COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
JOURNAL

Identification of the genetic basis of sporadic polydactyly in China by targeted sequencing

Bailing Zu [a,1], Xiaoqing Zhang [a,1], Yunlan Xu [b,1], Ying Xiang [a], Zhigang Wang [b], Haiqing Cai [b], Bo Wang [a,*], Guoling You [c,*], Qihua Fu [a,*]

[a] Pediatric Translational Medicine Institute, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[b] Department of Pediatric Orthopedic, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[c] Department of Laboratory Medicine, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

ABSTRACT

Purpose: Polydactyly is a highly heterogeneous group of skeletal deformities in clinical and genetic background. The variation spectrum in Chinese sporadic polydactyly has not been comprehensively analyzed. To elucidate genetic variation spectrum and genotype-phenotype correlations in Chinese patients with polydactyly, we conducted comprehensive genetic analysis of patients nationwide using targeted sequencing.
Methods: A total of 181 patients diagnosed with polydactylies were recruited. We designed a targeted capture panel for sequencing 721 genes that are associated with the pathogenesis of skeletal dysplasia. We performed rigorous variant- and gene-level filtrations to identify potentially damaging variants, followed by enrichment analysis and gene prioritization.
Results: A total of 568 deleterious variants of 293 genes were identified in 173 of 181 patients with a positive rate of 95.6% by targeted sequencing. For each sample, an average of 3.17 deleterious variants were identified. Especially, 14 pathogenic or likely pathogenic variants were identified in 10 genes in 14 patients out of the 181 patients, providing a positive molecular diagnostic rate of 7.7%.
Conclusion: Targeted sequencing analysis provides a high efficiency approach for the genetic diagnosis of polydactyly. This is the largest next generation sequencing study performed to date in patients with polydactyly and represents the genetic basis of polydactyly typically encountered in genetics clinics.

1. Introduction

Polydactyly is a digital deformity with additional digits in hands and/or feet, which is one of the most frequently observed limb malformations [1]. The prevalence of polydactyly is estimated to be 0.3–3.6/1000 in live-births, with higher incidence rate for male than female [1,2]. According to the Chinese birth defect monitoring report in 2012, the overall incidence of polydactyly was 16.74/10000 in live births, accounting for the second most common birth defects in China [3].

Polydactyly is a highly clinically heterogeneous skeletal deformity. A single patient's phenotype may be symmetrical or asymmetrical, preaxial or postaxial, non-syndromic or syndromic, unilateral or bilateral, involving distinct metacarpals or phalanxes [4]. Polydactyly is commonly observed as the isolated appearance, which usually is unilateral and asymmetrical, while syndromic polydactyly could be bilateral and symmetrical [5]. The classification scheme of Temtamy and McKusick is being widely used nowadays to classify polydactyly patients into pre-axial, post-axial, and complex types [1]. Pre-axial polydactyly includes four subtypes: thumb/hallux polydactyly (type I, PPD1), polydactyly of triphalangeal thumb (type II, PPD2), polydactyly of index finger (type III, PPD3), and polysyndactyly or crossed polydactyly (type IV, PPD4). Sequence variants in *GLI1* , *pre-ZRS* and *ZRS*, involved in Hedgehog pathway, may result in PPD1[6], and sequence variants in *ZRS, LMBR1, PTCH1*, and *SHH* can cause PPD2 [7]. Besides, *ZRS* and *GLI3* variations are regarded as be involved in PPD4 formation [7]. Post-axial polydactyly is generally divided into post-axial type A (PAPA) and post-axial type B (PAPB), and PAPA is further categorized into eight subtypes: PAPA1 ~ PAPA7 and PAP8 with

* Corresponding authors.
E-mail addresses: wangbo@scmc.com.cn (B. Wang), youguoling@126.com (G. You), qihuafu@126.com (Q. Fu).
[1] The first three authors contributed equally to this work, and should be considered as co-first author.

or without Ellis–van Creveld syndrome (EVC) [7]. Gene variants of Hedgehog and cAMP signaling pathways or cilia formation and function, including *GLI3, IQCE, PTCH1, FAM92A, KIAA0825, ATP6V1B1, CREBBP, EP300* and *BBS10*, have been reported to be related with PAPA [8–12].

The primary etiology of polydactyly appears to be genetic [13]. The genetics of polydactyly is highly heterogeneous, and not only restricted to Mendelian inheritance. Most familial polydactyly are inherited in an autosomal dominant trait with variable penetrance, but autosomal recessive and X-linked inheritance patterns have also been reported [14–16]. Generally, phenotypes of autosomal dominant polydactyly are less severe with variable penetrance and incomplete expressivity, which are mainly caused by gene variants associated with the anterior-posterior (AP) patterning of limb development. Genetic diagnosis of polydactyly plays an important role in risk stratification of potentially life-altering skeletal dysplasia-related disease and recurrence risk prognostication after surgical intervention for the affected individual. Polydactyly is more than a phenotype of skeletal dysplasia, and it may indicate a systemic congenital disorder or connective tissue disease. For example, short rib polydactyly syndrome type II (SRPS) is a genetic disorder whose visible phenotypes include a narrow chest and polydactyly, but most of SRPS II patients do not survive for long because of their kidney, heart and blood vessel dysfunction. Genetic diagnosis of polydactyly enable clinicians to identify life-altering skeletal dysplasia-related disease. In spite of the significant progress of medical genetics in recent years, molecular diagnosis of polydactyly can still be challenging. A study through whole exome sequencing has been performed in a relatively small sample with preaxial polydactyly, with a diagnostic yield of approximately 15% [17].

Since targeted sequencing provides a more complete picture of disease-associated genes from one clinical sample, targeted sequencing of gene panels has been cost-effective diagnostic workup for heterogeneous disorders. In our study, we performed targeted sequencing analysis of 181 cases with polydactyly through a panel containing 721 genes, which were reported to be associated with skeletal dysplasia. We identified pathogenic or suspected diagnostic variants in 14 of the 181 patients, providing a positive molecular diagnostic rate of 7.7%. Our results provide a preliminary overview of the genetic etiology of polydactyly based on the largest cohort of polydactyly patients reported to date.

2. Materials and methods

2.1. Patients and collection of clinical information

A total of 181 consecutive sporadic patients were enrolled in our study from 2010 to 2018 in Shanghai Children's Medical Center (Table S1). Patients with identified chromosomal abnormity were excluded. X-ray photographs of hands, feet and chest of all the included individuals were acquired for phenotype diagnosis based on Temtamy and McKusick.

2.2. Design of skeletal dysplasia-related gene panel

We used a customized capture panel of 721 targeted genes. All these genes are associated with skeletal disorders (e.g., polydactyly, syndactyly, polysyndactyly, brachydactyly) (Table S2). Most of these genes were stemmed from the 2015 revision of Nosology and Classification of Genetic Skeletal Disorders and other genes were selected through a systematic review of public databases such as PubMed, OMIM, HGMD, HPO and mouse genomes informatics (MGI). For probe design, the parameters were set to

cover all the exonic regions and their flanking intron(+/- 10 bp) sequences of 721 genes. 40,582 capture probes, with a 99.4% predicted coverage of all the selected regions, were designed by Sure-Design (Agilent, Inc.).

2.3. Library preparation and targeted sequencing

Genomic DNA was extracted from peripheral blood samples using a QIAamp DNA Blood Mini Kit (Qiagen, Germany). A spectrophotometer (NanoDrop, USA) was used to determine the purity and concentration of the gDNA in the sample. The gDNA libraries were prepared by using Agilent Custom SureSelect Enrichment Kit (Agilent, USA) in accordance with the manufacturer's protocol. The capture library was sequenced on a HiSeq2500 Sequencer (Illumina, USA) with 2 × 150 bp paired-end reads.

2.4. Variant annotation and filtering

Clean sequence reads were aligned to the human reference genome (hg38) with BWA-MEM [18] and further processed using the GATK best practice workflows [19], which included duplication marking, indels realignment, and base quality score recalibration. Single nucleotide variants (SNVs) and small insertions/deletions (indels) were called with GATK Haplotype Caller in GVCF mode, and recalibrated by GATK Variant Quality Score Recalibration (VQSR) to reduce likely false-positive variants. The identified variants were annotated using ANNOVAR [20]. The preliminary filter process was performed to filter for rare (VAF $\leq 10^{-3}$ in 1000 Genomes, and gnomAD database) and high quality (pass GATK VQSR, minimum 10 total reads, GQ score $\geq$ 20, and a minimum 20% alternate allele ratio if alternate allele reads $\geq$ 10 or, if alternate allele reads is < 10, a minimum 30% alternate ratio) variants. The MetaSVM algorithm, annotated using dbNSFP [21], was used to predict deleteriousness of missense variants (D-Mis) [22]. Only LoF variants (nonsense, canonical splice-site, frameshift indels and start loss), D-Mis and non-frameshift indels were considered as potentially damaging to the disease. The potentially deleterious variants were further classified into pathogenic and likely pathogenic variants based on ACMG guidelines [23] for sequence variant interpretation.

2.5. Function, pathway, and network analysis

Gene Ontology analysis and KEGG pathway enrichment analysis of the complete list of genes which harbored LoF and damaging variants were performed using clusterProfiler package in R software, using *p* value < 0.05 as the cut-off criterion. To survey disease associations of these candidate genes, we carried out Disease Ontology (DO) and Disease Gene Network (DisGeNET) enrichment analysis with the R package DOSE [24], which provides semantic similarity computations among DO terms and genes, and allows biologists to explore the similarities of diseases and of gene functions in disease perspective. Protein-protein interactions (PPI) networks were constructed using Cytoscape software [25] based on the STRING database [26]. As nodes with a high degree of connectivity contribute more to the stability of the network, genes with a degree connectivity of greater than 20.0 were identified as hub genes by using the plugin CentiScaPe [27]. Molecular Complex Detection (MCODE) plug in [28] was applied to identify significant subnetworks with node score cutoff = 0.2, K-Core = 4, max depth = 100 and degree cut-off greater than 4 set as the cut-off criteria.

3. Results

3.1. Patient characteristics

In order to get an insight into phenotypic manifestations and genetic etiology, we enrolled a cohort of 181 sporadic probands with polydactyly. Among them, 104 (57.5%) were males and 77 (42.5%) were females, and the median age was 10 month old (range 5 months to 11.5 years). The demographic and basic clinical phenotype characteristics are shown in Table S1. Pre-axial polydactyly was the most common type (79.6%), and the proportion of post-axial type and complex polydactyly were 17.7% and 2.8%, respectively. In the pre-axial polydactyly, PPD1 was the most prevalent subtype (92.4%), PPD2 was diagnosed in 7.6% cases, while PPD3 and PPD4 were not found. Majority of the polydactyly cases were unilateral (86.2%). As shown in Table 1, involvement of upper/lower and right/left extremity was highly variable across the 181 established polydactyly cases.

3.2. Capture-based targeted sequencing

For each sample undergoing capture-based targeted sequencing, the average coverage was 272.49 X, sufficient depth to interrogate the sequence for variations. The percentage of interested regions with coverage greater than 10X was above 95.6% and with coverage greater than 20X was above 91.6%. An average of 97.7% of reads had a Phred-like quality score (Q score) of greater than 20 and 94.7% of reads had a Q score of greater than 30. At least 99.16% of targeted regions were covered by at least one read. These results suggested that capture-based targeted sequencing is sufficient to yield high-quality data for further analysis.

3.3. Genetic spectrums of polydactyly

After quality control and filtering process, 568 rare variants spanning 293 genes were identified to be deleterious in 173 patients, including 35 splicing variants, 28 nonsense variants, 16 frameshift insertions, 27 frameshift deletions, 13 non-frameshift insertions, 15 non-frameshift deletions and 434 missense variants (Fig. 1A, Table S3). For each sample, an average of 3.14 deleterious variants were identified, which indicated polydactyly as a polygenic disorder. Of 293 mutated genes, FLNB and FBN2 was the most frequently mutated genes, and 132 genes showed recurrent variations (Fig. 1B, Table S4).

To establish genotype-phenotype correlation, we examined the overlapped genes identified in PPD1, PPD2, PAPA, PAPB and complex polydactyly involve upper/lower or right/left extremity (Fig. 2, Fig. S1). The analysis showed that 27 genes were shared between PPD1 and PPD2; 18 genes were shared between PAPA and PAPB, 3 genes (TCIRG1, FLNB, TBXAS1) were shared across four types of polydactyly (PPD1, PPD2, PAPA, and PAPB) (Table S5).

We curated a set of 7 human non-syndromic polydactyly (H-nsynPD) genes and loci, 222 human syndromic polydactyly (H-synPD) genes from OMIM and published data, and human orthologs of 96 additional mouse polydactyly (M−PD) genes in MGI. Thus, 325 known polydactyly (KnownPD) genes and 396 genes of other subtypes of skeletal dysplasia (H-OS-SD) were included for the custom capture panel (Table S6). In our study, we identified 161 heterozygous variants in 83H-synPD genes as rare deleterious variants, including 53 variants in 26 dominant H-synPD genes, 90 variants in 51 recessive H-synPD genes, and 6 variants in 3 X-linked H-synPD genes. Apart from the above variants, we also identified 99 deleterious heterozygous variants in 56 dominant H-OS-SD genes, and 119 variants in 64 human orthologs of M−PD genes (Table S3).

Especially, 14 variations in 10 genes were further classified as pathogenic or likely pathogenic in 14 patients, providing a positive molecular diagnostic rate of 7.7% (Table 2). Among them, we identified four patients with pathogenic or likely pathogenic variants in the same gene. GLI3 is intracellular mediators of hedgehog signaling, which modulates limb bud development. Four deleterious variations were identified in GLI3 gene of four patients, including two splicing variations (c.1497 + 1G > C, c.2104-3C > A), one frameshift deletion (p.Gly552ValfsTer9) and one missense variation (p. Met948Ile). Patient N355 was detected with both GLI3 frameshift variations and FGFR3 frameshift variations, whose pre-axial polydactyly phenotype involves both upper/lower and right/left extremity (Table 2).

3.4. Gene enrichment and pathway analysis

To achieve a deeper understanding of the candidate 293 genes in cases, we performed GO function and KEGG pathway enrichment analysis using the ClusterProfiler package as implemented in R (Table S7). In the BP ontology, the candidate 293 genes were significantly enriched in two major categories: skeletal system development (limb morphogenesis, ossification, and skeletal system morphogenesis) and cilia function (cilium organization, cilium assembly, and non-motile cilium assembly) (Fig. 3A); In the CC ontology, cilia items constitute the majority of enriched GO categories, including ciliary part, ciliary basal body, ciliary membrane and photoreceptor connecting cilium (Fig. 3B). In the MF ontology, the binding-related items constitute the majority of enriched GO categories, including proximal promoter sequence-specific DNA binding, RNA polymerase II transcription factor binding, chromatin binding, growth factor binding (Fig. 3C). The enriched GO terms of the candidate 293 genes were similar with that of 325 polydactyly associated genes and 721 skeletal associated genes (Fig. 4). KEGG pathway enrichment showed that several pathways were related to skeletal system development, such as Hedgehog signaling pathway, Wnt signaling pathway, TGF-beta signaling pathway, and PI3K-Akt signaling pathway (Fig. 3D). Hedgehog signaling pathway plays a major role in specifying the antero-posterior pattern of the digits [29]. The initiation of limb bud development requires the interactions of the retinoic acid (RA), FGF, and WNT signaling pathways, and TBX transcription factors [30]. PI3K/AKT pathways regulate limb blastema formation [31]. The pathway analysis suggested that these pathways might be critical in the development of polydactyly.

To identify disease associations of these candidate genes, we carried out DO and DisGeNET enrichment analysis with the R package DOSE. For DO enrichment, the candidate genes were significantly enriched in bone development disease and polydactyly (Fig. 5A). For DisGeNET enrichment, 18 genes were reported in DisGeNET as associated to polydactyly and 22 genes were enriched in skeletal dysplasia, 18 genes were reported in DisGeNET as associated to syndactyly (Fig. 5B, Table S8).

To investigate the relationship of the candidate 293 genes between modules, PPI networks were constructed using Cytoscape software based on the STRING database. In total, 283 nodes and 1672 PPI relationships were obtained (Fig. 6). 51 genes with a degree of connectivity ≥ 20 were defined as hub genes (Table S9) for polydactyly. As the most intensive hub gene, SHH interacts with 74 genes. For the 51 hub genes, we identified 119 rare variations (Table S3). Of these 51 hub genes, GO analysis showed that the most significant GO biological process term found in PPI was limb morphogenesis (22 genes, p.adjust = 1.32E-33), followed by smoothened signaling pathway (20 genes, p.adjust = 7.00E-31), 18 genes were associated with cilium assembly and organization, 24 genes were associated with ciliary part, basal body, membrane, 9 genes were associated with chromatin binding (Table S10). These

Table 1
Distribution of affected limbs in polydactylies.

Phenotype	Male	Female	Unilateral	Bilateral	Symmetrical	236 limbs						
						LH	RH	BH	LF	RF	BF	BHF
Preaxial-1	81	52	123	10	8	47	84	6	7	5	3	1
Preaxial-2	6	5	11	0	0	8	4	0	0	1	0	0
Postaxial-A	8	12	13	7	5	4	3	2	13	12	5	0
Postaxial-B	7	5	6	6	2	4	3	3	5	6	1	0
Complex	2	3	3	2	2	1	2	1	3	1	1	0

LH, left hand; RH, right hand; BH, both hands; LF, left foot; RF, right foot; BF, both feet; BHF, both hands and feet.
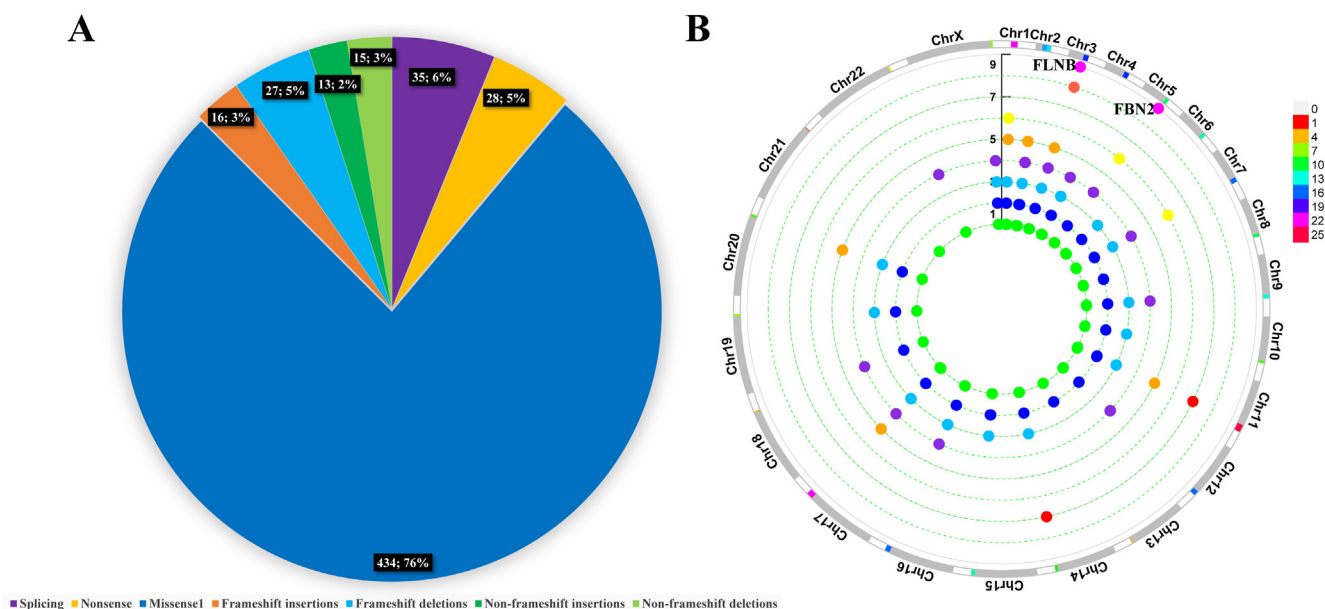
Fig. 1. **Summary of rare variants identified in 181 probands with polydactyly.** A, summary of rare variants; B, summary of rare variants identified by genes.
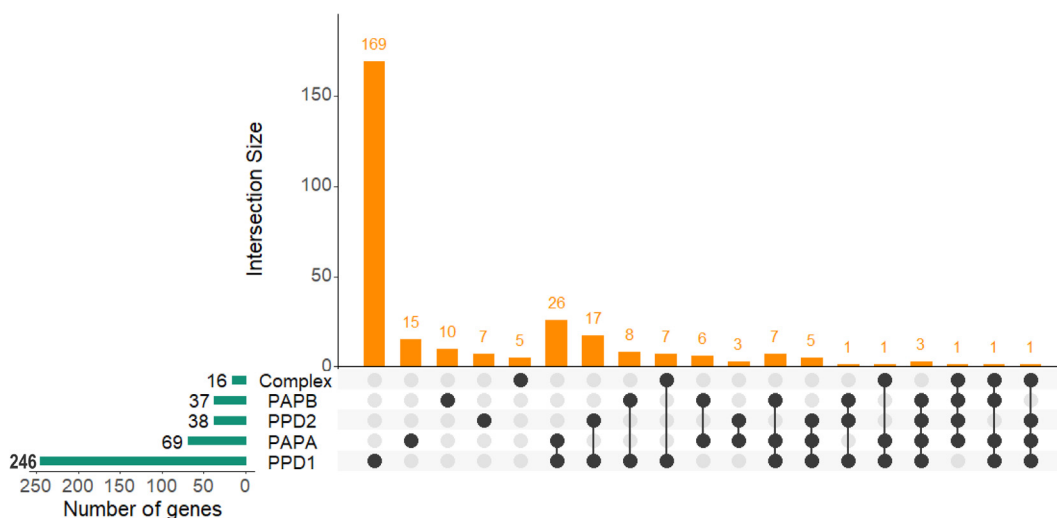

Fig. 2. An UpSetR plot of gene across five types of polydactyly.

hub genes might play an important role in polydactyly and should be further studied. A total of 7 subnetworks from the PPI satisfied the criteria of an MCODE computed node score >4 and number of nodes >4 (Fig. 6). 100 MCODE related genes were identified and

which contained the 39 hub genes (Table S11). Notably, 67 genes are known polydactyly-causing genes. Subnetwork 1 with the highest score included 24 nodes and 220 PPI pairs, the network genes were found associated with cilium cilium assembly, cilium

Table 2
Summary of causative variations identified in the investigated cases.

| No.	Sample	Variation type	Gene	AAChange	Protein	Phenotype	OMIM	Classification
1	X62	Missense	*ALX4*	NM_021926.3:c.19G > A	p.Val7Ile	RF-PPD1	605,420	Likely pathogenic
2	B66	Nonsense	*BMP4*	NM_001347913.1:c.777G > A	p.Trp259Ter	RF-PAPB/LF-PAPA	112,262	Pathogenic
3	B116, X384	Frameshift insertion	*CD96*	NM_001318889.1:c.54dupT	p. Val19CysfsTer22	RF-PPD2, LH-PPD1/RH-PPD1	606,037	Pathogenic
4	N319	Nonsense	*CREBBP*	NM_004380.2:c.1775G > A	p.Trp592Ter	RF-PPD1/LF-PPD1	120,140	Pathogenic
5	X565	Missense	*FGFR1*	NM_001354369.1: c.1753C > T	p.His585Tyr	RH-PPD1	136,350	Likely pathogenic
6	N90	Missense	*FGFR1*	NM_001354369.1: c.1686G > C	p.Lys562Asn	LH-PPD1	136,350	Likely pathogenic
7	N355	Frameshift deletion	*FGFR3*	NM_022965.3:c.2072delG	p. Gly691AlafsTer17	LF-PPD1/RF-PPD1/RH-PPD1/ LH-PPD1	134,934	Pathogenic
8	X452	Splicing	*FUZ*	NM_001352262.1: c.233 + 1G > A		RF-PAPB	610,622	Likely pathogenic
9	N250	Splicing	*GLI3*	NM_000168.5: c.1497 + 1G > C		RF-PPD1/LF-PPD1	165,240	Pathogenic [33]
10	N355	Frameshift deletion	*GLI3*	NM_000168.5:c.1653delA	p. Gly552ValfsTer9	LF-PPD1/RF-PPD1/RH-PPD1/ LH-PPD1	165,240	Pathogenic
11	X197	Splicing	*GLI3*	NM_000168.5:c.2104-3C > A		RF-PAPA/LF-PAPA	165,240	Likely pathogenic
12	X343	Missense	*GLI3*	NM_000168.5:c.2844G > A	p.Met948Ile	RH-PPD1	165,240	Likely pathogenic [6]
13	B21	Missense	*NSD1*	NM_022455.4:c.6505A > G	p.Lys2169Glu	RH-PPD1	606,681	Likely pathogenic
14	N336	Splicing	*SETBP1*	NM_015559.2:c.4001-1G > T	–	RH-PPD1	611,060	Pathogenic
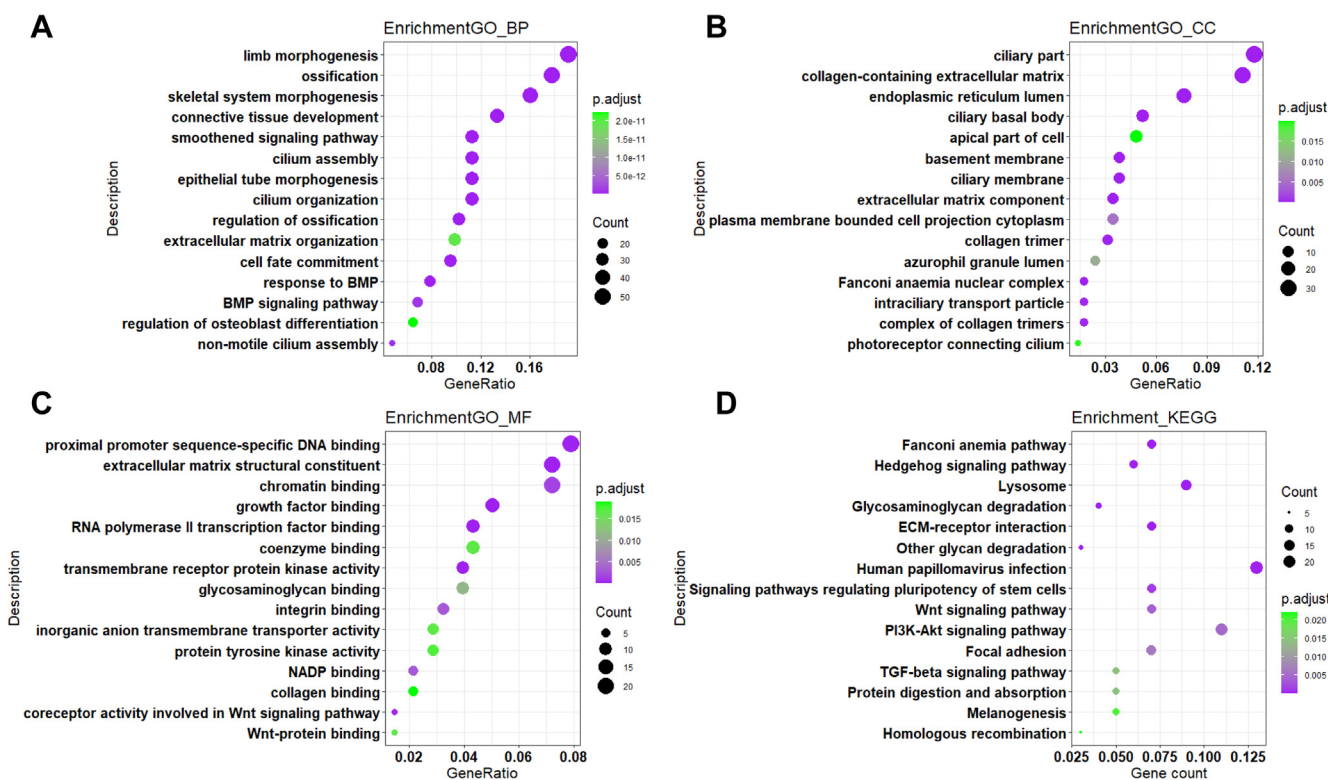
Fig. 3. Functional enrichment analyses of candidate genes. A, biological processes (BP) enrichment; B, cellular components (CC) enrichment; C, molecular functions (MF) enrichment; D, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

organization, and ciliary part. The second higthest score is subnetwork 2, which included 17 nodes and 96 PPI pairs and was associated with limb morphogenesis (p.adjust = 9.86E-08) (Table S12).

4. Discussion

Polydactyly is considered as skeletal dysplasia, while skeletal dysplasia contains a large group of heritable diseases of the cartilage or bone abnormalities. Moreover, some other subtypes of skeletal dysplasia may be with a polydactyly phenotype, e.g.,

short-rib thoracic dysplasia and sive skeletal dysplasia. Over time, it's being recognized that there are clinical and molecular overlaps among different subtypes of skeletal dysplasia. Hence, we designed a custom capture panel containing 721 skeletal dysplasia-associated genes for 181 polydactyly patients. Based on the population data, computational and predictive data, functional data and other database, 14 detected pathogenic or likely pathogenic variants were identified in 14 polydactyly patients, providing a positive molecular diagnostic rate of 7.7%. To the best of our knowledge, this is the largest study interrogating the genetic spec-
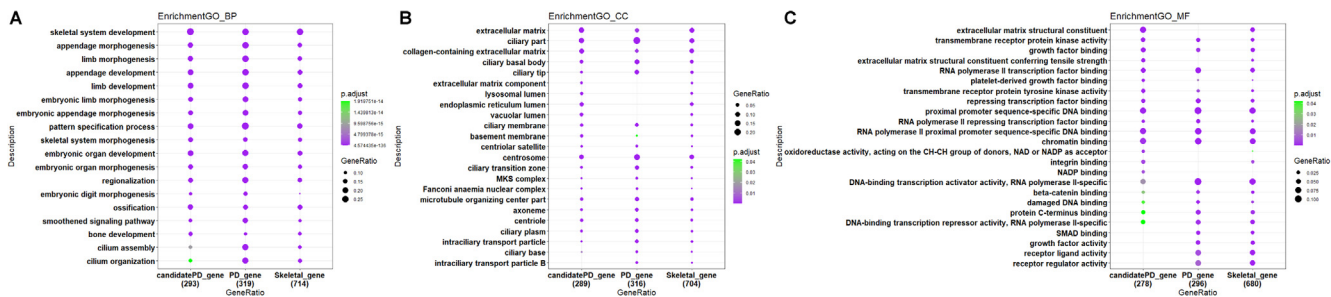
Fig. 4. The compared GO terms of the candidate 293 genes with that of 325 polydactyly associated genes and 721 skeletal associated genes. A, biological processes (BP) enrichment; B, cellular components (CC) enrichment; C, molecular functions (MF) enrichment.
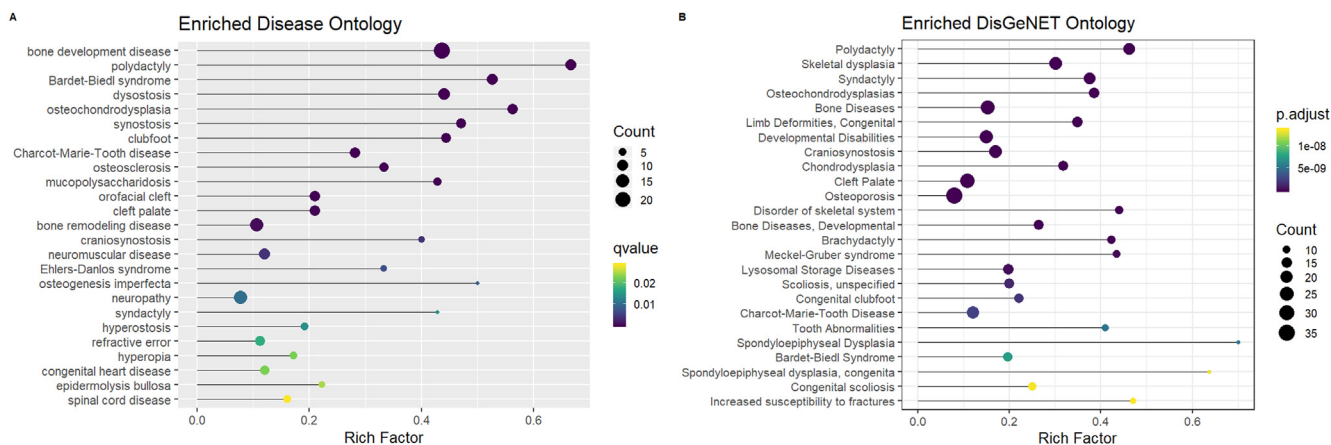
Fig. 5. Disease Ontology (DO) and Disease Gene Network (DisGeNET) enrichment analysis. A, DO enrichment analysis; B, DisGeNET enrichment analysis.

trum in patients with polydactyly. 568 deleterious variants of 293 genes may be implicated with sporadic polydactyly, which tend to provide foundation for further segregation analysis and specific experiments to verify the new disease-causing genes and genetic model. The further comprehensive analysis of deleterious variants will pave the way for more effective genetic diagnosis for polydactyly.

SHH, BMPs, FGFs and WNTs pathways have been demonstrated to have a substantial role in the development of limbs. The SHH-GLI3 activated pathway is highly conserved and essential for AP limb patterning, which is associated with polydactyly. The key genes in the SHH-GLI3 pathway include *GLI3*, *SHH*, *PTCH1*, *IQCE*, *SMO* and *BMP4*.

Variations in *GLI3* gene are known to be associated with several clinical disorders including pre-axial polydactyly, post-axial polydactyly, Greig cephalopolysyndactyly syndrome (GCPS) and Pallister-Hall syndrome (PHS) [32]. In our study, four deleterious variations in *GLI3* gene were identified, including a missense, a frameshift deletion and two splicing variations. Two novel LoF variations, locating in the zinc finger domain of *GLI3* (c.1653delA, c. 1497 + 1G > C), and one nonsynonymous variation (c.2844G > A) were identified in three patients with PPD1. One patient with PAPA was found carrying the splicing variation (c.2104-3C > A). Though c.1497 + 1G > C variation has been reported in probands with features of PHS or GCPS [33], none of them in our cohort had typical GCPS and PHS phenotypes, such as polysyndactyly, hypothalamic hamartoma, atresia and bifid epiglottis, craniofacial abnormalities, and developmental delay.

PTCH1 plays a critical role in SHH signaling, and loss of PTCH1 function usually results in elevated shh pathway activity. More generally, mutaions in *PTCH1* have been reported in Gorlin Syn-

drome, with primary clinical manifestations of polysyndactyly [34]. However, few non-syndromic polydactyly phenotype with *PTCH1* mutation was reported. In this cohort, three deleterious variations of *PTCH1* gene were identified in patients with PPD1 and PAPA phenotype.

Variations in *BMP4* were shown to be associated with ocular malformation, digital anomalies, and brain anomalies. Here, we identified two missense variations and a nonsense variation in the *BMP4* gene in three cases with three different polydactyly phenotypes, respectively. However, no additional phenotypes of other system were found in those patients. Our data might provide additional support for *BMP4* variations to be associated with non-syndromic polydactyly phenotype.

Synpolydactyly (SPD) is the second most frequent syndactyly type and is inherited in an autosomal dominant condition. Variations in *HOXD13* and *FBLN1* can lead to SPD. *HOXD13* encodes a transcription factor, which initiates Shh expression during the early phase and mediates the Shh morphogenic signal within the limb during the second phase. Haploinsufficiency of the *FBLN1* could lead to the observed limb malformations. In this study, we identified a missense *HOXD13* variation and a missense *FBLN1* variation in PAPA. Further follow-up and assessment will be needed for the final interpretation of these variants.

We identified several variations in syndromic polydactyly genes, such as *SALL1, ALX4, FUZ, GLI2, EVC, COL2A1, FGFR3, IHH, NSD1, SALL4, TBX1, TBX3* and *TP63*, in non-syndromic polydactyly cases, while no additional pathological features were observed in these probands. Further follow-up will be needed, since patients in this cohort, median age as 1.6 years old, can be too young to manifest developmental anomaly outcomes, even if a syndrome existed
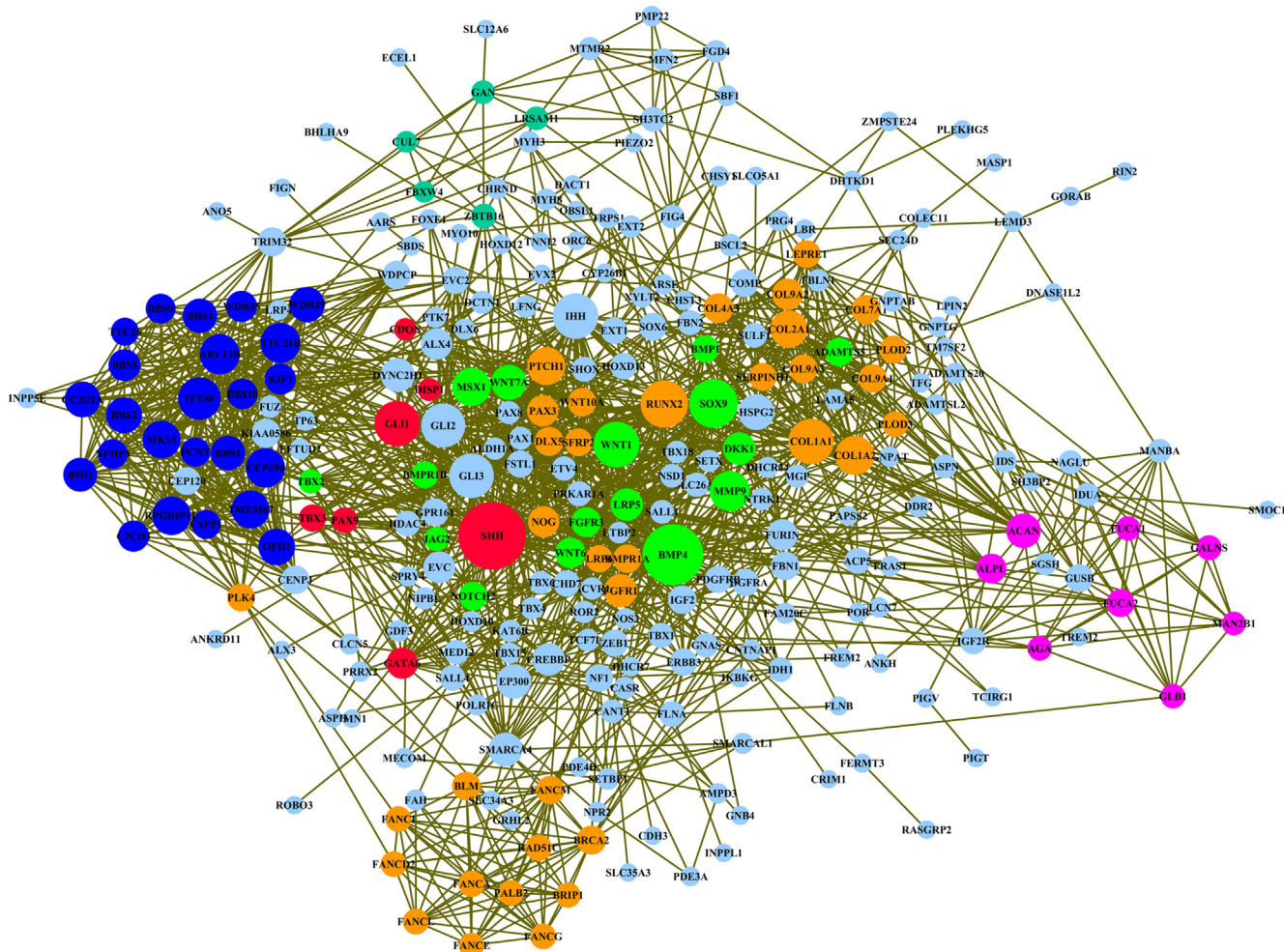
Fig. 6. Protein–protein interaction (PPI) network construction and subnetwork analysis. The blue circles represent subnetwork 1, the orange circles represent subnetwork 2, the green circles represent subnetwork 3, the pink circles represent subnetwork 4, the darkgreen circles represent subnetwork 5, the red circles represent subnetwork 6. The size of circles represent the degee of nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We also identified variations in human orthologs of mouse polydactyly genes, such as *DKK1, FBN2, SPRY4,* et al, and no polydactyly phenotypes by these genes have been reported in human before. However, a comprehensive in silico analysis of functional and structural impact for the coordinate gene variations pinpoints possible variational effects to support their pathogenic role. Our data tend to provide foundation for further specific experiments.

As a consequence of less evolutionary pressure, non-coding sequences allow for much more genetic variations than coding sequences, and thus pathogenic variants search from non-coding regions is extremely challenging given the vastness of noncoding genome variation. What's more, variations in non-coding sequences are much more difficult to interpret, because of the intrinsic difficulty in assessing the role of noncoding sequences and our still primitive understanding of noncoding sequences. Currently, clinical genomics analyses focus on protein-coding sequences. However, the pathogenic possibilities of variation in non-coding regions should not be overlooked. For example, *ZRS* enhancer has been proved for its pathogenic variations for polydactyly in human patients. *ZRS*, an enhancer regulating the expression of *SHH* gene, has been reported to be associated with pre-axial polydactyly and triphalangeal thumb (TPT) [35]. In this region,

more than 20 point variants, 10 duplications, 1 triplication, and a base pair insertion have been reported to be associated with PPD. In the present cohort, we identified a novel deleterious point variation at *ZRS* 382 T > C (g.156791524 T > C) in one patient with PPD1, and this is the second report identifying the genetic variation causing non-syndromic PPD1. The previously reported patient with *ZRS* 446 T > A variation showed duplicated thumb of distal phalanx, while the affected case in our study presented duplicated thumbs of proximal and distal phalanx.

PPI network analysis provided detailed interaction among the candidate 293 genes, 100 genes with a degree of connectivity >15 were defined as hub genes. The top 10 hub genes were *SHH, BMP4, SOX9, RUNX2, GLI1, GLI3, WNT1, IFT88, GLI2, IHH*. GO analysis showed that these hub genes are associated with limb development, cilium assembly and organization, which have been reported to have a substantial role in polydactly.

Our data will help to explain diverse clinical phenotypes and genetic heterogeneity of polydactyly by establishing the landscape of genetic variation basis by targeted high-throughput sequencing. Hub gene analysis has shown that polydactyly phenotypes are associated with a pathway network of limb development, cilium and chromatin. However, the additive and epistatic effect of multi-

ple genetic variants, SNP or rare variants, especially involving oligo- or polygenic genetic modifiers, hamper mechanistic dissection of genotype–phenotype correlation by bioinformatic analysis.

5. Conclusions

In summary, we identified 568 rare variants in 181 Chinese patients by targeted sequencing in the largest cohort of patients with polydactyly described to date. Our diagnostic rate of 7.7% demonstrates the effectiveness of targeted sequencing as a diagnostic tool for polydactyly. In addition, the results extended the gene variation spectrum of polydactyly. The identification of novel disease-causing genes and genetic model for polydactyly could not only be important to molecular diagnostics and genetic counseling, but also help to elucidate the limb patterning and digit specification mechanisms.

6. Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Shanghai Children's Medical Center. Every participant provided informed consent according to the Declaration of Helsinki.

CRediT authorship contribution statement

Bailing Zu: Data curation, Validation, Investigation, Writing - original draft. **Xiaoqing Zhang:** Data curation, Writing - original draft. **Yunlan Xu:** Data curation, Resources, Writing - review & editing. **Ying Xiang:** Data curation. **Zhigang Wang:** Data curation, Resources. **Haiqing Cai:** Data curation, Resources. **Bo Wang:** Data curation. **Guoling You:** Conceptualization, Methodology, Software, Investigation, Data curation, Writing - original draft. **Qihua Fu:** Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.06.014.

References

[1] Malik S. Polydactyly: phenotypes, genetics and classification. Clin Genet 2014;85(3):203–12.

[2] Castilla E, Paz J, Mutchinick O, Munoz E, Giorgiutti E, Gelman Z. Polydactyly: a genetic study in South America. Am J Hum Genet 1973;25(4):405–12.

[3] China Birth Defects Prevention Report (2012). In.: Beijing: The Ministry of Health issued; 2012: 1-24.

[4] Ahmed H, Akbari H, Emami A, Akbari MR. Genetic Overview of Syndactyly and Polydactyly. Plast Reconstr Surg Glob Open 2017;5(11):e1549. https://doi.org/10.1097/GOX.0000000000001549.

[5] Malik S, Ullah S, Afzal M, Lal K, Haque S. Clinical and descriptive genetic study of polydactyly: a Pakistani experience of 313 cases. Clin Genet 2014;85(5):482–6.

[6] Xiang Y, Jiang L, Wang Bo, Xu Y, Cai H, Fu Q. Mutational screening of GLI3, SHH, preZRS, and ZRS in 102 Chinese children with nonsyndromic polydactyly. Dev Dyn 2017;246(5):392–402.

[7] Umair M, Ahmad F, Bilal M, Ahmad W, Alfadhel M. Clinical genetics of polydactyly: an updated review. Front Genet 2018;9.

[8] Verma PK, El-Harouni AA. Review of literature: genes related to postaxial polydactyly. Front Pediatr 2015;3:8.

[9] Ullah I, Kakar N, Schrauwen I, Hussain S, Chakchouk I, Liaqat K, et al. Variants in KIAA0825 underlie autosomal recessive postaxial polydactyly. Hum Genet 2019;138(6):593–600.

[10] Estrada-Cuzcano A, Etard C, Delvallée C, Stoetzel C, Schaefer E, Scheidecker S, et al. Novel IQCE variations confirm its role in postaxial polydactyly and cause ciliary defect phenotype in zebrafish. Hum Mutat 2020;41(1):240–54.

[11] Putoux A, Mougou-Zerelli S, Thomas S, Elkhartoufi N, Audollent S, Le Merrer M, et al. BBS10 mutations are common in 'Meckel'-type cystic kidneys. J Med Genet 2010;47(12):848–52.

[12] Yu S, Wu B, Qian Y, Zhang P, Lu Y, Dong X, et al. Clinical exome sequencing identifies novel CREBBP variants in 18 Chinese Rubinstein-Taybi Syndrome kids with high frequency of polydactyly. Mol Genet Genomic Med 2019;7(12). https://doi.org/10.1002/mgg3.v7.1210.1002/mgg3.1009.

[13] Deng H, Tan T, Yuan L. Advances in the molecular genetics of non-syndromic polydactyly. Expert Rev Mol Med 2015;17:e18.

[14] Kalsoom U-e, Klopocki E, Wasif N, Tariq M, Khan S, Hecht J, et al. Whole exome sequencing identified a novel zinc-finger gene ZNF141 associated with autosomal recessive postaxial polydactyly type A. J Med Genet 2013;50(1):47–53.

[15] Umair M, Shah K, Alhaddad B, Haack TB, Graf E, Strom TM, et al. Exome sequencing revealed a splice site variant in the IQCE gene underlying postaxial polydactyly type A restricted to lower limb. Eur J Hum Genet 2017;25(8):960–5.

[16] Wentzensen IM, Johnston JJ, Patton JH, Graham JM, Sapp JC, Biesecker LG. Exome sequencing identifies a mutation in OFD1 in a male with Joubert syndrome, orofaciodigital spectrum anomalies and complex polydactyly. Hum Genome Var 2016;3:15069.

[17] Wang T, Xuan Z, Dou Y, Liu Y, Fu Y, Ren J, et al. Identification of novel mutations in preaxial polydactyly patients through whole-exome sequencing. Mol Genet Genomic Med 2019;7(6). https://doi.org/10.1002/mgg3.2019.7.issue-610.1002/mgg3.690.

[18] Mohapatra B, Casey B, Li H, Ho-Dawson T, Smith L, Fernbach SD, et al. Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. Hum Mol Genet 2009;18(5):861–71.

[19] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43(5):491–8.

[20] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.

[21] Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med 2020;12(1):103.

[22] Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015;24(8):2125–37.

[23] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17(5):405–23.

[24] Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 2015;31(4):608–9.

[25] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.

[26] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017;45(D1):D362–8.

[27] Scardoni G, Petterlini M, Laudanna C. Analyzing biological network parameters with CentiScaPe. Bioinformatics 2009;25(21):2857–9.

[28] Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinf 2003;4:2.

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

[30] Zuniga A, Zeller R, Probst S. The molecular basis of human congenital limb malformations. Wiley Interdiscip Rev Dev Biol 2012;1(6):803–22.

[31] Suzuki M, Satoh A, Ide H, Tamura K. Transgenic Xenopus with prx1 limb enhancer reveals crucial contribution of MEK/ERK and PI3K/AKT pathways in blastema formation during limb regeneration. Dev Biol 2007;304(2):675–86.

[32] Al-Qattan MM, Shamseldin HE, Salih MA, Alkuraya FS. GLI3-related polydactyly: a review. Clin Genet 2017;92(5):457–66.

[33] Johnston JJ, Sapp JC, Turner JT, Amor D, Aftimos S, Aleck KA, et al. Molecular analysis expands the spectrum of phenotypes associated with GLI3 mutations. Hum Mutat 2010;31(10):1142–54.

[34] Smith MJ, Beetz C, Williams SG, Bhaskar SS, O'Sullivan J, Anderson B, et al. Germline mutations in SUFU cause Gorlin syndrome-associated childhood medulloblastoma and redefine the risk associated with PTCH1 mutations. J Clin Oncol 2014;32(36):4155–61.

[35] Potuijt JWP, Baas M, Sukenik-Halevy R, Douben H, Nguyen P, Venter DJ, et al. A point mutation in the pre-ZRS disrupts sonic hedgehog expression in the limb bud and results in triphalangeal thumb-polysyndactyly syndrome. Genet Med 2018;20(11):1405–13.