

Received: 2020.04.18

Accepted: 2020.06.08

Available online: 2020.06.22

Published: 2020.08.23

Integrated Bioinformatics Analysis Reveals Potential Pathway Biomarkers and Their Interactions for Clubfoot

Authors' Contribution:

Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

BCDEF Jing Ding*
BCDE Zhenpeng Liang*
EF Weijia Feng
AG Qixun Cai
AG Ziming Zhang

Department of Pediatric Orthopaedics, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, P.R. China

* Jing Ding and Zhenpeng Liang contributed equally and share first author

Corresponding Authors: Qixun Cai, e-mail: caiqixun@xinhuaumed.com.cn, Ziming Zhang, e-mail: zhangziming@xinhuaumed.com.cn

Source of support: This research was supported by the Foundation of Science and Technology Commission of Shanghai Municipality (17DZ1930507 and 17411965800)

Background: Congenital talipes equinovarus (clubfoot), one of the most regular pediatric congenital skeletal anomalies, seriously affects the normal growth and development of about 1 in 1000 newborns. Although it has been investigated widely, the etiology and pathogenesis of clubfoot are still controversial.

Material/Methods: g: Profiler, NetworkAnalyst and WebGestalt were used to probe the enriched signaling pathways by using the Gene Ontology (GO), Human Phenotype Ontology (HP), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome (REAC), and WikiPathways (WP) databases. Large numbers of enriched signaling pathways were identified using the integrated bioinformatics enrichment analyses.


Results: Apoptosis or programmed cell death (PCD), disease, muscle contraction, metabolism, and immune system were the top functions. Embryo or organ morphogenesis and development, cell or muscle contraction, and apoptosis were the top biological processes, and cell/muscle contraction and apoptosis were the top molecular functions using enriched GO terms analysis. There were a large number of complex interactions in the genes, enriched pathways, and transcription factor (TF)-miRNA co-regulatory networks. Transcription factors such as FOXN3, GLI3, HOX, and NCOR2 family regulated the gene expression of APAF1, BCL2, BID, CASP, MTHFR, and TPM family.

Conclusions: The results of bioinformatics enrichment analysis not only supported the previously proposed hypotheses, e.g., extracellular matrix (ECM) abnormality, fetal movement reducing, genetic abnormality, muscle abnormality, neurological abnormality, skeletal abnormality and vascular abnormality, but also indicated that cellular or immune responses to external stimulus, molecular transport and metabolism may be new etiological mechanisms in clubfoot.

MeSH Keywords: **Clubfoot • Databases, Genetic • Neuroimmunomodulation**

Abbreviations: **AdjP** – adjusted *P* value; **BP** – biological process; **BH** – Benjamini-Hochberg; **CC** – cellular component; **ECM** – extracellular matrix; **GO** – Gene Ontology; **GRN** – Gene Regulatory Networks; **HP** – Human Phenotype Ontology; **MF** – molecular function; **ORA** – Overrepresentation Enrichment Analysis; **PCD** – programmed cell death; **PPI** – protein-protein interactions; **REAC** – Reactome; **TF** – transcription factor; **WP** – WikiPathways; **-log₁₀(AdjP)** – negative log₁₀

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/925249>

 3745

 3

 8

 42



Background

Congenital talipes equinovarus, also named clubfoot, is one of the most common major congenital distal skeletal abnormalities in newborn infants with a prevalence of 1.13 per 1000 livebirths [1]. A previous study confirmed that about 20% of clubfoot patients are connected with chromosomal abnormalities [2]. Clubfoot is frequently observed in several neurological disorders, such as congenital myotonic dystrophy, distal arthrogryposis, myelomeningocele, and myotonic dystrophy. Therefore, it is regarded as a common component of these nervous system diseases [3]. However, about 80% of cases of congenital malformations are isolated clubfoot and this type termed as idiopathic clubfoot. Although the clinical presentation in foot is consistent or similar, clinical treatments and proposed genetic pathology mechanism are obvious different between idiopathic clubfoot and syndromic or secondary clubfoot. Syndromic cases seem to be due to neuromuscular, while idiopathic clubfoot may derive from fetal abnormalities [4,5].

Although the etiology and pathogenesis of clubfoot are still controversial, the genetic mechanisms of clubfoot are progressively being demonstrated with the application of genomics, proteomics, and other omics technologies. Recently, numerous abnormal expressed genes associated with clubfoot were gradually screened and identified. For examples, collagen encoding gene COL9A1, regulated by SOX9, has been proposed as a candidate gene for clubfoot [6]. In addition, some rare mutations were recognized in copy number variants in multiplex clubfoot pedigrees, such as developmentally expressed transcription factors and their regulators CHD1, HOXC13, PITX1, RIPPLY2, TBX4, and UTX [7]. And single nucleotide polymorphisms (SNPs) in apoptosis related genes, such as APAF1, BCL-2, BID, CASP3, CASP8, CASP9, and CASP10 are also significantly associated with clubfeet [8]. HOXD family genes have been suggested as susceptible genes for clubfeet and their mutations were association with clubfoot [9,10]. Recently, Pavone et al. reviewed the genetic and environmental risk factors of clubfoot and pointed out some genes expressed in embryogenesis and morphogenesis, embryonic development/morphogenesis, striated muscle contraction, apoptosis, necrosis, and inflammation processes play key roles in etiology of clubfoot, such as CASP, HOX, GLI3, MTHFR, PITX-TXB4, T-box, and troponin (TN) family genes [11]. Several potential markers of genes and/or pathways with clubfoot have been identified. Based on the results of these studies, several theories of clubfoot have been proposed, such as abnormal development of nerves, blood vessels or skeleton, connective tissue fibrosis, and abnormal extracellular matrix proteins [12–15]. However, these studies do not further explore the interactions among these genes, signaling pathways, and the gene upstream and downstream regulatory networks. Recently, we found that the results of bioinformatics analyses using abnormal proteins provided experimental

evidence for previously proposed hypotheses, e.g., the abnormalities of ECM, genetic, muscle, neurological, skeletal, and vascular [16]. Additionally, cellular or immune responses to external stimulus, molecular transport and metabolism were suggested as new promising etiological mechanisms in clubfoot [16]. However, the enriched results of bioinformatics analysis with abnormal genes in clubfoot were not explored. Therefore, the main purpose of this investigation is to study the interactions among the abnormal genes and abnormally active pathways which focus on clubfoot-related pathway enrichment analysis, generic gene-coding protein-protein interactions (PPI), gene-miRNA interactomes, transcription factor (TF)-gene interactions and TF-miRNA coregulatory interactions through bioinformatics analyses.

Material and Methods

Clubfoot-associated gene screening and inclusion

PubMed and Science Direct, 2 widely used databases, were adopted to screen by keywords on May 7, 2019. Keywords were “(clubfeet or clubfoot or congenital talipes equinovarus) and (embryology or etiology or etiopathogenesis or genetics or genomics or pathology or pathophysiology)”. After duplicates were excluded, we then further screened these studies using the title and abstract reading according to the inclusion and exclusion criteria as described in our recently work [16]. Finally, a total of 31 studies were performed for subsequent genetic screening analysis. After removing duplicate genes described in these candidate studies, 43 genes were firstly assessed. CASP and HOXD were removed due to attend a more detailed classification. CAND2 was excluded because of the mutation was silent (Figure 1). TNNT3 was not applied to analyze because of not associated with clubfeet. WNT7 was not included due to the mutation was in non-coding regions. We ultimately employed 37 genes shown in Table 1 [7–10,17–32] for further bioinformatics analysis.

WebGestalt analysis

The analysis was performed according to our previous work [16]. The interesting list contains 37 user IDs in which all of them are unambiguously mapped to 37 unique entrezgene IDs.

g: Profiler analysis

GO analysis [GO molecular function (GO: MF), GO cellular component (GO: CC), and GO biological process (GO: BP)] was performed as described previously [16]. Biological pathway databases were used Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome (REAC), WikiPathways (WP), and Human Phenotype Ontology (HP). The software version was e94_eg41_p11_9f195a1.

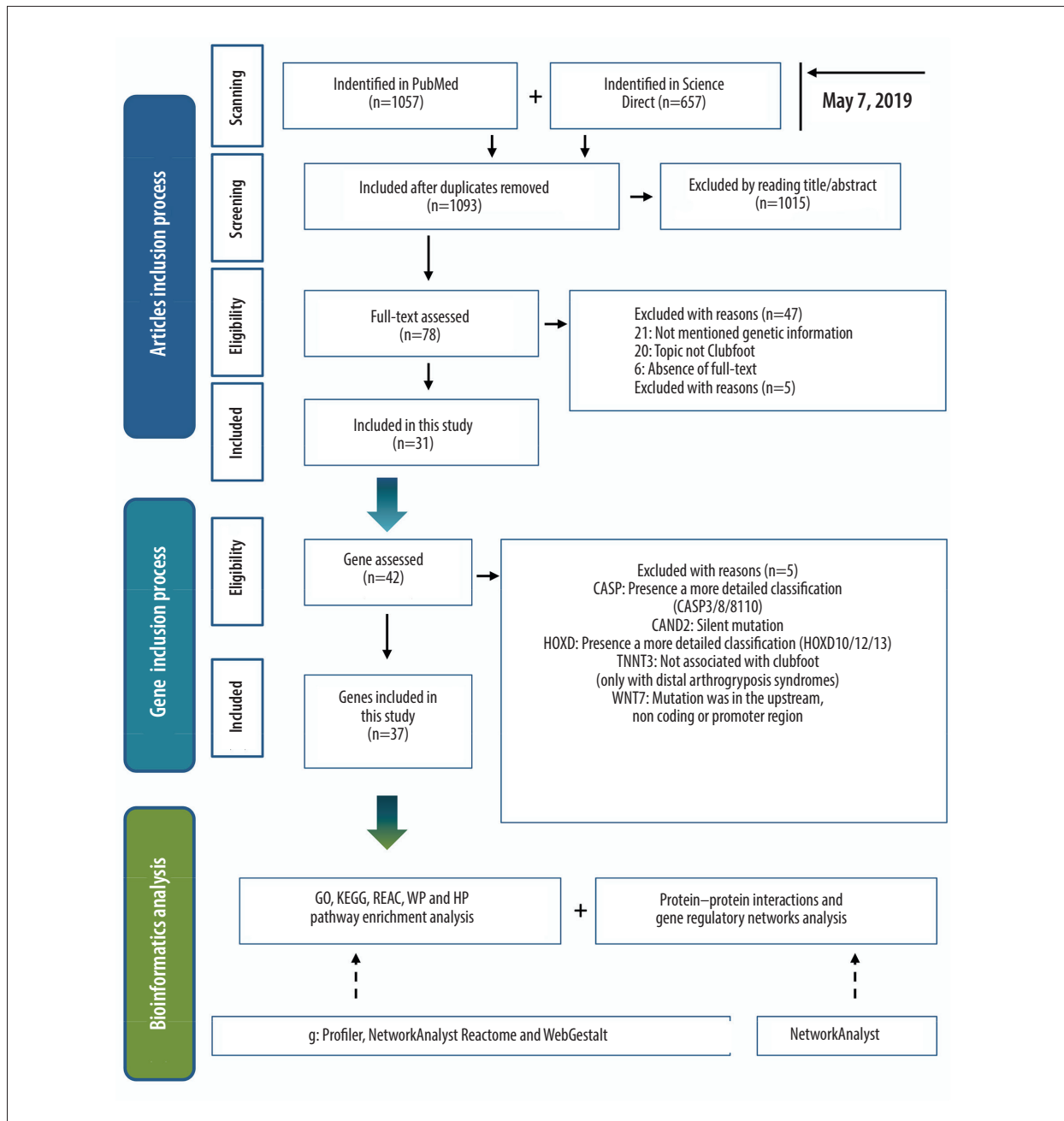


Figure 1. A flowchart of the inclusion of candidate articles/genes and process of bioinformatics analyses.

NetworkAnalyst analysis

The interactions were analyzed and mapped NetworkAnalyst analysis according to our recently study [16].

PPI analysis

The database was MEx Interactome. The analysis processes were performed as described previously [16].

Gene-miRNA interactome analysis

Gene-miRNA interaction was carried out according to our previous work [16] using TarBase and miRTarBase databases.

TF-gene interaction analysis

The analysis was performed as previously described in our previous work [16]. The database was ENCODE.

Table 1. Gene information.

User Id	Gene symbol	Gene name	Entrez gene	References
ANXA3	ANXA3	Annexin A3	306	17
APAF1	APAF1	Apoptotic peptidase activating factor 1	317	8
BCL2	BCL2	BCL2, apoptosis regulator	596	8
BID	BID	BH3 interacting domain death agonist	637	8
CASP10	CASP10	Caspase 10	843	8, 18
CASP3	CASP3	Caspase 3	836	8, 19
CASP8	CASP8	Caspase 8	841	8
CASP9	CASP9	Caspase 9	842	8
CHD1	CHD1	Chromodomain helicase DNA binding protein 1	1105	7
COL9A1	COL9A1	Collagen type IX alpha 1 chain	1297	20
FLNB	FLNB	Filamin B	2317	21
FOXP3	FOXP3	Forkhead box N3	1112	22
GLI3	GLI3	GLI family zinc finger 3	2737	23
HOXA9	HOXA9	Homeobox A9	3205	24
HOXC13	HOXC13	Homeobox C13	3229	7
HOXD10	HOXD10	Homeobox D10	3236	9, 10, 25, 26
HOXD12	HOXD12	Homeobox D12	3238	9
HOXD13	HOXD13	Homeobox D13	3239	9
IGFBP4	IGFBP4	Insulin like growth factor binding protein 4	3487	19
MMP7	MMP7	Matrix metalloproteinase 7	4316	22
MTHFR	MTHFR	Methylenetetrahydrofolate reductase	4524	17
MYH1	MYH1	Myosin heavy chain 1	4619	27
MYH2	MYH2	Myosin heavy chain 2	4620	27
MYH3	MYH3	Myosin heavy chain 3	4621	27
MYH8	MYH8	Myosin heavy chain 8	4626	27
NAT1	NAT1	N-acetyltransferase 1	9	28
NAT2	NAT2	N-acetyltransferase 2	10	28
NCOR2	NCOR2	Nuclear receptor corepressor 2	9612	22
PITX1	PITX1	Paired like homeodomain 1	5307	7, 29
RIPPLY2	RIPPLY2	Ripply transcriptional repressor 2	134701	7
SORCS1	SORCS1	Sortilin related VPS10 domain containing receptor 1	114815	22
TBX4	TBX4	T-box 4	9496	7, 30, 31
TMEM123	TMEM123	Transmembrane protein 123	114908	22
TNNC2	TNNC2	Troponin C2, fast skeletal type	7125	32
TPM1	TPM1	Tropomyosin 1	7168	32
TPM2	TPM2	Tropomyosin 2	7169	24
ZNF664	ZNF664	Zinc finger protein 664	144348	22

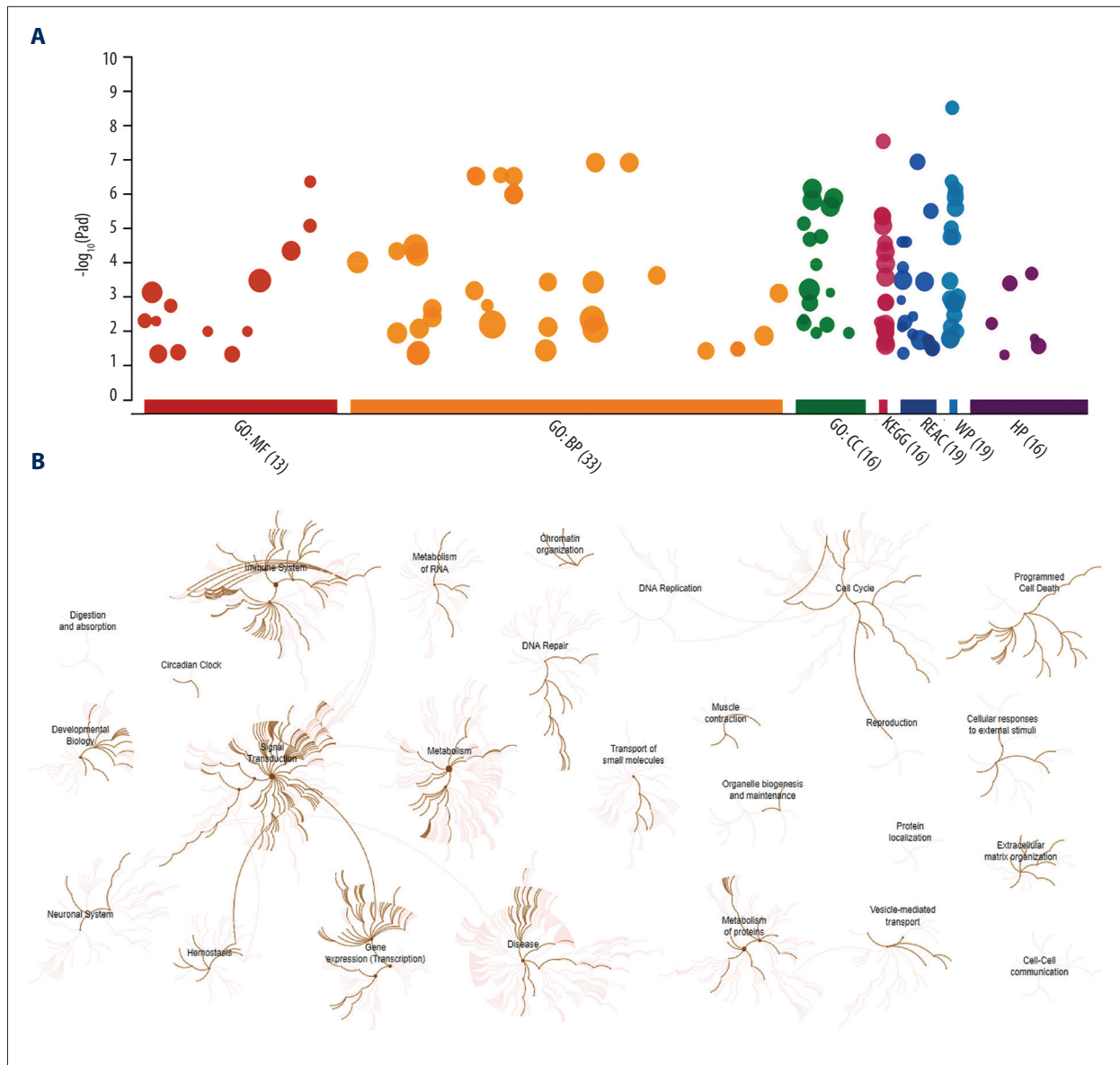


Figure 2. Overall sketch results of enrichment analysis with the candidate genes with g: Profiler and REAC. **(A)** Significant signaling pathways were analyzed by GO, KEGG, REAC, WP, and HP databases. **(B)** Significantly signaling pathways and interactions obtained using REAC databases. REAC – Reactome; GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes; WP – WikiPathways; HP – Human Phenotype Ontology.

TF-miRNA coregulatory network analysis

The coregulatory network analysis was carried out to our recently work [16] and the interaction data was used RegNetwork repository.

Statistics

Only annotated genes in the NCBI gene database were included to analysis. Adjusted *P*-value (*AdjP*) was used the method

of conventional BH procedure. A significant difference was used an *AdjP* value <0.05.

Results

Thirty-seven genes were obtained using article screening

We searched the 2 major medical databases (PubMed and Science Direct) by keywords on May 7, 2019 and 1093 articles were found after duplicates excluded. We then further screened these

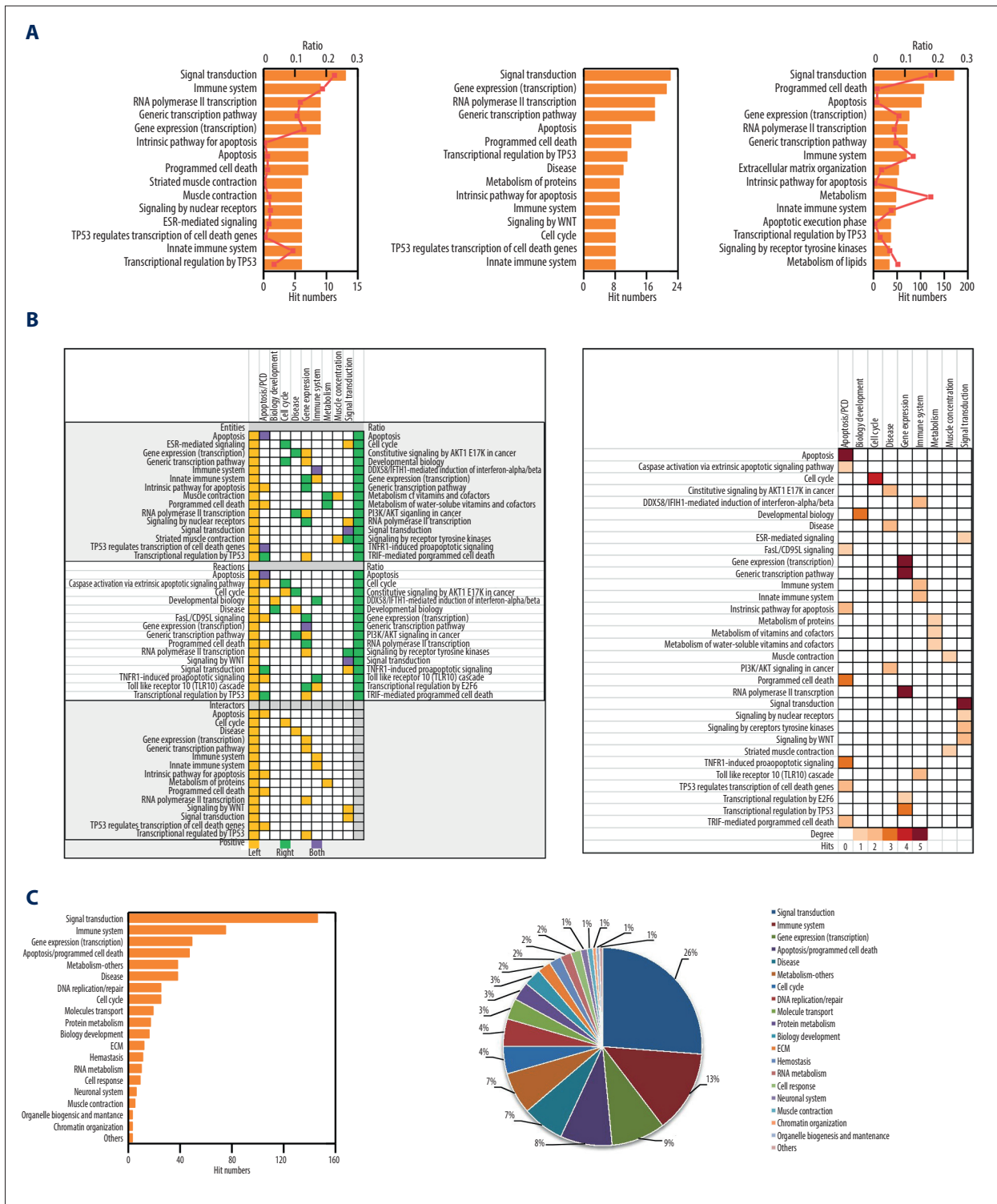
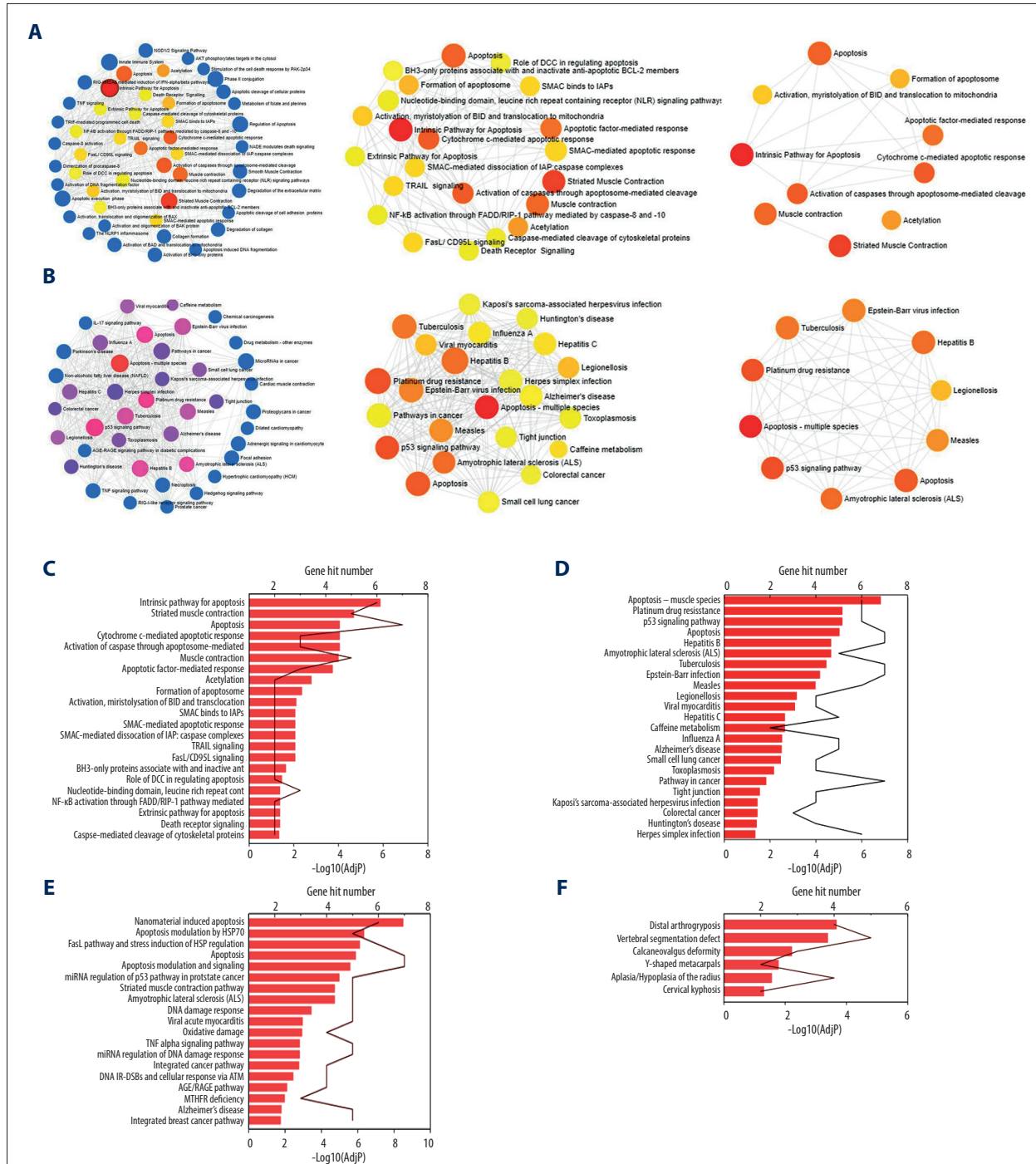


Figure 3. Classification statistics of top significant enrichment pathways from REAC analysis results as shown in Figure 2B. **(A)** Classification statistics of top 15 enrichment pathways. The hits of pathways (entities) and their ratio were shown in the **left** histogram and line combination chart. Interactors were in the **middle**, and reactions and their ratios were shown in the **right**. The hit numbers were presented as the histogram and the ratio was presented as a line chart. Top 15 pathways were used to analyze due to some pathways jointly topped 10. **(B)** Distribution of the top 15 pathways. **(C)** Subtotal analysis results of pathways enriched by REAC. **Left:** distribution, and **Right:** frequency. REAC – Reactome.

studies using the title and abstract reading according to the inclusion criteria as described in Materials and Methods section. Seventy-eight articles were retained behind and were accessed for full text. After reading the full text, 21 articles were excluded for not describing genetic information and 20 texts were also excluded for that the topic was not for clubfeet. And 6 articles were removed for not accessed for full text. Finally, a total of 31 studies were performed for subsequent genetic screening analysis.

Forty-three genes were firstly assessed following excluding duplicate genes described in these candidate studies. TNNT3 was not applied to analyze because of not associated with clubfeet. CAND2 was excluded owing to the mutation was silent. WNT7 was not included because of the mutation was in the non-coding region. CASP and HOXD were removed due to attend a more detailed classification. Finally, 37 genes were selected to further bioinformatics analysis (Table 1).



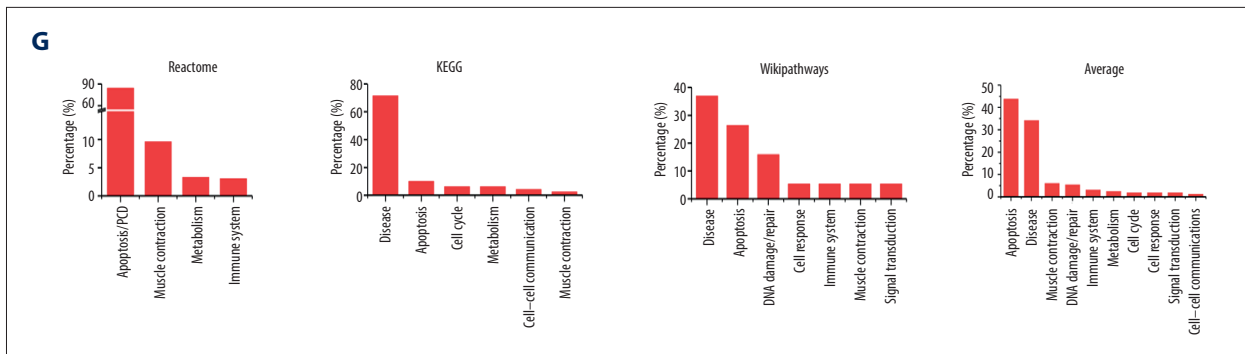


Figure 4. Pathway enrichment analysis. The size of the icon means gene hit numbers in pathways. The bigger the size, the more the gene hit numbers. Blue icons: $P < 0.05$ but adjusted P (Adj P) > 0.05 . Other colors: both P value and Adj P value < 0.05 . The darker the color, the larger the $-\log_{10}(\text{Adj}P)$. (A) REAC enriched analysis using NetworkAnalyst. The interactions of the signalings enriched with REAC with a $P < 0.05$ but Adj $P > 0.05$ (left), $P < 0.05$ and Adj $P < 0.05$ (middle), and top 10 interactions (right). (B) KEGG enrichment analysis by NetworkAnalyst. The interactions of the pathways enriched by REAC with a $P < 0.05$ but Adj P value > 0.05 (left), $P < 0.05$ and Adj $P < 0.05$ (middle), and top 10 interactions (right). Statistical results of REAC (C), KEGG (D), and WP (E) enrichment analysis by g: Profiler with a $P < 0.05$ and Adj $P < 0.05$. (F) HP enrichment analysis with g: Profiler. The $-\log_{10}(\text{Adj}P)$ value was presented as the histogram and the hit gene numbers was presented as a line chart as shown in C–F. (G) The percentage of the signaling pathways in databases. The value was the average of percentage of pathways in REAC or KEGG analyzed by g: Profiler, NetworkAnalyst, and WebGestalt (Supplementary Figure 1) shown in the 3 histograms on the left. Average of percentage of pathways in REAC, KEGG, and WP was placed on the right. REAC – Reactome; GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes; WP – WikiPathways; HP – Human Phenotype Ontology.

Overall views of results of pathway enrichment analysis

There were numerous pathways identified by g: Profiler (Figure 2A). Meanwhile, we further utilized another bioinformatics analysis software REAC Pathway Browser and found numerous numbers of enrichment pathways (Figure 2B). Additionally, the result of enrichment pathways by REAC Pathway Browser showed a more intuitive representation of the interactions and connections between pathways shown in Figure 2B. The higher the enrichment, the darker the color of the pathways.

In order to further analyze the most enriched pathways, a total of 557 pathways from REAC Pathway Browser analysis were ranked. Apoptosis or programmed cell death, immune system, signal transduction, and transcription were the top 10 pathway hit numbers (Figure 3A left and Supplementary Table 1). Additionally, muscle contraction, signal transduction and metabolism were also ranked the top (Figure 3A left and Supplementary Table 1). The ratio of the pathway aforementioned hits (entities) was also ranked the top (Figure 3A left and Supplementary Table 2). And biology development, disease and cell cycle were also the top rank (Figure 3A left and Supplementary Table 1). The top pathway interactors were mostly distributed in signal transduction, gene expression (transcription), apoptosis/PCD, disease, immune system, metabolism of proteins and cell cycle (Figure 3A middle and Supplementary Table 1). Apoptosis or PCD, biology development, cell cycle, disease, immune system, signal transduction, and transcription were also the top reactions and ratios (Figure 3A right and Supplementary Table 1). To intuitively show the situation

of top 15 pathways, we further classified these pathways. We can clearly find that apoptosis, gene expression (transcription), generic transcription pathway, RNA polymerase II transcription, signal transduction and cell cycle were top 5 pathways in REAC database (Figure 3B). According to the pathway classification, signal transduction, immune system, gene expression (transcription), apoptosis/programmed cell death and metabolism (including RNA and protein) were top 5 pathway classification and the sum of the 5 was 68.6% (Figure 3C). Additionally, the sum of top 10 pathways was 90.7% (Figure 3C). Besides, biology development, cell cycle, disease, DNA replication/repair, and molecular transport were also enriched.

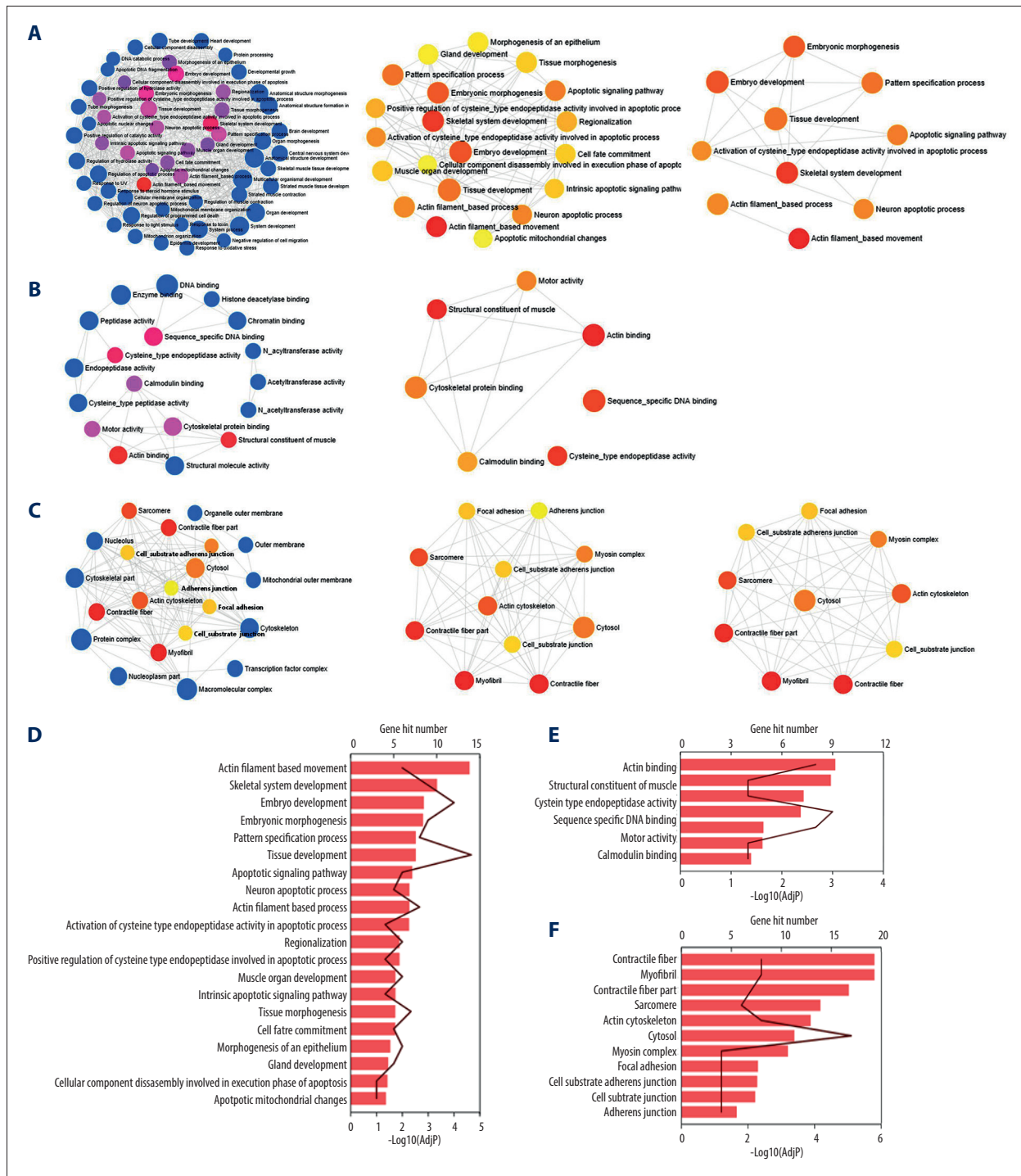
Results of pathway enrichment analysis

There were 22, 23, 19, and 6 remarkably altered pathways that found in REAC, KEGG, WP, and HP databases with Adj $P < 0.05$ (Figure 4A–4F). Among of them, apoptosis/PCD, muscle contraction, metabolism and immune system were significantly enriched pathways in REAC database using NetworkAnalyst (Figure 4C, 4G and Supplementary Table 2). Consistent results were obtained using g: Profiler and WebGestalt (Supplementary Figure 1). Apoptosis/PCD was the top significantly enriched pathway with a percentage of 84.1% (Supplementary Table 2). Using the same analysis method, disease and apoptosis were the top pathways in KEGG (Figure 4D, 4G and Supplementary Table 2) and WP databases (Figure 4E, 4G and Supplementary Table 2). After summarizing the data from the 3 databases (REAC, KEGG, and WP), apoptosis or PCD and disease were the top 2 signaling pathways and the sum was 76% (Figure 4G and Supplementary Table 2).

Results of enriched GO terms analysis

There were 20, 7, and 11 significantly enriched GO terms (AdjP <0.05) found using NetworkAnalyst in BP, MF, and CC terms, respectively (Figure 5A–5F). Among of the BP terms, embryo or organ morphogenesis and development, cell or muscle contraction, and apoptosis were the top processes (Figure 5D, 5G),

such as limb development, appendage development, muscle filament sliding, actin-myosin filament sliding and embryonic appendage morphogenesis. The sum of the percentage of top 2 BPs was more than 90% (Figure 5G). In addition, cell/muscle contraction and apoptosis were the top molecular functions and the percentage of the 2 molecular functions reached 75% of the total shown in Figure 5E, 5G, for example,



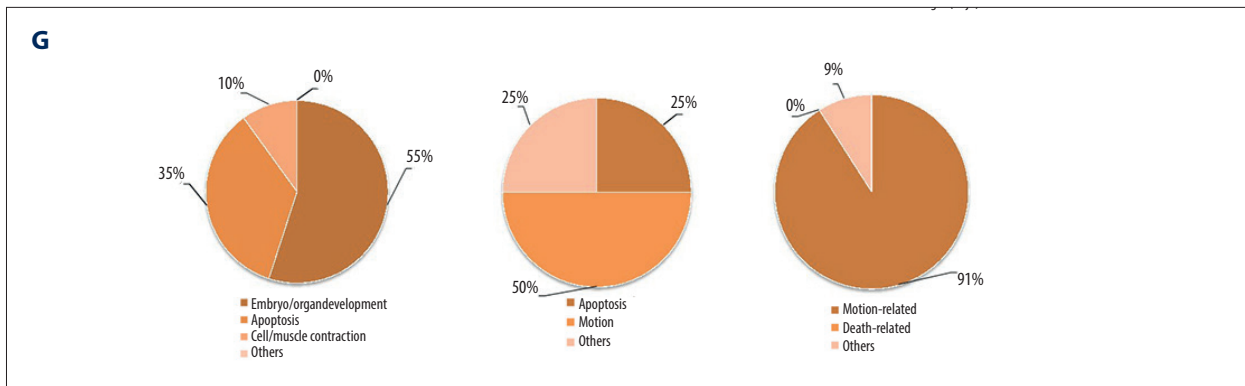


Figure 5. The enriched GO terms in the BP, MF, and CC categories for included candidate genes of clubfoot with NetworkAnalyst. The size of the icon means gene hit numbers in pathways. The bigger the size, the more the gene hit numbers. Blue icons: $P < 0.05$ but adjusted P ($AdjP$) > 0.05 . Other colors: both P value and $AdjP < 0.05$. The darker the color, the larger the $-\log_{10}(AdjP)$. (A) Biological process (BP). The interactions of BP enriched by GO with a $P < 0.05$ but $AdjP > 0.05$ (left), $P < 0.05$ and $AdjP < 0.05$ (middle), and top 10 interactions (right). The size of the icon means gene hit numbers in pathways. The bigger the size, the more the gene hit numbers. Blue icons: $P < 0.05$ but $AdjP > 0.05$. Other colors: both P and $AdjP < 0.05$. The darker the color, the larger the $-\log_{10}(AdjP)$. (B) Molecular function (MF): the interactions of MF enriched by GO with a $P < 0.05$ but $AdjP > 0.05$ (left) and $P < 0.05$ and $AdjP < 0.05$ (right). (C) Cellular component (CC): the interactions of CC enriched by GO with $P < 0.05$ but $AdjP$ value > 0.05 (left), $P < 0.05$ and $AdjP < 0.05$ (middle), and top 10 interactions (right). (D) Statistical results of the enriched Gene Ontology (GO) terms in BP (D), MF (E), and CC (F) categories with $P < 0.05$ and $AdjP < 0.05$. The $-\log_{10}(AdjP)$ value was presented as the histogram and the hit gene numbers was presented as a line chart. (G) Summary results of the enriched GO terms. Left: BP; middle: MF; right: CC.

cysteine-type endopeptidase activity involved in apoptosis and actin filament binding. Besides, motion cellular component was the top cell component, such as sarcomere, contractile fiber, myofibril, muscle myosin, myosin filament, and death-inducing signaling complex (Figure 5F). The sum of the percentage of the top cellular component was more than 90% (Figure 5G). Similar results were also detected in GO enrichment analysis by NetworkAnalyst (Supplementary Figure 2).

Results of PPI and gene regulatory networks (GRN) analysis

We then further analyzed the generic PPI of these candidate genes using NetworkAnalyst. Extensive and complex PPI was present among these proteins (Figure 6A). The 29 candidate gene coding proteins (excepted COL9A1, IGFBP4, MTHFR, RIPPLY2, SORCS1, TBX4, TNNC2, and ZNF664) were interacted with a total of 617 predicted proteins. Among of the total interactions ($N=1779$), the percentage of the top 10 interactions that were comprised of CASP8, BCL2, NCOR2, CASP3, TPM1, FLNB, BID, CASP9, CASP10, and HOXA9 was 40% (Figure 6A).

We then further analyzed the interaction between the candidate genes and miRNAs. A total of 31 genes except MYH3, MYH8, NAT2, SORCS1, TNNC2, and ZNF664 were included to analyze these interactions using NetworkAnalyst. Principal 10 miRNAs were hsa-mir-26b-5p, hsa-mir-21-5p, hsa-mir-16-5p, hsa-let-7a-5p, hsa-mir-1-1, hsa-mir-15a-5p, hsa-mir-15b-5p, hsa-mir-195-5p, hsa-mir-34a-5p, and hsa-mir-497-5p. Although

the proportion of these 10 miRNAs is less than 2%, the proportion of interactions formed by them is close to 7%. Additionally, the percentage of interactions formed by top 10 genes (HOXA9, BCL2, IGFBP4, FOXN3, APAF1, NCOR2, MTHFR, TPM2, CASP3, and CASP8) was 64% (Figure 6B).

TF-gene interactions were further determined using NetworkAnalyst. Twenty-six genes except BID, CASP9, GLI3, MYH1, MYH2, MYH3, MYH8, NAT1, NAT2, NCOR2, and SORCS1 were included. The top genes were MAHFR, TMEM123, HOXA9, HOXD12, TPM1, HOXD13, TNNC2, IGFBP4, FLNB, and HOXC13 (Figure 6C). The percentage of TF-gene interactions formed by the top 10 genes was greater than 36%. Additionally, we further analyzed TF-miRNA coregulatory interactions. Thirty-four genes except MYH8, NAT2, and ZNF664 were included in this analysis. HOXA9, GLI3, BCL2, FOXN3, NCOR2, CHD1, SORCS1, HOXD10, CASP8, and CASP3 were the top 10 genes and the percentage of interactions formed by top 10 genes was greater than 45% (Figure 6D). Transcription factors such as GLI3, NCOR2, FOXN3, and HOX family regulated the gene expression of APAF1, BCL2, BID, MTHFR, TPM, and CASP family.

Finally, 4 types of these interactions aforementioned were consolidated and further analyzed. A total of 25 genes were participated in these interactions (Figure 6E). The top 10 genes were HOXA9, BCL2, CASP3, CASP8, FLNB, FOXN3, IGFBP4, MATHFR, NCOR2, and TPM1 (Figure 6F). The ratio of hit interactions by top 10 genes was 62.5%.

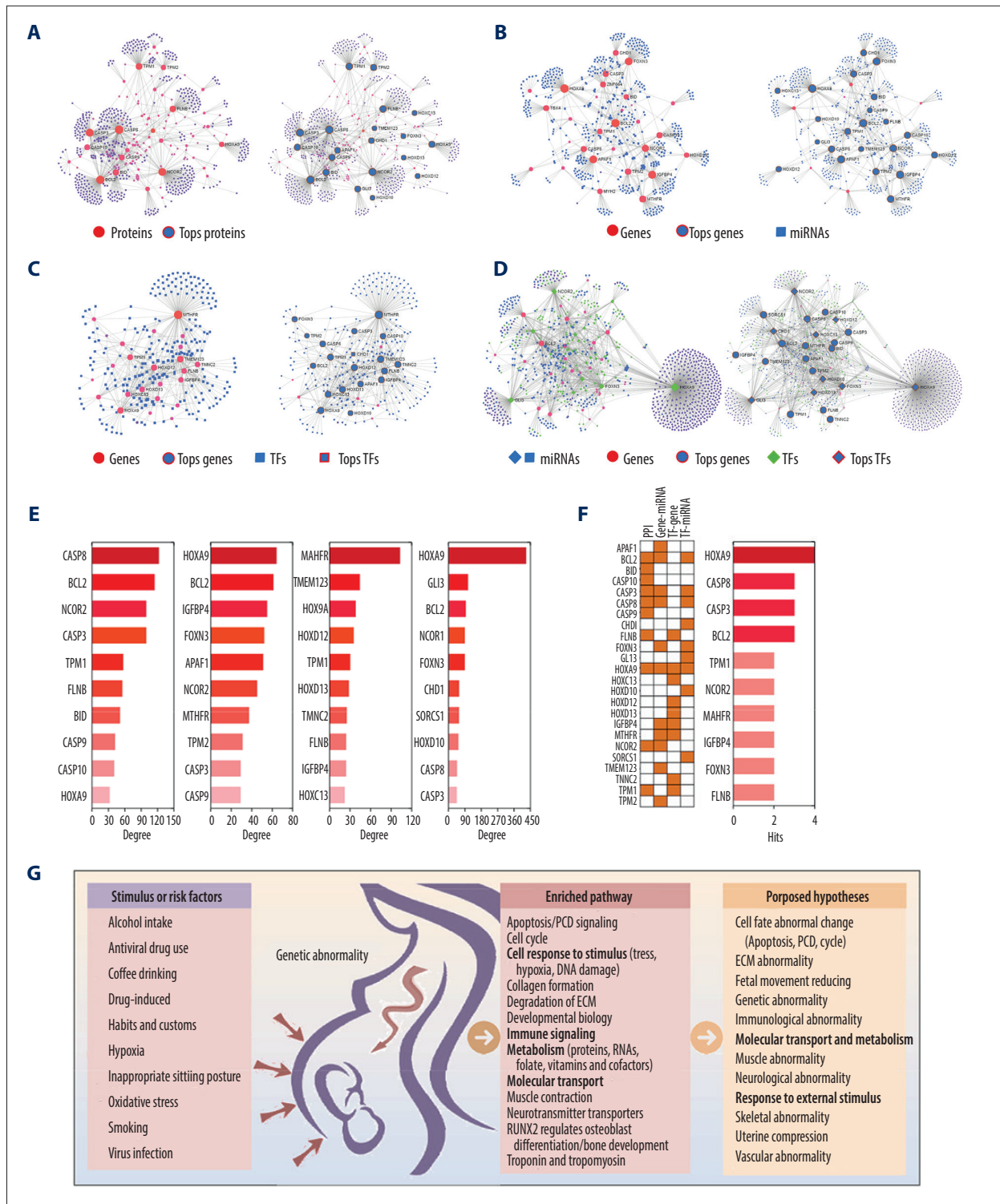


Figure 6. Protein-protein Interactions (PPI) and Gene Regulatory Networks (GRN) of the candidate genes by NetworkAnalyst. (A) Generic PPI. (B) Gene-miRNA Interactome. (C) Transcription factor (TF)-gene interaction database. (D) TF-miRNA coregulatory interaction. Right was the top 10 genes in the 4 aforementioned interactions in A-D. (E) Top 10 interactions. The order from left to right is PPI, Gene-miRNA, TF-gene interaction, and TF-miRNA coregulatory interaction. (F) The distribution and frequency of the top 10 genes in the 4 aforementioned interactions. Left: distribution. Right: frequency. (G) Proposed theories of clubfoot.

Discussion

Although emerging technologies including various omics have been widely used in the study of clubfoot, the molecular etiology of clubfoot remains to be revealed. In the present study, we found that the pathways in biology development, apoptosis/programmed cell death, immune system, cell cycle, muscle contraction, and ECM play key role in clubfoot through the enriched pathway analysis of abnormal genes by REAC, KEGG, WP, and HP enrichment analysis. And there are extensive interactions among these pathways. Additionally, there are also extensive interactions between the top TFs, such as CHD1, FOXN3, GLI3, HOXA9, HOXC13, HOXD10, HOXD13, NCOR2, and the top genes, such as APAF1, BCL2, BID, CASPs, MTHFR, SORCS1, and TPM1.

There are 557 pathways enriched by REAC analysis. These pathways focus on the signal transduction, immune system, gene transcription, apoptosis/programmed cell death, metabolism, etc. The sum of percentage of top 5 is almost 70%. Besides, there are a lot of pathways concentrated in cell cycle, DNA replication/repair, molecule transport, biology development, ECM, and muscle contraction. REAC, KEGG, and WP enriched analyses using g: Profiler, NetworkAnalyst, and WebGestalt were further carried out. Apoptosis and disease-associated pathways are dominant. There are also numerous pathways participated in muscle contraction, DNA replication/repair, immune system, metabolism, and cell cycle. GO enrichment analysis was then performed using NetworkAnalyst. Obviously, 90% of the GO terms focus on biology development, apoptosis/programmed cell death, and muscle contraction. Additionally, wide interaction occurs between embryo and tissue development and apoptosis pathways.

Apoptosis or programmed cell death signaling pathway composed of APAF1, BCL2, BID, CASP3, CASP8, and CASP9. During the limb development, apoptosis related genes were firstly associated with clubfoot [8,18]. Large numbers of significant signaling pathways were identified, such as SMAC, XIAP-regulated apoptotic response, caspase activation via extrinsic apoptotic pathway, cytochrome c-mediated apoptotic response, cell death regulated by TP53, intrinsic pathway for apoptosis, and PCD. These results further demonstrate that apoptosis or programmed cell death is a key signaling in clubfoot.

Cell cycle signaling was also enriched by REAC database. For example, PTK6 regulates cell cycle, TP53 regulates transcription of genes involved in G1/2 cell cycle arrest, and cell cycle checkpoints and mitotic pathway. Li et al. reported that growth factors were expressed higher in the contracted tissues compared with control tissues, and growth factor blockade could reduce collagen expression and proliferation in the cells obtained from the more contracted tissues [33]. Although more

evidence is needed, it seems that abnormally activated growth factors may change cell cycle in clubfoot.

Biology development abnormalities, for instance, muscle abnormality, neurological abnormality, skeletal abnormality, and vascular abnormality, have been proposed as playing key roles in clubfoot by several research teams or scholars [3–5,11,15,34]. COL9A1, NCOR2, HOX, and MYH family genes participated in several pathways in biology development, such as pervasive developmental disorders, activation of anterior HOX genes in hindbrain development, RUNX2 regulates osteoblast differentiation and bone development. In addition, HOXD10, MYH3, and TPM2 participated in the regulation of calcaneovalgus deformity, and GLI3, MYH3, MYH8, and TPM2 were involved in distal arthrogyposis. We also found that MYH3, MYH8, TPM2, TPM1, and TNNC2 participated in the regulation of myofibril, contractile fiber, myofibril, muscle thin filament tropomyosin, actin cytoskeleton, and myosin filament functions and thus caused abnormal muscle contraction. These data supported that striated muscle contraction was a common proposed hypothesis in the pathogenetic mechanisms of clubfoot [34]. Naturally, fetal movement draws support from normal contraction and relaxation of muscles and may participate in the development of joints, and reduced fetal movement was proposed a new hypothesis by Hester et al. in the morphogenesis of the forefoot [5].

Abnormal change of extracellular matrix proteins such as collagen was associated with clubfoot [20]. And collagen gene COL9A1 was a susceptible gene in clubfoot [15]. We found that MMP7, CASP3, and COL9A1 genes were involved in the progress of collagen formation and degradation of the extracellular matrix. In addition, interleukin-13 (IL-13) may induce myofibroblasts to produce ECM which indicates that excess stimulation induced by chronic inflammation promotes excessive accumulation of ECM [35]. These bioinformatics results supported the previous hypothesis that ECM abnormality and immune responses participated in the etiopathogenetic mechanism of clubfoot.

Molecular transportation is critical for material transportation, signaling transduction, neurotransmitter transporter and cell response to external stimulants [36]. Several pathways were enriched, such as amino acid transport across the plasma membrane, Na⁺/Cl⁻ dependent neurotransmitter transporters, PIPs transport between early and late endosome and Golgi membranes, regulation of insulin-like growth factor (IGF) transport, transport of inorganic cations/anions and amino acids/oligopeptides, and vesicle-mediated transport. These disorders of molecular transportation are very serious for cells. Therefore, we proposed that metabolism abnormality may be a new theory of clubfoot.

Metabolism of cofactors, folate, proteins, RNAs and vitamins, play key roles in the cell function [37], and were also enriched in the present enrichment analyses. BCL2, IGFBP4, MTHFR, NCOR2, and NAT family genes participated in the regulation of these molecules' metabolism processes. Additionally, vitamins and folic acid were associated with neural tube defects [38]. These findings suggested that metabolism abnormality may be a new hypothesis for clubfoot.

It is essential for cells to maintain homeostasis when they suffer from external or internal molecular and physical stresses, such as oxidative stress and hypoxia. Inappropriate cell response to stimulus or incorrect DNA damage response can cause unexpected damage or variation in cells. It has been confirmed that smoking and higher levels of alcohol and coffee intake are closely related to the disease [39]. These xenobiotic metabolisms, such as acetylation of toxins including free radicals in cigarette by N-acetyltransferase (NAT) genes, were considered to be a booster for clubfoot [28]. More importantly, the use of antiviral drugs in pregnancy is more frequent in clubfoot cases than in controls with an odds ratio=4.22 [40]. We found that FLNB was involved in the antiviral mechanism of IFN-stimulated genes and ISG15, OAS (oligoadenylate synthetase) antiviral response or Epstein-Barr virus infection. These data indicated that activation of the antiviral signaling induced virus infection or antiviral drug use may be an initiation or propellant of clubfoot. Additionally, the changes of immune signaling pathways were inevitably affected by these responses. We found that APAF1, BCL2, CASP family, FLNB, and MYH2 were enriched in the regulation of immune signaling. Besides, clubfoot is intimately related to genetic abnormality [41]. This evidence indicated that external or internal stimulus can increase the risk of clubfoot. The adverse stimulus causes abnormal cellular response, such as cell apoptosis, programmed cell death, cell cycle alteration, and immune disorder, and thus induces the formation and development of clubfoot. Cellular responses to external stimulus enriched by WP database, such as oxidative stress, hypoxia, and DNA damage induced by smoking, alcohol, coffee, or viral infection, may be a new hypothesis in the etiopathogenetic mechanism of clubfoot. Although more direct evidence is still needed, it seems that inappropriate cell response may cause unnecessary uterine compression, and eventually cause fetal movement reducing.

Based on the results of the present study and previous work, we mapped risk factors, signaling pathways, and proposed hypotheses for clubfoot (Figure 6G). Risk factors of clubfoot were alcohol intake, antiviral drug use, coffee drinking, drug-induced, habits and customs, hypoxia, inappropriate sitting posture, oxidative stress, smoking, virus infection and genetic factors [5,38–42]. The signaling pathways, enriched by integrated bioinformatics analysis, were apoptosis/PCD signaling, cell cycle, cell response to stimulus, collagen formation, degradation

of ECM, developmental biology, immune signaling, metabolism (proteins, RNAs, folate, vitamins and cofactors), molecular transport, muscle contraction, neurotransmitter transporters, RUNX2 regulates osteoblast, differentiation/bone development, troponin and tropomyosin signaling [7–10,1540,27–34]. Cell fate abnormal change, ECM abnormality, fetal movement reducing, genetic abnormality, immunological abnormality, metabolism abnormality, muscle abnormality, neurological abnormality, response to external stimulus, skeletal abnormality, uterine compression and vascular abnormality were proposed as the hypothesis in the etiopathogenetic mechanism of clubfoot [3–5,11,15,36,41]. Recently, we found that the enriched results of bioinformatics analysis using abnormally expressed proteins in clubfoot not only supported etiological hypotheses previously proposed, e.g., ECM abnormality, genetic abnormality, immune responses to stimuli, muscle abnormality, neurological abnormality, but also suggested that the abnormal changes in cellular responses to stimuli, molecular transport and metabolism may be new potential pathogenic mechanisms in clubfoot [16]. Although more direct evidence is needed to verify, the enriched results of bioinformatics analysis with abnormal genes and proteins indicated that cellular or immune responses to external stimuli, molecular transport and metabolism may be new etiological mechanisms in clubfoot.

False positives or omissions were inevitable in the bioinformatics analysis due to these candidate genes are not obtained from the same study. Therefore, direct experimental evidence is needed to prove these hypotheses. Clinical study is in progress to confirm the proposed hypotheses.

Conclusions

Large numbers of signalings were identified with bioinformatic analysis. Apoptosis or PCD, diseases, muscle contraction, metabolism and immune system were the top functions. The GO analysis results indicated that apoptosis, cell or muscle contraction, embryo or organ morphogenesis and development were the top biological processes and apoptosis and cell/muscle contraction were the top molecular functions. Results of PPI and GRN analysis suggested that there were many complex interactions in the genes, signalings, and TF-miRNA co-regulatory network. Transcription factors such as FOXN3, GLI3, NCOR2, and HOX family regulated the gene expression of APAF1, BCL2, BID, CASP, MTHFR, and TPM family. The results of this study consistent with the results of our recently study [16] not only supported etiological hypotheses previously proposed, e.g., ECM abnormality, genetic abnormality, immune responses to stimuli, muscle abnormality, and neurological abnormality, but also suggested the abnormal changes in the cellular responses to external stimuli, molecular transport and metabolism may be new potential etiological mechanisms of clubfoot.

Availability of data and materials

All data generated or analyzed during this study are in this published article (and its supplementary information files).

Conflict of interest

None.

Supplementary Data

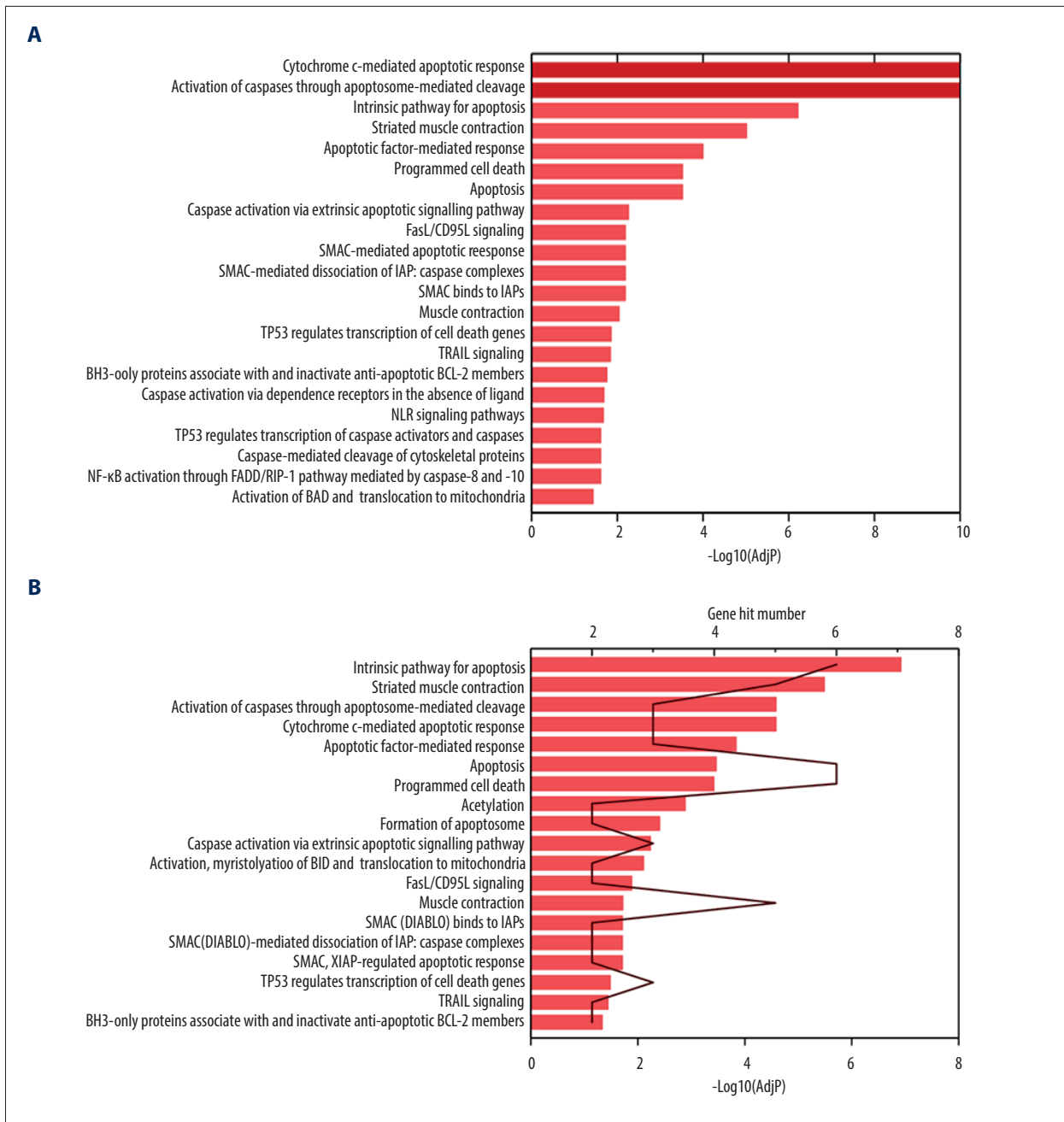
Supplementary Table 1. Top 15 significantly changed pathways enriched by REAC database.

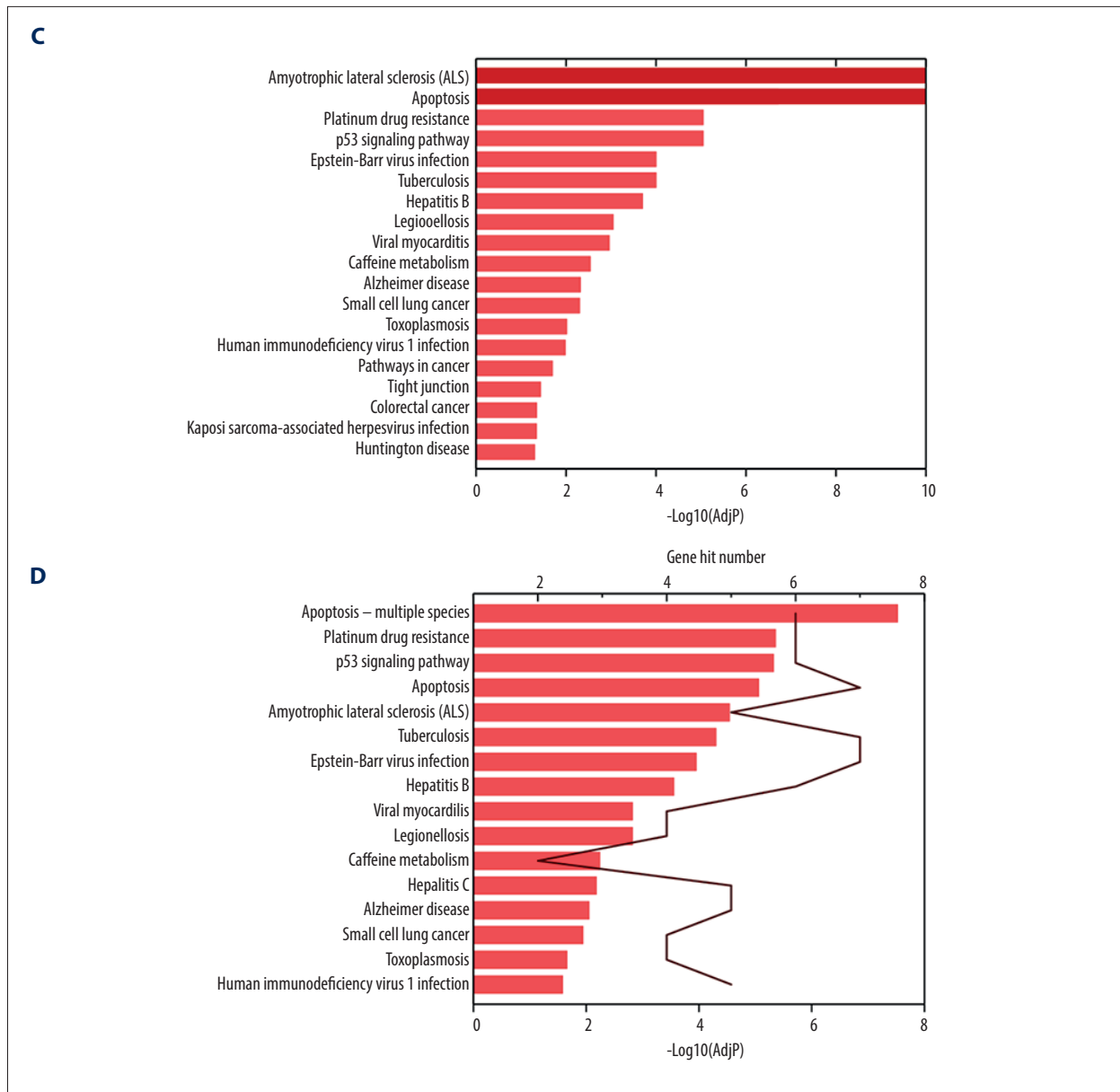
Pathway identifier	Pathway name	Entities		Entities ratio		Interactors		Reactions		Reactions ratio	
		NO.	RANK	Ratio	RANK*	NO.	RANK*	NO.	RANK*	Ratio	RANK*
R-HSA-162582	Signal transduction	13	1	22.62%	1	22	1	170	1	18.27%	1
R-HSA-73857	RNA polymerase II transcription	9	2	12.87%	3	18	3	75	3	8.11%	3
R-HSA-212436	Generic transcription pathway	9	3	11.76%	4	18	4	71	4	7.20%	4
R-HSA-74160	Gene expression (transcription)	9	4	10.77%	5	21	2	71	5	6.69%	5
R-HSA-168256	Immune system	9	5	0.28%	13	9	8	4	13	0.03%	15
R-HSA-109581	Apoptosis	7	6	18.81%	2	12	5	70	6	12.60%	2
R-HSA-109606	Intrinsic pathway for apoptosis	7	7	0.59%	9	9	9	14	9	0.56%	9
R-HSA-5357801	Programmed cell death	7	8	0.42%	12	12	6	49	7	0.50%	10
R-HSA-3700989	Transcriptional regulation by TP53	6	9	1.32%	8	11	7	101	2	1.15%	6
R-HSA-168249	Innate immune system	6	10	2.24%	6	8	10	12	10	1.08%	7
R-HSA-5633008	TP53 regulates transcription of cell death genes	6	11	1.80%	7	8	11	12	11	0.91%	8
R-HSA-390522	Striated muscle contraction	6	12	0.49%	10	0	15	29	8	0.44%	11
R-HSA-8939211	ESR-mediated signaling	6	13	0.48%	11	4	12	2	14	0.30%	12
R-HSA-9006931	Signaling by nuclear receptors	6	14	0.23%	14	4	13	6	12	0.21%	13
R-HSA-397014	Muscle contraction	6	15	0.14%	15	1	14	1	15	0.07%	14

* Ranked in the top 15 entities found, not in total significantly changed pathways.

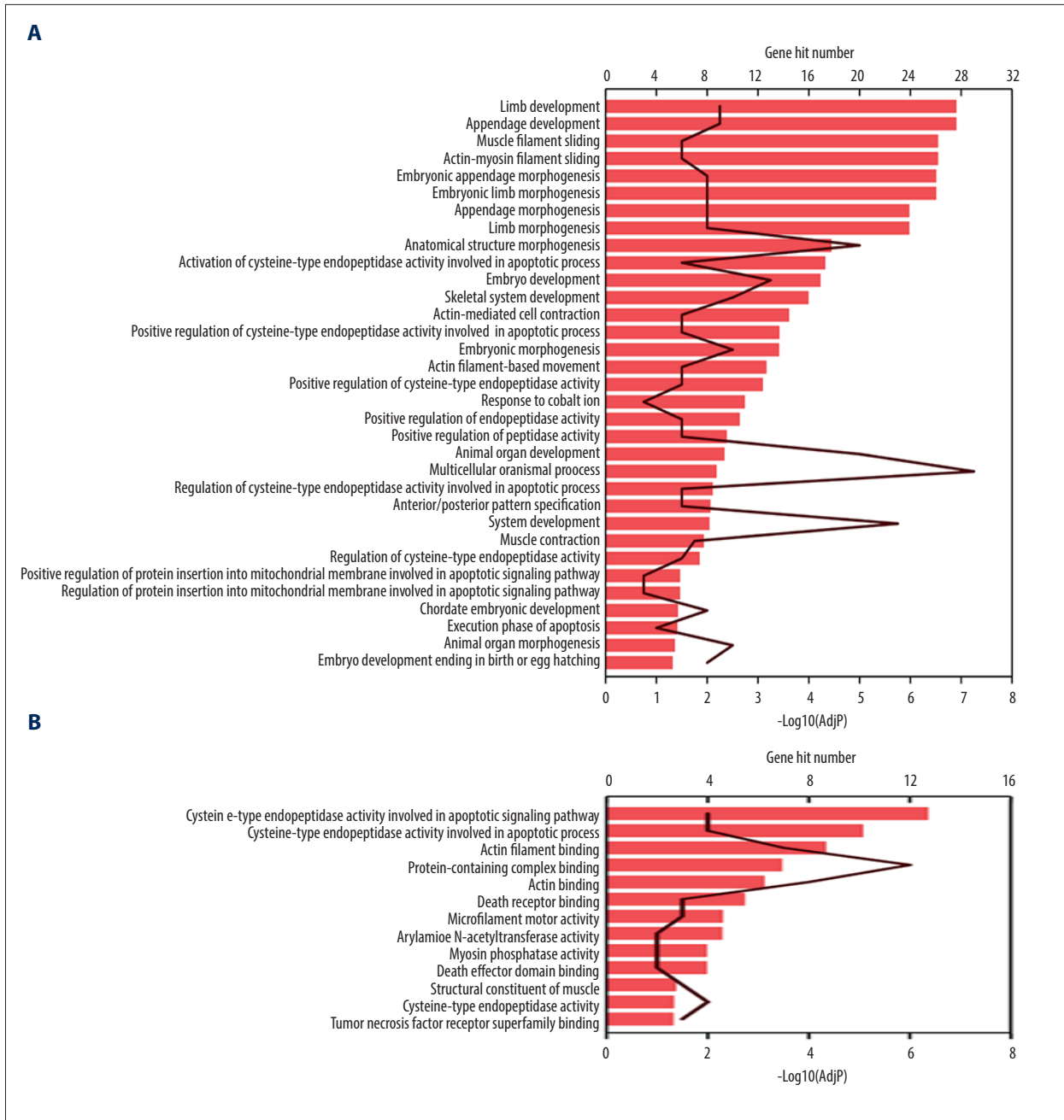
Supplementary Table 2. Top functions enriched.

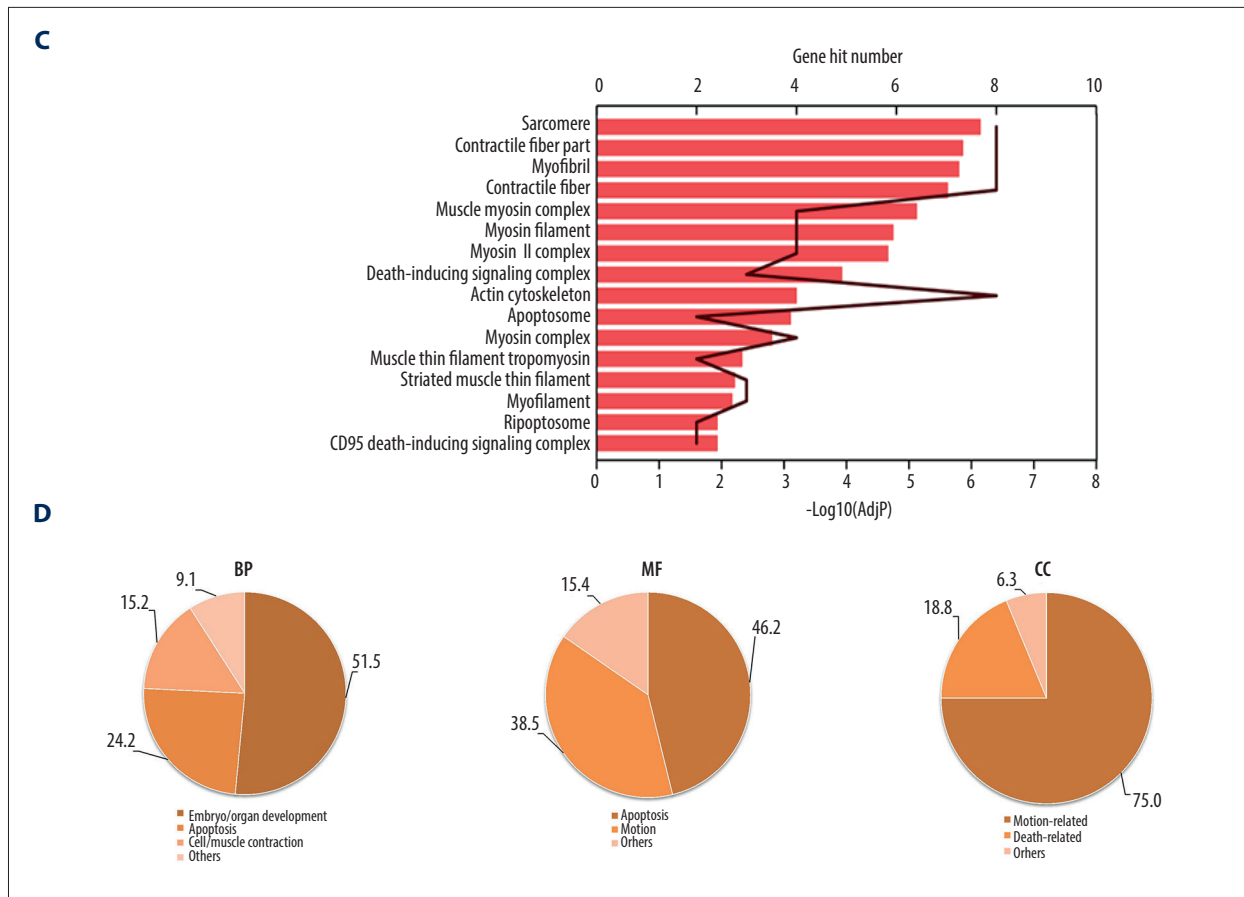
Database	Name	NetworkAnalyst	WebGestalt	g: Profiler	Average
REAC	Apoptosis/PCD	81.8	86.4	84.2	84.1
	Muscle contraction	9.1	9.1	10.5	9.6
	Metabolism	4.5	0.0	5.3	3.3
	Immune system	4.5	4.5	0.0	3.0
KEGG	Disease	64.3	75.0	75.0	71.4
	Apoptosis/PCD	7.1	10.0	12.5	9.9
	Cell cycle	7.1	5.0	6.3	6.1
	Metabolism	7.1	5.0	6.3	6.1
	Cell-cell communication	7.1	5.0	0.0	4.1
	Muscle contraction	7.1	0.0	0.0	2.4
WP	Disease	ND	ND	36.8	36.8
	Apoptosis/PCD	ND	ND	26.3	26.3
	DNA damage/repair	ND	ND	15.8	15.7
	Muscle contraction	ND	ND	5.3	5.3
	Immune system	ND	ND	5.3	5.3
	Cell response	ND	ND	5.3	5.3
	Signal Transduction	ND	ND	5.3	5.3
		REAC	KEGG	WP	Average
Mixed	Apoptosis/PCD	84.1	9.9	26.3	40.1
	Disease	0.0	71.4	36.8	36.1
	Muscle contraction	9.6	2.4	5.3	5.8
	DNA damage/repair	0.0	0.0	15.7	5.2
	Metabolism	3.3	6.1	0.0	3.1
	Immune system	3.0	0.0	5.3	2.8
	Cell cycle	0.0	6.1	0.0	2.0
	Cell response	0.0	0.0	5.3	1.8
	Signal Transduction	0.0	0.0	5.3	1.8
	Cell-cell communication	0.0	4.1	0.0	1.4





Supplementary Figure 1. Pathway enrichment analysis. (A) REAC by WebGestalt. (B) REAC by g:Profiler. (C) KEGG by WebGestalt. (D) KEGG by g:Profiler. Dark color was $-\log_{10}(\text{Adj}P) > 50$. The $-\log_{10}(\text{Adj}P)$ value was presented as the histogram and the hit gene numbers was presented as line chart. REAC – Reactome; KEGG – Kyoto Encyclopedia of Genes and Genomes.





Supplementary Figure 2. GO enriched analysis with g: Profiler. (A) BP, (B) MF, and (C) CC. The $-\log_{10}(\text{Adj}P)$ value was presented as the histogram and the hit gene numbers was presented as line chart. (D) Summary results of GO enriched analysis. GO – Gene Ontology; BP – biological process; MF – molecular function; CC – cellular component.

References:

- Wang H, Barisic I, Loane M et al: Congenital clubfoot in Europe: A population-based study. *Am J Med Genet A*, 2019; 179(4): 595–601
- Gurnett CA, Boehm S, Connolly A et al: Impact of congenital talipes equinovarus etiology on treatment outcomes. *Dev Med Child Neurol*, 2008; 50(7): 498–502
- Basit S, Khoshhal KI: Genetics of clubfoot; Recent progress and future perspectives. *Eur J Med Genet*, 2018; 61(2): 107–13
- Lovell ME, Morcuende JA: Neuromuscular disease as the cause of late clubfoot relapses: report of 4 cases. *Iowa Orthop J*, 2007; 27: 82–84
- Hester TW, Parkinson LC, Robson J et al: A hypothesis and model of reduced fetal movement as a common pathogenic mechanism in clubfoot. *Med Hypotheses*, 2009; 73(6): 986–88
- Wang Z, Yan N, Liu L et al: SOX9 overexpression plays a potential role in idiopathic congenital talipes equinovarus. *Mol Med Rep*, 2012; 7(3): 821–25
- Alvarado DM, Buchan JG, Frick SL et al: Copy number analysis of 413 isolated talipes equinovarus patients suggests role for transcriptional regulators of early limb development. *Eur J Hum Genet*, 2013; 21(4): 373–80
- Ester AR, Tyerman G, Wise CA et al: Apoptotic gene analysis in idiopathic talipes equinovarus (clubfoot). *Clin Orthop Relat Res*, 2007; 462: 32–37
- Wang LL, Jin CL, Liu LY et al: [Analysis of association between 5' HOXD gene and idiopathic congenital talipes equinovarus.] *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, 2005; 22(6): 653–56 [in Chinese]
- Shrimpton AE, Levinsohn EM, Yozawitz JM et al: A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-tooth disease. *Am J Hum Genet*, 2004; 75(1): 92–96
- Vito P, Emanuele C, Andrea V et al: The etiology of idiopathic congenital talipes equinovarus: A systematic review. *J Orthop Surg Res*, 2018; 13(1): 206
- Miedzybrodzka Z: Congenital talipes equinovarus (clubfoot): A disorder of the foot but not the hand. *J Anat*, 2003; 202(1): 37–42
- Sodre H, Bruschini S, Mestriner LA: Arterial abnormalities in talipes equinovarus as assessed by angiography and the Doppler technique. *J Pediatr Orthop*, 1990; 10(1): 101–4
- Chesney D, Barker S, Maffulli N: Subjective and objective outcome in congenital clubfoot; A comparative study of 204 children. *BMC Musculoskeletal Disord*, 2007; 8: 53
- Eckhardt A, Novotny T, Doubkova M: Novel contribution to clubfoot pathogenesis: The possible role of extracellular matrix proteins. *J Orthop Res*, 2019; 37(3): 769–78
- Cai G, Yang X, Chen T et al: Integrated bioinformatics analysis of potential pathway biomarkers using abnormal proteins in clubfoot. *Peer J*, 2020; 8: e8422
- Zhang Z, Kong Z, Zhu M et al: Whole genome sequencing identifies ANXA3 and MTHFR mutations in a large family with an unknown equinus deformity associated genetic disorder. *Mol Biol Rep*, 2016; 43(10): 1147–55
- Heck AL, Bray MS, Scott A et al: Variation in CASP10 gene is associated with idiopathic talipes equinovarus. *J Pediatr Orthop*, 2005; 25(5): 598–602
- Ester AR, Weymouth KS, Burt A et al: Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. *Am J Med Genet A*, 2009; 149(12): 2745–52

20. Zhao XL, Wang YJ, Wu YL et al: Role of COL9A1 genetic polymorphisms in development of congenital talipes equinovarus in a Chinese population. *Genet Mol Res*, 2016; 15(4): gmr15048773
21. Yang H, Zheng Z, Cai H et al: Three novel missense mutations in the filamin B gene are associated with isolated congenital talipes equinovarus. *Hum Genet*, 2016; 135(10): 1181–89
22. Zhang TX, Haller G, Lin P et al: Genome-wide association study identifies new disease *loci* for isolated clubfoot. *J Med Genet*, 2014; 51(5): 334–39
23. Zhang X, Jin CL, Liu LY et al: [Association and mutation analysis of GLI3 gene in idiopathic congenital talipes equinovarus.] *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2006; 23(5): 551–54 [in Chinese]
24. Weymouth KS, Blanton SH, Powell T et al: Functional assessment of clubfoot associated HOXA9, TPM1, and TPM2 variants suggests a potential gene regulation mechanism. *Clin Orthop Relat Res*, 2016; 474(7): 1726–35
25. Alvarado DM, McCall K, Hecht JT et al: Deletions of 5' HOXC genes are associated with lower extremity malformations, including clubfoot and vertical talus. *J Med Genet*, 2016; 53(4): 250–55
26. Dobbs MB, Gurnett CA, Pierce B et al: HOXD10 M319K mutation in a family with isolated congenital vertical talus. *J Orthop Res*, 2006; 24(3): 448–53
27. Shyy W, Wang K, Sheffield VC et al: Evaluation of embryonic and perinatal myosin gene mutations and the etiology of congenital idiopathic clubfoot. *J Pediatr Orthop*, 2010; 30(3): 231–34
28. Hecht JT, Ester A, Scott A et al: NAT2 variation and idiopathic talipes equinovarus (clubfoot). *Am J Med Genet A*, 2007; 143(19): 2285–91
29. Alvarado DM, McCall K, Aferol H et al: Pitx1 haploinsufficiency causes clubfoot in humans and a clubfoot-like phenotype in mice. *Hum Mol Genet*, 2011; 20(20): 3943–52
30. Lu W, Bacino CA, Richards BS et al: Studies of TBX4 and chromosome 17q23.1q23.2: An uncommon cause of nonsyndromic clubfoot. *Am J Med Genet A*, 2012; 158(7): 1620–27
31. Peterson JF, Ghaloul-Gonzalez L, Madan-Khetarpal S et al: Familial microduplication of 17q23.1-q23.2 involving TBX4 is associated with congenital clubfoot and reduced penetrance in females. *Am J Med Genet A*, 2014; 164(2): 364–69
32. Weymouth KS, Blanton SH, Bamshad MJ et al: Variants in genes that encode muscle contractile proteins influence risk for isolated clubfoot. *American Am J Med Genet A*, 2011; 155(9): 2170–79
33. Li C, Nguyen Q, Cole WG et al: Potential treatment for clubfeet based on growth factor blockade. *J Pediatr Orthop*, 2001; 21(3): 372–77
34. Herceg MB, Weiner DS, Agamanolis DP et al: Histologic and histochemical analysis of muscle specimens in idiopathic talipes equinovarus. *J Pediatr Orthop*, 2006; 26(1): 91–93
35. Kerling A, Stoltenburg-Didinger G, Grams L et al: The congenital clubfoot – immunohistological analysis of the extracellular matrix. *Orthop Res Rev*, 2018; 10: 55–62
36. Hu L, Papoian GA: Molecular transport modulates the adaptive response of branched actin networks to an external force. *J Phys Chem B*, 2013; 117(42): 13388–96
37. Esteban-Martínez L, Sierra-Filardi E, Boya P: Mitophagy, metabolism, and cell fate. *Mol Cell Oncol*, 2017; 4(5): e1353854
38. Seller MJ: Vitamins, folic acid and the cause and prevention of neural tube defects. *Ciba Found Symp*, 1994; 181: 161–79
39. Werler MM, Yazdy MM, Kasser JR et al: Maternal cigarette, alcohol, and coffee consumption in relation to risk of clubfoot. *Paediatr Perinat Epidemiol*, 2015; 29(1): 3–10
40. Werler MM, Yazdy MM, Kasser JR et al: Medication use in pregnancy in relation to the risk of isolated clubfoot in offspring. *Am J Med Genet A*, 2014; 180(1): 86–93
41. Wynne-Davies R: Family studies and aetiology of club foot. *J Med Genet*, 1965; 2(4): 227–32
42. Stewart SF: Clubfoot: Tts incidence, cause, and treatment;an anatomical-physiological study. *J Bone Joint Surg Am*, 1951; 33-A(3): 577–90