.2041C>T	ENSP00000348498.3:p.Arg681Ter
17	29557401	G	A	splice_donor_variant	NF1	protein_coding	0	ENST00000356175.3:c.3113+1G>A	.
17	29562981	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.3916C>T	ENSP00000348498.3:p.Arg1306Ter
17	29576111	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.4084C>T	ENSP00000348498.3:p.Arg1362Ter
17	29679366	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.7486C>T	ENSP00000348498.3:p.Arg2496Ter
17	29684326	C	T	stop_gained	NF1	protein_coding	0	ENST00000356175.3:c.7846C>T	ENSP00000348498.3:p.Arg2616Ter

17	39768464	ACT	A	frameshift_variant	KRT16	protein_coding	0	ENST00000301653.4:c.475_476del	ENSP00000301653.3:p.Ser159Ter
17	39768467	G	GCA	frameshift_variant	KRT16	protein_coding	0	ENST00000301653.4:c.473_474insTG	ENSP00000301653.3:p.Ser159AlafsTer17
19	17955240	C	T	splice_acceptor_variant	JAK3	protein_coding	0	ENST00000458235.1:c.-13-1G>A	.
19	40740862	G	A	stop_gained	AKT2	protein_coding	0	ENST00000579047.1:c.1270C>T	ENSP00000471369.1:p.Arg424Ter
19	49657846	G	A	stop_gained	HRC	protein_coding	0	ENST00000252825.4:c.649C>T	ENSP00000252825.3:p.Arg217Ter
19	54778581	GTT	G	frameshift_variant	LILRB2	protein_coding	0	ENST00000314446.5:c.1748_1749del	ENSP00000319960.5:p.Glu583AlafsTer4
19	54778586	CT	C	frameshift_variant	LILRB2	protein_coding	0	ENST00000314446.5:c.1744del	ENSP00000319960.5:p.Arg582GlyfsTer68
19	55086356	C	CGT	frameshift_variant	LILRA2	protein_coding	0	ENST00000251376.3:c.511_512insGT	ENSP00000251376.3:p.His171ArgfsTer16
19	55086358	TGC	T	frameshift_variant	LILRA2	protein_coding	0	ENST00000251376.3:c.514_515del	ENSP00000251376.3:p.Ala172ProfsTer27
19	55143609	G	A	stop_gained	LILRB1	protein_coding	0.0007068	ENST00000324602.7:c.582G>A	ENSP00000315997.7:p.Trp194Ter
19	55146627	G	A	stop_gained	LILRB1	protein_coding	0.0001643	ENST00000324602.7:c.1559G>A	ENSP00000315997.7:p.Trp520Ter
19	55179243	AG	A	frameshift_variant&splice_region_variant	LILRB4	protein_coding	0.000492	ENST00000270452.2:c.1200+1del	.
19	55179246	T	A	splice_donor_variant	LILRB4	protein_coding	0.0005468	ENST00000270452.2:c.1200+2T>A	.
20	42195065	ACC	A	frameshift_variant	SGK2	protein_coding	5.44E-05	ENST00000341458.4:c.115_116del	ENSP00000340608.4:p.Pro39AsnfsTer17

Figure Legends

Fig1. Identification of differentially expressed genes associated with CPT in the transcriptome sequencing (a) Volcano plot for the comparison between the 3 CPT (JG159_nonNF1, JG250_NF1 and JG252_NF1) and 2 healthy individuals (GDZ001_Control and GDZ002_Control). The cutoff values fold change >2 and FDR < 0.01 were utilized to identify differentially expressed genes. Non-changed genes were shown in black color. Red color is indicative of the up-regulated genes and the down-regulated genes. (b) Heatmap plot of differentially expressed genes. The Pearson correlation distance metric and the average linkage clustering algorithm were used. (c) the GO enrichment of the 179 up-regulated genes in the transcriptome. X-axis indicates the -log(P) in which P is the probability that the genes are enriched in the biological process in GO database. Here the threshold of -log(P) is set as 2, indicating the P value 0.01.

Fig2. The 179 up-regulated genes in patients with CPT significantly enriched in 40 biological processes in the GO biological processes (the Bonferroni corrected P= 0.05/100=5e-4) among which 21 revealed the LoF excesses (under the criteria OR >10 and P > 0.01). X-axis indicates the biological processes from GO database and Y-axis indicates the odd ratio (OR) that the

LoF mutations enrich in the biological process between 159 CPT cases and 208 controls. The number above the bar in each process is the P values indicating the enrichment of LoF mutations in this process.

Fig3. The KEGG pathway enrichment analysis and the LoF excesses of the transcriptome high expression (THE) gene set suffering from rare genomic LoF variations. (a) The KEGG enrichment of THE carrying the LoF mutations. X-axis indicates the -log(P) in which P is the probability that the genes are enriched in the pathway. Here the threshold of -log(P) is set as 2, indicating the P value 0.01. (b) the LoF excesses of the significant KEGG pathways enriched by THE. X-axis indicates the pathways from KEGG database and Y-axis indicates the odd ratio(OR) that the LoF mutations enrich in the pathway between 159 CPT cases and 208 controls. The numbers above the bars are the -log(P) in which P indicate the enrichment of LoF mutations in the pathway. Here the threshold of -log(P) is 2, indicating the threshold P 0.01.

Data availability

Both RNA-seq and WES data are at the service for request.

Authors information

Hui Huang, Huanhuan Peng and Haibo Mei conceived and designed the study. Guanghui Zhu and Yu Zheng collected the samples and performed the experimental analysis. Nan Li performed the data analysis and wrote the manuscript. All the authors read and approved the final manuscript.

Declaration of conflicting interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics declarations

The experimental protocol was approved by the Research Ethics Committee of Hunan Children's Hospital. Informed consent was obtained from every patient.

Fundings

Opening fundings of Hunan Provincial Key Laboratory of Pediatric Orthopedics

References

[1]. Hefti, F., et al., Congenital pseudarthrosis of the tibia: history, etiology, classification, and epidemiologic data. J Pediatr Orthop B, 2000. 9(1): p. 11-5.

[2]. Shannon, C.E., A.J. Huser and D. Paley, Cross-Union Surgery for Congenital Pseudarthrosis of the Tibia. Children (Basel), 2021. 8(7).

[3]. Han, J., et al., A benign form of congenital anterolateral bowing of the tibia associated with ipsilateral polydactyly of the hallux: case report and literature review. Am J Med Genet A, 2012. 158A(7): p. 1742-9.

[4]. Zheng, Y., et al., Case series of congenital pseudarthrosis of the tibia unfulfilling neurofibromatosis type 1 diagnosis: 21% with somatic NF1 haploinsufficiency in the periosteum. Hum Genet, 2022. 141(8): p. 1371-1383.

[5]. Xu, J., et al., Clinical characteristics and in silico analysis of congenital pseudarthrosis of the tibia combined with neurofibromatosis type 1 caused by a novel NF1 mutation. Front Genet, 2022. 13: p. 991314.

[6]. Brekelmans, C., et al., Neurofibromatosis type 1-related pseudarthrosis: Beyond the pseudarthrosis site. Hum Mutat, 2019. 40(10): p. 1760-1767.

[7]. Zhu, G., et al., Identification and characterization of NF1 and non-NF1 congenital pseudarthrosis of the tibia based on germline NF1 variants: genetic and clinical analysis of 75 patients. Orphanet J Rare Dis, 2019. 14(1): p. 221.

[8]. Lee, S.M., et al., Is double inactivation of the Nf1 gene responsible for the development of congenital pseudarthrosis of the tibia associated with NF1? J Orthop Res, 2012. 30(10): p. 1535-40.

[9]. El-Hoss, J., et al., A murine model of neurofibromatosis type 1 tibial pseudarthrosis featuring proliferative fibrous tissue and osteoclast-like cells. J Bone Miner Res, 2012. 27(1): p. 68-78.

[10]. Sant, D.W., et al., Evaluation of somatic mutations in tibial pseudarthrosis samples in neurofibromatosis type 1. J Med Genet, 2015. 52(4): p. 256-61.

[11]. Stevenson, D.A., et al., Double inactivation of NF1 in tibial pseudarthrosis. Am J Hum Genet, 2006. 79(1): p. 143-8.

[12]. Leskelä, H.V., et al., Congenital pseudarthrosis of neurofibromatosis type 1: impaired osteoblast differentiation and function and altered NF1 gene expression. Bone, 2009. 44(2): p. 243-50.

[13]. Sakamoto, A., et al., Congenital pseudarthrosis of the tibia: analysis of the histology and the NF1 gene. J Orthop Sci, 2007. 12(4): p. 361-5.

[14]. Kamiya, N., et al., Targeted Disruption of NF1 in Osteocytes Increases FGF23 and Osteoid With Osteomalacia-like Bone Phenotype. J Bone Miner Res, 2017. 32(8): p. 1716-1726.

[15]. Yang, G., et al., Serum-derived exosomes from neurofibromatosis type 1 congenital tibial pseudarthrosis impaired bone by promoting osteoclastogenesis and inhibiting osteogenesis. Exp Biol Med (Maywood), 2021. 246(2): p. 130-141.

[16]. Julien, A., et al., FGFR3 in Periosteal Cells Drives Cartilage-to-Bone Transformation in Bone Repair. Stem Cell Reports, 2020. 15(4): p. 955-967.

[17]. Qu, H., et al., The effects of vasoactive intestinal peptide on RANKL-induced osteoclast formation. Ann Transl Med, 2021. 9(2): p. 127.

[18]. Yang, G., et al., Serum-derived exosomes from neurofibromatosis type 1 congenital tibial pseudarthrosis impaired bone by promoting osteoclastogenesis and inhibiting osteogenesis. Experimental Biology and Medicine, 2021. 246(2): p. 130-141.

[19]. Liu, Y., et al., New insights into pathogenesis of congenital pseudarthrosis of tibia in children using periosteum proteomics analysis. Rapid Commun Mass Spectrom, 2022. 36(21): p. e9374.

[20]. Kenkre, J.S. and J. Bassett, The bone remodelling cycle. Ann Clin Biochem, 2018. 55(3): p. 308-327.

[21]. Salhotra, A., et al., Mechanisms of bone development and repair. Nature Reviews Molecular Cell Biology, 2020. 21(11): p. 696-711.

[22]. Chan, W., et al., Regulation and Role of Transcription Factors in Osteogenesis. Int J Mol Sci, 2021. 22(11).

[23]. Yahara, Y., et al., The origins and roles of osteoclasts in bone development, homeostasis and repair. Development, 2022. 149(8).

[24]. Cushion, T.D., et al., De novo mutations in the beta-tubulin gene TUBB2A cause simplified gyral patterning and infantile-onset epilepsy. Am J Hum Genet, 2014. 94(4): p. 634-41.

[25]. Guella, I., et al., De Novo Mutations in YWHAG Cause Early-Onset Epilepsy. Am J Hum Genet, 2017. 101(2): p. 300-310.

[26]. Kanca, O., et al., De Novo Variants in WDR37 Are Associated with Epilepsy, Colobomas, Dysmorphism, Developmental Delay, Intellectual Disability, and Cerebellar Hypoplasia. Am J Hum Genet, 2019. 105(3): p. 672-674.

[27]. Schoch, K., et al., A Recurrent De Novo Variant in NACC1 Causes a Syndrome Characterized by Infantile Epilepsy, Cataracts, and Profound Developmental Delay. Am J Hum Genet, 2017. 100(2): p. 343-351.

[28]. Heyne, H.O., et al., De novo variants in neurodevelopmental disorders with epilepsy. Nat Genet, 2018. 50(7): p. 1048-1053.

[29]. Goos, J., et al., A de novo substitution in BCL11B leads to loss of interaction with transcriptional complexes and craniosynostosis. Hum Mol Genet, 2019. 28(15): p. 2501-2513.

[30]. Ehmke, N., et al., De Novo Mutations in SLC25A24 Cause a Craniosynostosis Syndrome with Hypertrichosis, Progeroid Appearance, and Mitochondrial Dysfunction. Am J Hum Genet, 2017. 101(5): p. 833-843.

[31]. Timberlake, A.T., et al., De novo mutations in inhibitors of Wnt, BMP, and Ras/ERK signaling pathways in non-syndromic midline craniosynostosis. Proc Natl Acad Sci U S A, 2017. 114(35): p. E7341-E7347.

[32].Awadalla, P., et al., Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am J Hum Genet, 2010. 87(3): p. 316-24.

[33]. O'Roak, B.J., et al., Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun, 2014. 5: p. 5595.

[34]. O'Roak, B.J., et al., Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet, 2011. 43(6): p. 585-9.

[35]. Uddin, M., et al., Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. Nat Genet, 2014. 46(7): p. 742-7.

[36].Iossifov, I., et al., The contribution of de novo coding mutations to autism spectrum disorder. Nature, 2014. 515(7526): p. 216-21.

[37]. Neale, B.M., et al., Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature, 2012. 485(7397): p. 242-5.

[38]. O'Roak, B.J., et al., Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature, 2012. 485(7397): p. 246-50.

[39]. Sanders, S.J., et al., De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature, 2012. 485(7397): p. 237-41.
[40]. Sanders, S.J., et al., Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron, 2011. 70(5): p. 863-85.
[41]. Luo, T., et al., De novo mutations in folate-related genes associated with common developmental disorders. Comput Struct Biotechnol J, 2021. 19: p. 1414-1422.
[42]. Watkins, W.S., et al., De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. Nat Commun, 2019. 10(1): p. 4722.
[43]. Jin, S.C., et al., Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat Genet, 2017. 49(11): p. 1593-1601.
[44]. Sifrim, A., et al., Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet, 2016. 48(9): p. 1060-5.
[45]. Homsy, J., et al., De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science, 2015. 350(6265): p. 1262-6.
[46]. Zaidi, S., et al., De novo mutations in histone-modifying genes in congenital heart disease. Nature, 2013. 498(7453): p. 220-3.
[47]. García-García, A. and I. Martin, Extracellular Matrices to Modulate the Innate Immune Response and Enhance Bone Healing. Front Immunol, 2019. 10: p. 2256.
[48]. Wang, J., et al., Spatial-Temporal Patterns and Inflammatory Factors of Bone Matrix Remodeling. Stem Cells Int, 2021. 2021: p. 4307961.
[49]. Ajibade, A.A., H.Y. Wang and R.F. Wang, Cell type-specific function of TAK1 in innate immune signaling. Trends Immunol, 2013. 34(7): p. 307-16.
[50]. Guo, Y.F., et al., The role of autophagy in bone homeostasis. J Cell Physiol, 2021. 236(6): p. 4152-4173.
[51]. Montaseri, A., et al., The Role of Autophagy in Osteoclast Differentiation and Bone Resorption Function. Biomolecules, 2020. 10(10).
[52]. Yang, G., et al., Are Children Suffering From Congenital Pseudarthrosis of the Tibia Associated With Decreased Bone Strength? Front Pediatr, 2022. 10: p. 859580.
[53]. Li, Z., et al., Differential expression and effect analysis of lncRNA-mRNA in congenital pseudarthrosis of the tibia. Front Genet, 2023. 14: p. 1094298.
[54]. Tan, Q., et al., The neurofibromatosis type I gene promotes autophagy via mTORC1 signalling pathway to enhance new bone formation after fracture. J Cell Mol Med, 2020. 24(19): p. 11524-11534.
[55]. Wang, F.S., et al., Biophysical Modulation of the Mitochondrial Metabolism and Redox in Bone Homeostasis and Osteoporosis: How Biophysics Converts into Bioenergetics. Antioxidants (Basel), 2021. 10(9).
[56]. Kim, J.M., et al., Osteoblast-Osteoclast Communication and Bone Homeostasis. Cells, 2020. 9(9).
[57]. Delgado-Calle, J. and T. Bellido, The osteocyte as a signaling cell. Physiol Rev, 2022. 102(1): p. 379-410.
[58]. Peng, Y., et al., Type H blood vessels in bone modeling and remodeling. Theranostics, 2020. 10(1): p. 426-436.
[59]. Tuckermann, J. and R.H. Adams, The endothelium-bone axis in development, homeostasis and bone and joint disease. Nat Rev Rheumatol, 2021. 17(10): p. 608-620.
[60]. Stucker, S., et al., Bone Angiogenesis and Vascular Niche Remodeling in Stress, Aging, and Diseases. Front Cell Dev Biol, 2020. 8: p. 602269.

[61]. Diomede, F., et al., Functional Relationship between Osteogenesis and Angiogenesis in Tissue Regeneration. Int J Mol Sci, 2020. 21(9).

[62]. Kim, H.D., et al., Biomimetic whitlockite inorganic nanoparticles-mediated in situ remodeling and rapid bone regeneration. Biomaterials, 2017. 112: p. 31-43.

[63]. Zhang, R., et al., Unique roles of phosphorus in endochondral bone formation and osteocyte maturation. J Bone Miner Res, 2011. 26(5): p. 1047-56.

[64]. Fischer, V. and M. Haffner-Luntzer, Interaction between bone and immune cells: Implications for postmenopausal osteoporosis. Semin Cell Dev Biol, 2022. 123: p. 14-21.

[65]. Goltzman, D. and G.N. Hendy, The calcium-sensing receptor in bone--mechanistic and therapeutic insights. Nat Rev Endocrinol, 2015. 11(5): p. 298-307.

[66]. Qin, L., et al., Molecular mechanosensors in osteocytes. Bone Res, 2020. 8: p. 23.

[67]. Anderson, P.H., Vitamin D Activity and Metabolism in Bone. Curr Osteoporos Rep, 2017. 15(5): p. 443-449.

[68]. Nakamichi, Y., et al., Mechanisms involved in bone resorption regulated by vitamin D. J Steroid Biochem Mol Biol, 2018. 177: p. 70-76.

[69]. Zuo, H. and Y. Wan, Nuclear Receptors in Skeletal Homeostasis. Curr Top Dev Biol, 2017. 125: p. 71-107.

[70]. Kim, H.J., et al., Phospholipase D2 controls bone homeostasis by modulating M-CSF-dependent osteoclastic cell migration and microtubule stability. Exp Mol Med, 2022. 54(8): p. 1146-1155.

[71]. Vykhovanets, E.P., et al., Concentration of several osteotropic growth factors, markers of osteogenesis and biologically active molecules in the blood serum of patients with congenital pseudarthrosis of tibia during orthopaedic treatment with combined technologies. Biomeditsinskaya Khimiya, 2018. 64(6): p. 525-533.

[72]. Heervä, E., et al., Osteoclasts in neurofibromatosis type 1 display enhanced resorption capacity, aberrant morphology, and resistance to serum deprivation. Bone, 2010. 47(3): p. 583-90.

[73]. Stevenson, D.A., et al., Bone resorption in syndromes of the Ras/MAPK pathway. Clin Genet, 2011. 80(6): p. 566-73.

[74]. Sharma, R., et al., Hyperactive Ras/MAPK signaling is critical for tibial nonunion fracture in neurofibromin-deficient mice. Hum Mol Genet, 2013. 22(23): p. 4818-28.

[75]. Nedeva, I.R., et al., Role of OSCAR Signaling in Osteoclastogenesis and Bone Disease. Front Cell Dev Biol, 2021. 9: p. 641162.

[76]. Park, D.R., et al., Osteoclast-associated receptor blockade prevents articular cartilage destruction via chondrocyte apoptosis regulation. Nature Communications, 2020. 11(1): p. 4343.

[77]. Place, D.E., et al., Osteoclast fusion and bone loss are restricted by interferon inducible guanylate binding proteins. Nature Communications, 2021. 12(1): p. 496.

[78]. Chen, X., et al., LY3023414 inhibits both osteogenesis and osteoclastogenesis through the PI3K/Akt/GSK3 signalling pathway. Bone Joint Res, 2021. 10(4): p. 237-249.

[79]. Liu, P., et al., Effects of Mechanical Stress Stimulation on Function and Expression Mechanism of Osteoblasts. Front Bioeng Biotechnol, 2022. 10: p. 830722.

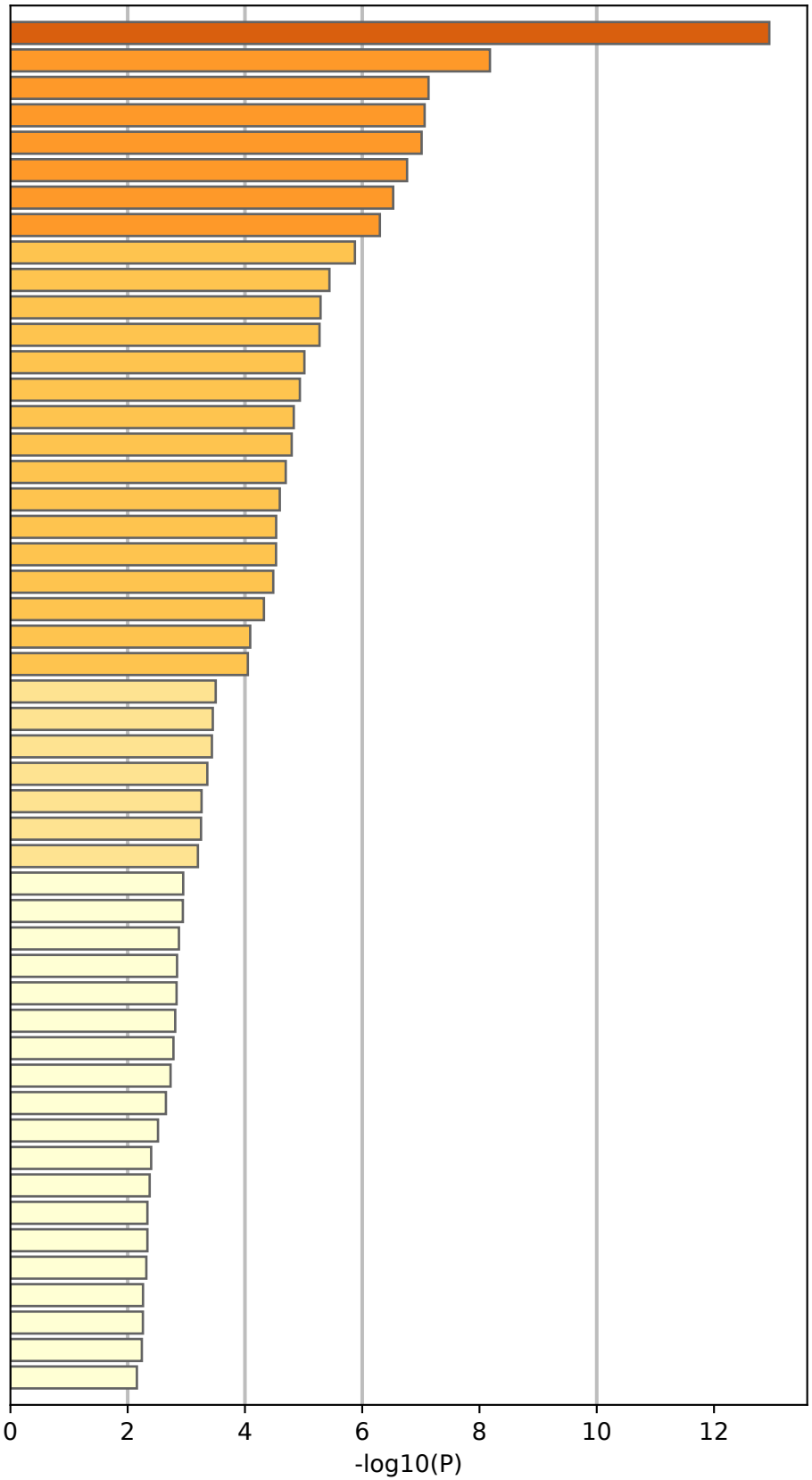
[80]. Wang, J., Y.X. Sun and J. Li, The role of mechanosensor Piezo1 in bone homeostasis and mechanobiology. Dev Biol, 2023. 493: p. 80-88.

[81]. Ma, Q., et al., Significance of mechanical loading in bone fracture healing, bone regeneration, and vascularization. J Tissue Eng, 2023. 14: p. 20417314231172573.

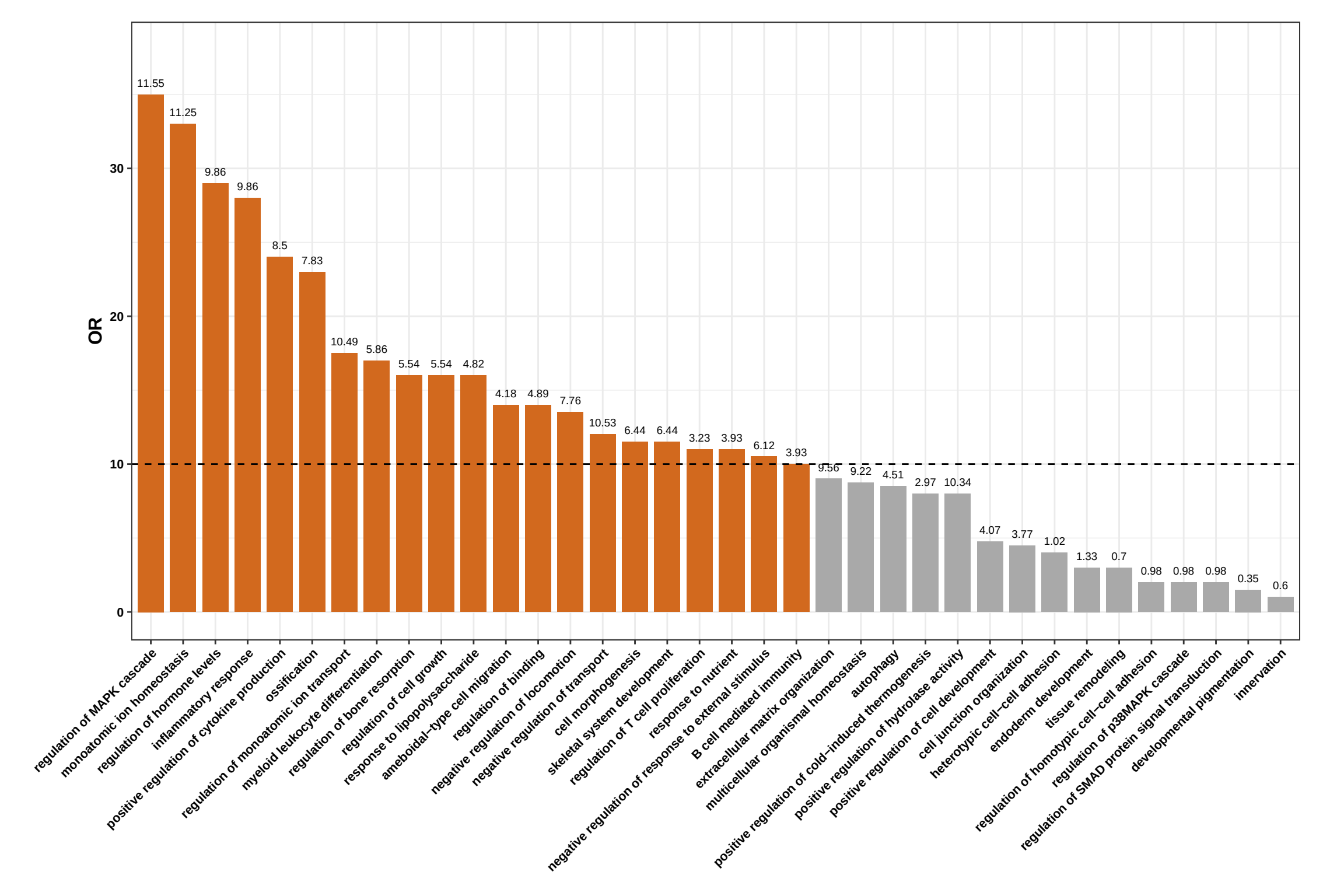
[82]. Zhou, H., et al., Research progress on the hedgehog signalling pathway in regulating bone formation

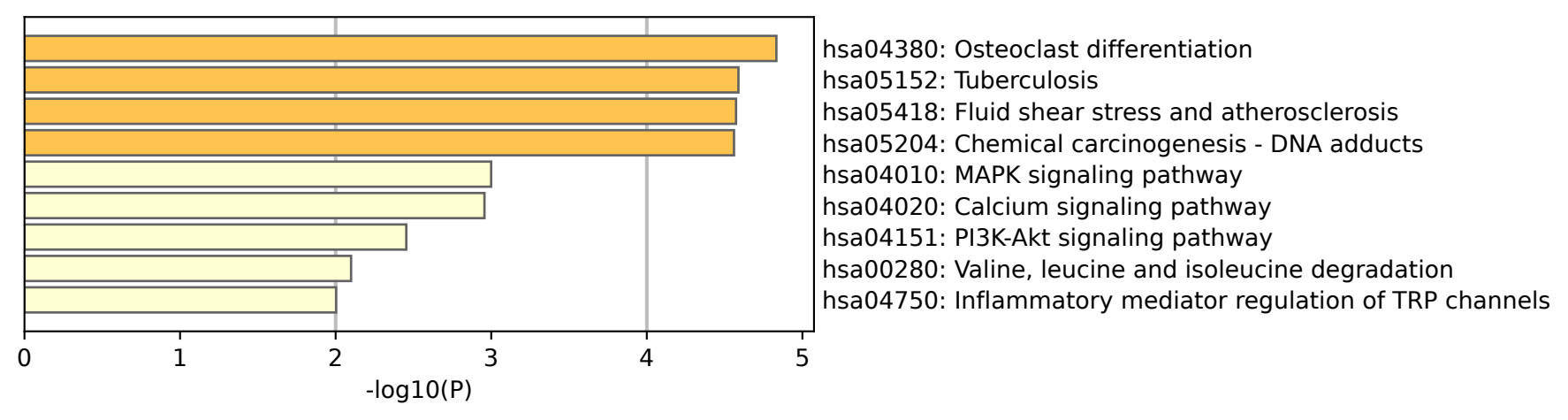
789 and homeostasis. Cell Prolif, 2022. 55(1): p. e13162.
790 [83]. Wang, H., et al., Hedgehog signaling orchestrates cartilage-to-bone transition independently of
791 Smoothened. Matrix Biol, 2022. 110: p. 76-90.
792 [84]. Hwang, S.H., et al., The G protein-coupled receptor Gpr161 regulates forelimb formation, limb
793 patterning and skeletal morphogenesis in a primary cilium-dependent manner. Development, 2018.
794 145(1).
795 [85]. Cai, W., Y. Su and G. Nan, Novel method for the treatment of congenital pseudarthrosis of the tibia
796 using the gastrocnemius flap: A preliminary study. J Child Orthop, 2022. 16(3): p. 167-173.
797 [86]. Siebert, M.J. and C.A. Makarewich, Anterolateral Tibial Bowing and Congenital Pseudoarthrosis
798 of the Tibia: Current Concept Review and Future Directions. Curr Rev Musculoskelet Med, 2022.
799 15(6): p. 438-446.
800 [87]. Friedman, J.M., Epidemiology of neurofibromatosis type 1. Am J Med Genet, 1999. 89(1): p. 1-6.
801 [88]. Schindeler, A. and D.G. Little, Recent insights into bone development, homeostasis, and repair in
802 type 1 neurofibromatosis (NF1). Bone, 2008. 42(4): p. 616-622.
803 [89]. Banchhor, H. and V. Chimurkar, Congenital Pseudoarthrosis of the Tibia: A Narrative Review.
804 Cureus, 2022. 14(12): p. e32501.
805 [90]. Ru, J.Y. and Y.F. Wang, Osteocyte apoptosis: the roles and key molecular mechanisms in
806 resorption-related bone diseases. Cell Death Dis, 2020. 11(10): p. 846.
807 [91]. Wang, S., et al., The Role of Autophagy and Mitophagy in Bone Metabolic Disorders. International
808 Journal of Biological Sciences, 2020. 16(14): p. 2675-2691.
809 [92]. Tang, A.D., et al., Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic
810 leukemia reveals downregulation of retained introns. Nature Communications, 2020. 11(1): p. 1438.
811 [93]. Haas, B.J., et al., De novo transcript sequence reconstruction from RNA-seq using the Trinity
812 platform for reference generation and analysis. Nature Protocols, 2013. 8(8): p. 1494-1512.
813 [94]. Zhou, Y., et al., Metascape provides a biologist-oriented resource for the analysis of systems-level
814 datasets. Nat Commun, 2019. 10(1): p. 1523.
815 [95]. Chen, Y., et al., SOAPnuke: a MapReduce acceleration-supported software for integrated quality
816 control and preprocessing of high-throughput sequencing data. Gigascience, 2018. 7(1): p. 1-6.
817 [96]. Li, H. and R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform.
818 Bioinformatics, 2009. 25(14): p. 1754-60.
819 [97]. McKenna, A., et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
820 generation DNA sequencing data. Genome Res, 2010. 20(9): p. 1297-303.
821 [98]. McLaren, W., et al., The Ensembl Variant Effect Predictor. Genome Biol, 2016. 17(1): p. 122.
822 [99]. Zhan, X., et al., RVTESTS: an efficient and comprehensive tool for rare variant association analysis
823 using sequence data. Bioinformatics, 2016. 32(9): p. 1423-1426.

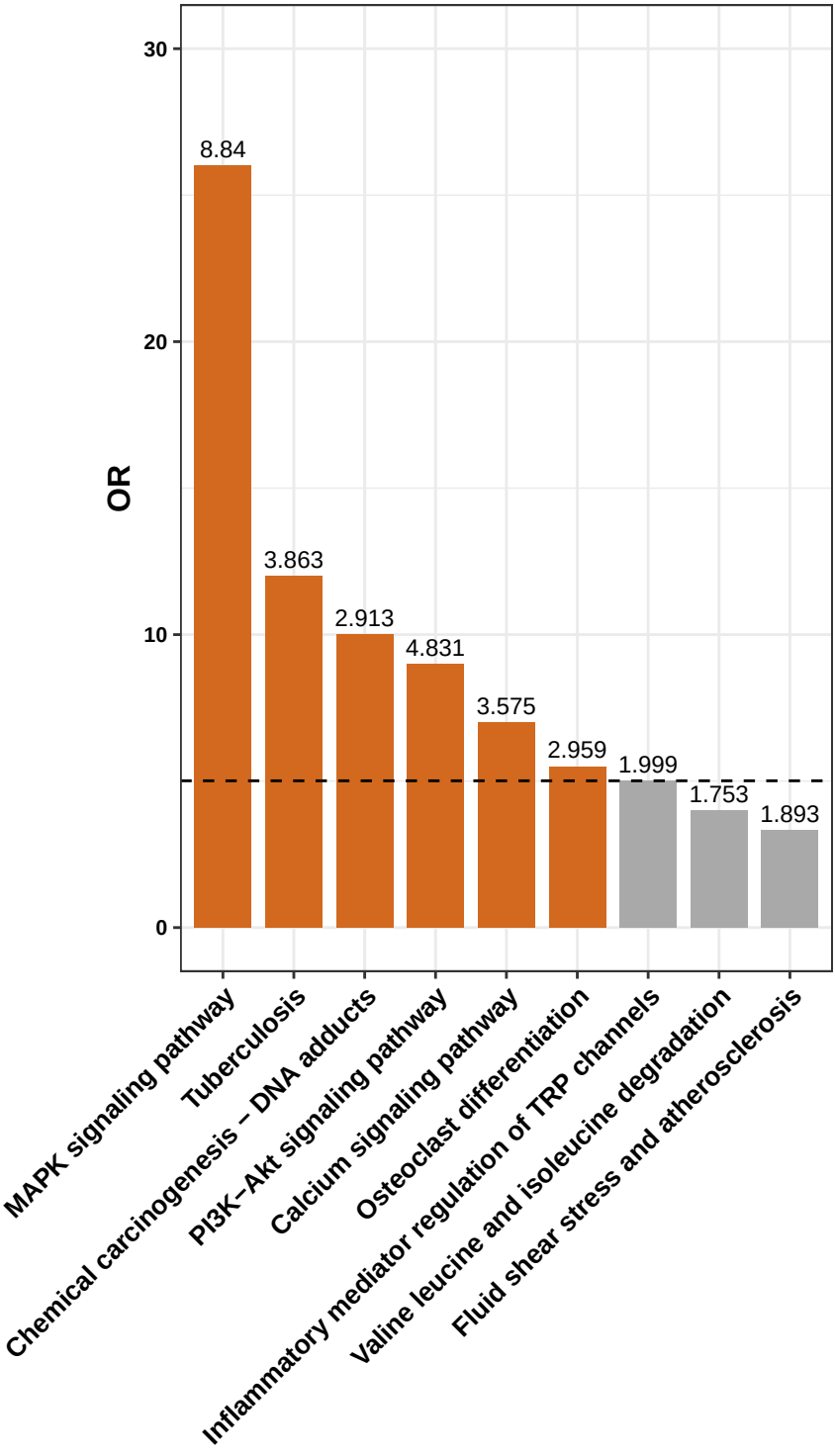
824

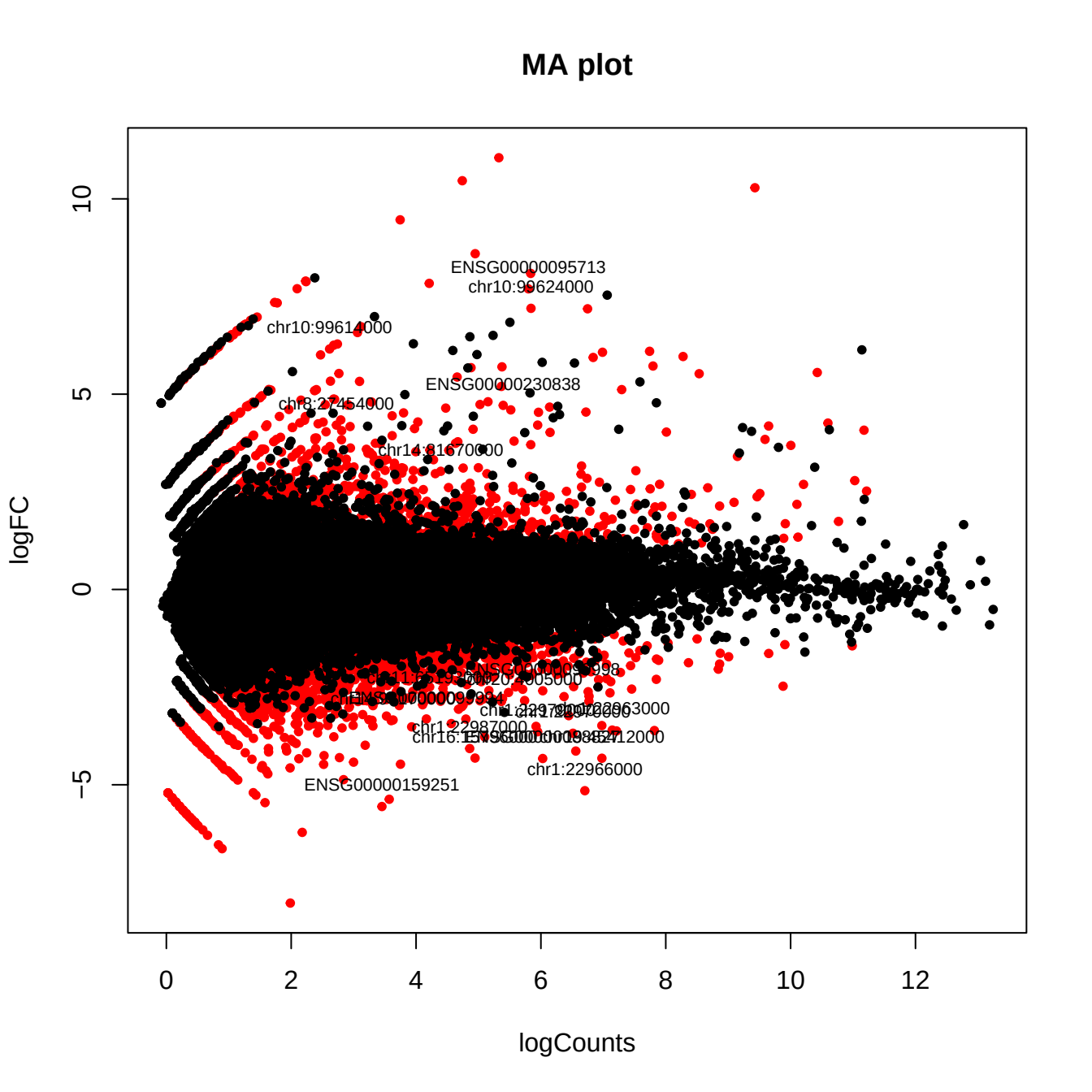


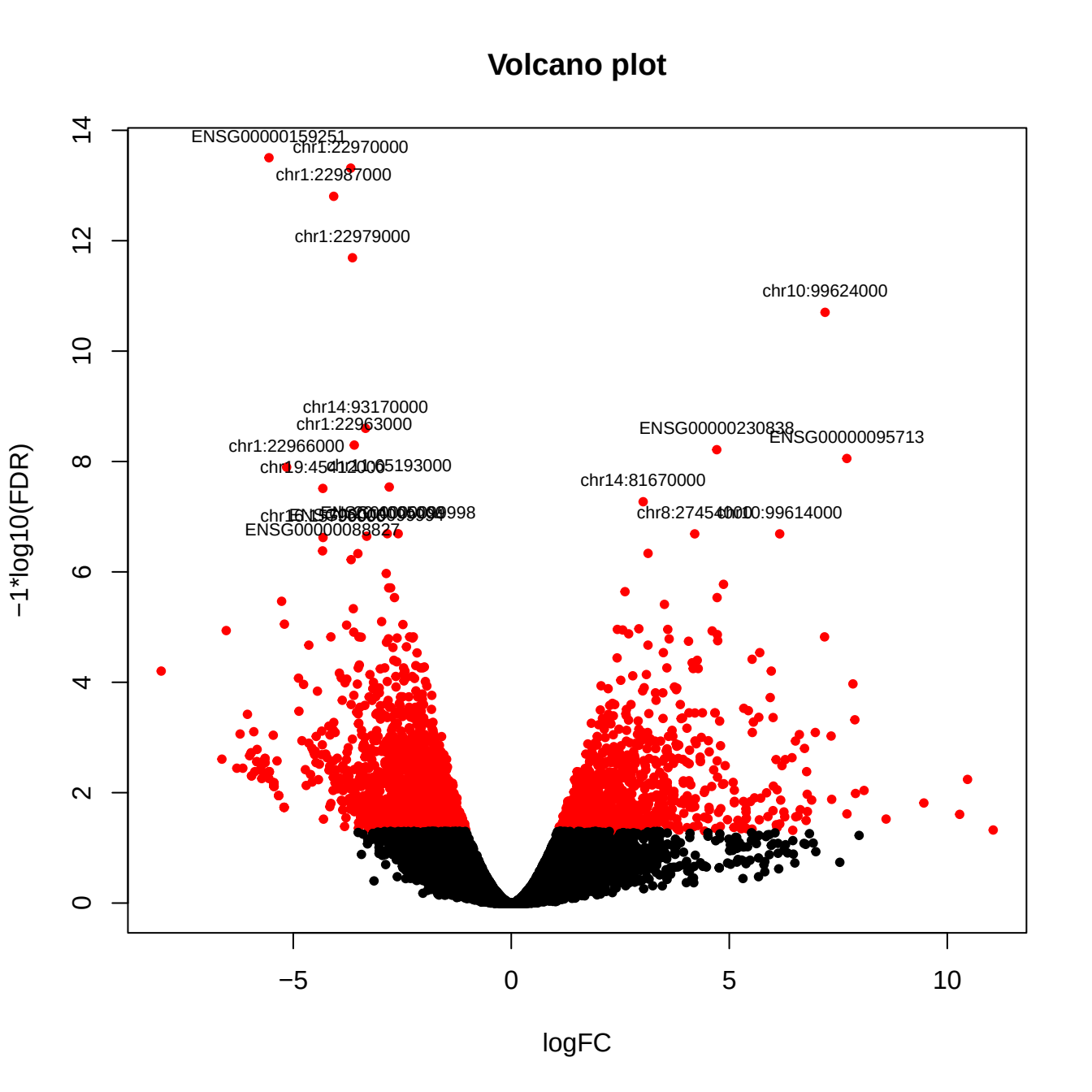
GO:0001501: skeletal system development
GO:0048771: tissue remodeling
GO:0043408: regulation of MAPK cascade
GO:0042129: regulation of T cell proliferation
GO:0002573: myeloid leukocyte differentiation
GO:0001667: ameboidal-type cell migration
GO:2001204: regulation of osteoclast development
GO:0001503: ossification
GO:0032102: negative regulation of response to external stimulus
GO:0048066: developmental pigmentation
GO:1900744: regulation of p38MAPK cascade
GO:0040013: negative regulation of locomotion
GO:0010817: regulation of hormone levels
GO:0006954: inflammatory response
GO:0060384: innervation
GO:0000902: cell morphogenesis
GO:0030198: extracellular matrix organization
GO:0001558: regulation of cell growth
GO:0050801: monoatomic ion homeostasis
GO:0061037: negative regulation of cartilage development
GO:0010720: positive regulation of cell development
GO:0034113: heterotypic cell-cell adhesion
GO:0007492: endoderm development
GO:0051051: negative regulation of transport
GO:0048871: multicellular organismal homeostasis
GO:0019724: B cell mediated immunity
GO:0051345: positive regulation of hydrolase activity
GO:0001819: positive regulation of cytokine production
GO:0032496: response to lipopolysaccharide
GO:0043269: regulation of monoatomic ion transport
GO:0034330: cell junction organization
GO:0051098: regulation of binding
GO:0034110: regulation of homotypic cell-cell adhesion
GO:0060390: regulation of SMAD protein signal transduction
GO:0071634: regulation of transforming growth factor beta production
GO:0007584: response to nutrient
GO:0007566: embryo implantation
GO:0045124: regulation of bone resorption
GO:0006914: autophagy
GO:0120162: positive regulation of cold-induced thermogenesis
GO:0007204: positive regulation of cytosolic calcium ion concentration
GO:0030282: bone mineralization
GO:0051091: positive regulation of DNA-binding transcription factor activity
GO:0035051: cardiocyte differentiation
GO:0009725: response to hormone
GO:0031334: positive regulation of protein-containing complex assembly
GO:2000027: regulation of animal organ morphogenesis
GO:0009135: purine nucleoside diphosphate metabolic process
GO:0098660: inorganic ion transmembrane transport
GO:0008037: cell recognition

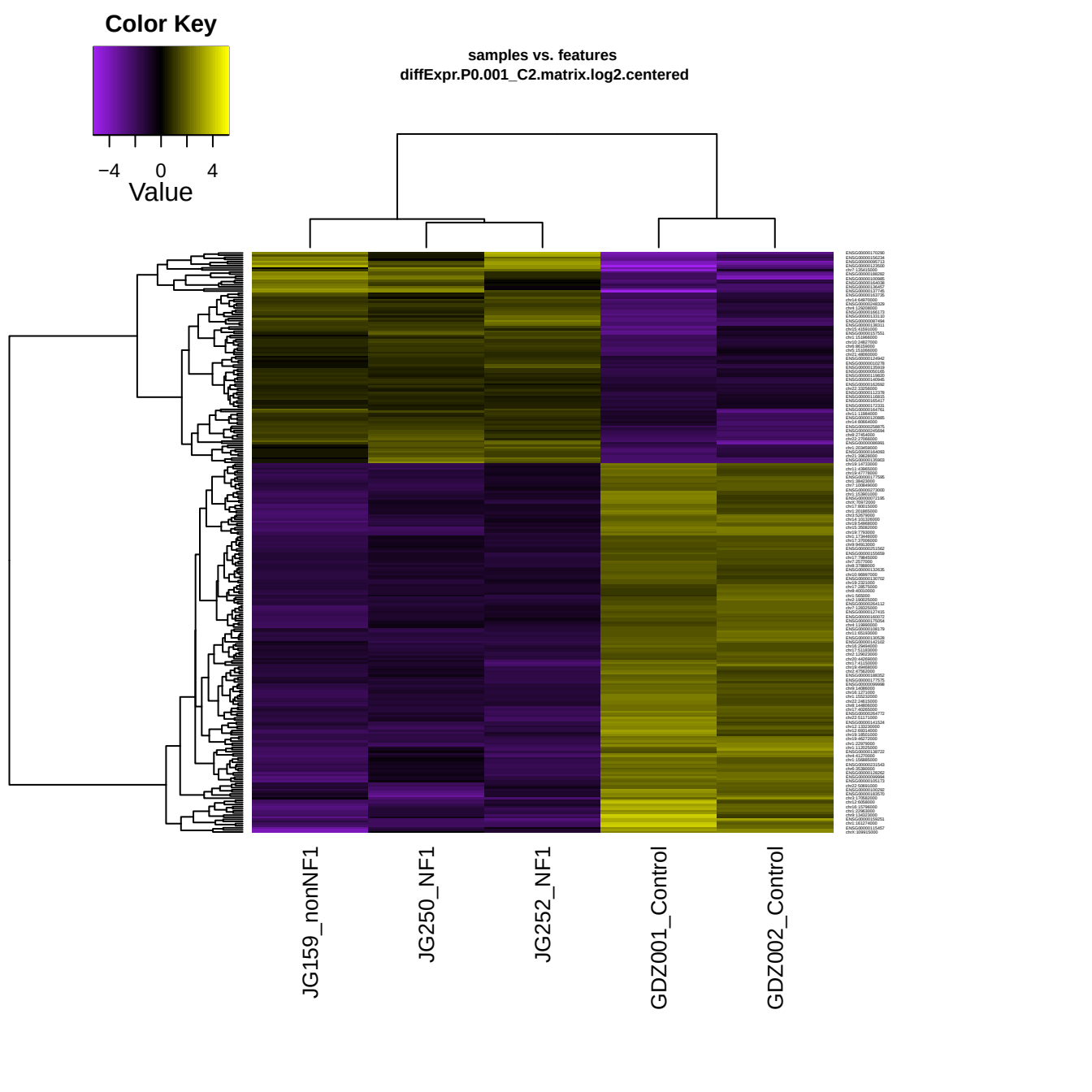

OR
30
20
10
0
11.55
11.25
9.86
9.86
8.5
7.83
10.49
5.86
5.54
5.54
4.82
4.18
4.89
7.76
10.53
6.44
6.44
3.23
3.93
6.12
3.93
9.56
9.22
4.51
2.97
10.34
4.07
3.77
1.02
1.33
0.7
0.98
0.98
0.98
0.35
0.6
regulation of MAPK cascade
monoatomic ion homeostasis
regulation of hormone levels
inflammatory response
positive regulation of cytokine production
ossification
regulation of monoatomic ion transport
myeloid leukocyte differentiation
regulation of bone resorption
regulation of cell growth
response to lipopolysaccharide
ameboidal−type cell migration
regulation of binding
negative regulation of locomotion
negative regulation of transport
cell morphogenesis
skeletal system development
regulation of T cell proliferation
response to nutrient
negative regulation of response to external stimulus
B cell mediated immunity
extracellular matrix organization
multicellular organismal homeostasis
autophagy
positive regulation of cold−induced thermogenesis
positive regulation of hydrolase activity
positive regulation of cell development
cell junction organization
heterotypic cell−cell adhesion
endoderm development
tissue remodeling
regulation of homotypic cell−cell adhesion
regulation of p38MAPK cascade
regulation of SMAD protein signal transduction
developmental pigmentation
innervation


hsa04380: Osteoclast differentiation
hsa05152: Tuberculosis
hsa05418: Fluid shear stress and atherosclerosis
hsa05204: Chemical carcinogenesis - DNA adducts
hsa04010: MAPK signaling pathway
hsa04020: Calcium signaling pathway
hsa04151: PI3K-Akt signaling pathway
hsa00280: Valine, leucine and isoleucine degradation
hsa04750: Inflammatory mediator regulation of TRP channels
0
1
2
3
4
5
-log10(P)


OR
30
20
10
0
8.84
3.863
2.913
4.831
3.575
2.959
1.999
1.753
1.893
MAPK signaling pathway
Tuberculosis
Chemical carcinogenesis – DNA adducts
PI3K−Akt signaling pathway
Calcium signaling pathway
Osteoclast differentiation
Inflammatory mediator regulation of TRP channels
Valine leucine and isoleucine degradation
Fluid shear stress and atherosclerosis


MA plot
logFC
logCounts
10
5
0
−5
0
2
4
6
8
10
12
ENSG00000095713
chr10:99624000
chr10:99614000
ENSG00000230838
chr8:27454000
chr14:91670000
chr1:22966000
ENSG00000159251


Volcano plot
ENSG00000159251
chr1:22970000
chr1:22987000
chr1:22979000
chr10:99624000
chr14:93170000
chr1:22963000
chr1:22966000
chr14:81670000
ENSG00000230838
ENSG00000095713
chr8:27454001
chr10:99614000
ENSG00000088827
−1*log10(FDR)
logFC
0
2
4
6
8
10
12
14
−5
0
5
10


Color Key
−4
0
4
Value
samples vs. features
diffExpr.P0.001_C2.matrix.log2.centered
JG159_nonNF1
JG250_NF1
JG252_NF1
GDZ001_Control
GDZ002_Control