



Integrated bioinformatics analysis of potential pathway biomarkers using abnormal proteins in clubfoot

Guiquan Cai¹, Xuan Yang², Ting Chen², Fangchun Jin², Jing Ding² and Zhenkai Wu²

¹Department of Orthopaedics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

²Department of Pediatric Orthopaedics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

Background. As one of the most common major congenital distal skeletal abnormalities, congenital talipes equinovarus (clubfoot) affects approximately one in one thousandth newborns. Although several etiologies of clubfoot have been proposed and several genes have been identified as susceptible genes, previous studies did not further explore signaling pathways and potential upstream and downstream regulatory networks. Therefore, the aim of the present investigation is to explore abnormal pathways and their interactions in clubfoot using integrated bioinformatics analyses.

Methods. KEGG, gene ontology (GO), Reactome (REAC), WikiPathways (WP) or human phenotype ontology (HP) enrichment analysis were performed using WebGestalt, g:Profiler and NetworkAnalyst.

Results. A large number of signaling pathways were enriched e.g. signal transduction, disease, metabolism, gene expression (transcription), immune system, developmental biology, cell cycle, and ECM. Protein-protein interactions (PPIs) and gene regulatory networks (GRNs) analysis results indicated that extensive and complex interactions occur in these proteins, enrichment pathways, and TF-miRNA coregulatory networks. Transcription factors such as SOX9, CTNNB1, GLI3, FHL2, TGFBI and HOXD13, regulated these candidate proteins.

Conclusion. The results of the present study supported previously proposed hypotheses, such as ECM, genetic, muscle, neurological, skeletal, and vascular abnormalities. More importantly, the enrichment results also indicated cellular or immune responses to external stimuli, and abnormal molecular transport or metabolism may be new potential etiological mechanisms of clubfoot.

Subjects Bioinformatics, Orthopedics, Pathology, Pediatrics, Computational Science

Keywords Congenital talipes equinovarus, Clubfoot, Etiological mechanisms, Bioinformatics analysis, Enrichment analysis

INTRODUCTION

As a common developmental malformation of newborns, clubfoot affects approximately 2% of newborns (*Wang et al., 2019*). If not actively and timely treated, this deformity will accompany the child for a lifetime, which will not only affect the appearance of the child, including walking difficulties or even a disability, but also cause serious adverse

Submitted 19 August 2019
Accepted 17 December 2019
Published 20 January 2020

Corresponding authors
Jing Ding,
dingjing@xinhumed.com.cn
Zhenkai Wu,
wuzhenkai@xinhumed.com.cn

Academic editor
Pedro Silva

Additional Information and
Declarations can be found on
page 14

DOI 10.7717/peerj.8422

© Copyright
2020 Cai et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

effects on the mental health of the child. Therefore, timely and active treatments are urgently needed. The treatment of clubfoot includes surgical and nonsurgical treatments. Surgical releases, such as muscle balance and anterior tibial tendon transposition, may cause many problems, such as large trauma complications, which seriously affect the healthy development of children (*Smith et al., 2014*). The number of extensive surgical cases of clubfoot has declined by 60% from 1996 to 2006 in the United States (*Zionts et al., 2010*). In contrast, the Ponseti method has received increasing attention in the treatment of clubfoot and has become the most commonly used method. Although the success rate of Ponseti's first correction is satisfactory after a long period of development, there are still many problems and challenges, such as having a recurrence of deformity, having a poor compliance, and determining the best time to stop wearing braces (*Miller et al., 2016; Thacker et al., 2005*).

Clubfoot is associated with neuromuscular lesions, heredity abnormalities, skeletal dysplasia, soft tissue contracture, vascular abnormalities, ECM abnormalities, and intrauterine growth retardation (*Chesney, Barker & Maffulli, 2007; Eckhardt et al., 2019; Miedzybrodzka, 2003; Ošťádal et al., 2015; Poon, Li & Alman, 2009; Sodre et al., 1990; Wang et al., 2013a; Wang et al., 2013b*). In addition, smoking and viral infections in pregnant women are also closely related to congenital clubfoot (*Palma et al., 2013; Robertson Jr & Corbett, 1997*). Although the direct cause of congenital clubfoot remains to be unified, these findings provide an important opportunity and basis for the treatments of clubfoot. Several studies have revealed that botulinum toxin injection can relieve muscle or soft tissue contracture and may replace percutaneous tendoachilles tenotomy in the treatment of clubfoot (*Alvarez et al., 2009; Howren, Jamieson & Alvarez, 2015*). Based on the progress of pathogenesis, appropriate drug treatment may improve patient compliance and ultimately improve the efficacy of the Ponseti method. In addition, familial occurrence and inter- and intraphenotypic variability of clubfoot is well documented (*Basit & Khoshhal, 2018*). Several genes were identified as susceptible genes in clubfoot, such as the HOX family, CASP family, PITX-TXB4 pathway, troponin (TN) family, GLI3, T-box and MTHFR genes (*Hecht et al., 2007; Shrimpton et al., 2004; Shyy et al., 2009; Weymouth et al., 2016; Zhang et al., 2016*). However, these studies do not further explore abnormally active signaling pathways and potential upstream and downstream regulatory networks. Therefore, the aim of the present investigation is to explore the abnormal pathways and their interactions using integrated bioinformatics analysis in clubfoot. We expect the results of this study will provide an update on the etiopathogenetic mechanism of idiopathic clubfoot.

MATERIALS AND METHODS

Inclusion of abnormal proteins

Two widely used databases, PubMed and Science Direct, were used for literature retrieval. Keywords were “(clubfoot or clubfeet or congenital talipes equinovarus) and (etiology or embryology or etiopathogenesis or genomics or genetics or pathology or pathophysiology)”. The search date was up to May 7, 2019. A total of 1,057 articles were found in PubMed, and 657 articles were found in Science Direct. After eliminating duplicate articles, a

total of 1,093 studies were retrieved. By reading the title and abstract, 1,015 articles were eliminated following the inclusion and exclusion criteria as described below. Inclusion criteria included clinical or preclinical studies written in English that were focused on the etiology of clubfeet. Investigations with a focus on secondary/syndromic clubfoot, such as distal arthrogryposis, myelomeningocele, and moebius syndrome, were excluded. We performed a full text assessment of the remaining 78 articles. A total of 47 articles were excluded because the study topic was not clubfoot, genetic information was not mentioned or full text of the study was not available. Finally, 8 articles were included in the present study. A total of 30 proteins that are shown in [Table 1](#) were used in this investigation after assessing the protein. The process of the inclusion of abnormal proteins was shown in [Fig. 1](#).

WebGestalt analysis

WebGestalt analysis was performed as described previously ([Wang et al., 2017](#)). The parameters for the enrichment analysis were as follows. A specific organism was chosen *H. sapiens* (human). Enrichment categories were used KEGG, Panther and Reactome pathways. A reference list was used for all mapped entrezgene IDs from the selected platform genome. The reference list was mapped to 61,506 entrezgene IDs, and 2266 IDs were annotated to the selected functional categories and used as the reference for the enrichment analysis. The minimum number of IDs in each category was 5, and the maximum number of IDs was 2000. Among 30 unique entrezgene IDs, 19 IDs were annotated to the selected functional categories and used for the enrichment analysis. The Gene Ontology (GO) Slim summary was based upon 30 unique entrezgene IDs. Fisher's exact test-based overrepresentation enrichment analysis (ORA) method was used for enrichment analysis. FDR was used for the Benjamani-Hochberg (BH) method.

g:Profiler analysis

The version of g:Profiler was e94_eg41_p11_9f195a1 (database updated on 01/24/2019). The parameters for the enrichment analysis were as follows. A specific organism was chosen *H. sapiens* (human). GO analyses (GO molecular function (GO: MF), GO cellular component (GO: CC), and GO biological process (GO: BP)) were carried out sequentially. The biological pathways used were the KEGG, Reactome, and WikiPathways databases. The protein databases used were the Human Protein Atlas and CORUM databases. The statistical domain scope was used only for annotated genes. The significance threshold was the g:SCS threshold. The user threshold was 0.05.

NetworkAnalyst analysis

The significantly changed genes from the previous analyses were mapped to the corresponding molecular interaction databases. The procedure typically produces one large subnetwork with several smaller ones. The website was upgraded and maintained until May 8, 2019 by the Xia Lab (<https://www.networkanalyst.ca/NetworkAnalyst/faces/home.xhtml>). The parameters for the enrichment analysis were as follows. A specific organism was chosen *H. sapiens* (human). The ID type was chosen Uniprot Protein ID. PPI analysis was performed using the MEx Interactome database. The parameters were referred

Table 1 Protein information.

Entry ID	Entry name	Protein names	Gene names	References
Q9BXN1	ASPN_HUMAN	Asporin	ASPN	<i>Ošťádal et al. (2015)</i>
P21810	PGS1_HUMAN	Biglycan	BGN	<i>Ošťádal et al. (2015)</i>
Q8IUL8	CILP2_HUMAN	Cartilage intermediate layer protein 2, CILP-2	CILP2	<i>Eckhardt et al. (2019)</i>
P35222	CTNB1_HUMAN	Catenin beta-1 (Beta-catenin)	CTNNB1	<i>Poon, Li & Alman (2009)</i>
P02461	CO3A1_HUMAN	Collagen alpha-1(III) chain	COL3A1	<i>Eckhardt et al. (2019)</i>
P20908	CO5A1_HUMAN	Collagen alpha-1(V) chain	COL5A1	<i>Eckhardt et al. (2019)</i>
P12109	CO6A1_HUMAN	Collagen alpha-1(VI) chain	COL6A1	<i>Eckhardt et al. (2019)</i>
Q99715	COCA1_HUMAN	Collagen alpha-1(XII) chain	COL12A1	<i>Eckhardt et al. (2019)</i>
Q05707	COEA1_HUMAN	Collagen alpha-1(XIV) chain (Undulin)	COL14A1	<i>Eckhardt et al. (2019)</i>
P08123	CO1A2_HUMAN	Collagen alpha-2(I) chain	COL1A2	<i>Eckhardt et al. (2019)</i>
P12110	CO6A2_HUMAN	Collagen alpha-2(VI) chain	COL6A2	<i>Eckhardt et al. (2019)</i>
P12111	CO6A3_HUMAN	Collagen alpha-3(VI) chain	COL6A3	<i>Eckhardt et al. (2019)</i>
P20849	CO9A1_HUMAN	Collagen alpha-1(IX) chain	COL9A1	<i>Wang et al. (2013a) and Wang et al. (2013b)</i>
O94907	DKK1_HUMAN	Dickkopf-related protein 1, Dickkopf-1	DKK1	<i>Poon, Li & Alman (2009)</i>
Q06828	FMOD_HUMAN	Fibromodulin	FMOD	<i>Eckhardt et al. (2019)</i>
P02751	FN1_HUMAN	Fibronectin (Cold-insoluble globulin, CIG)	FN1	<i>Ošťádal et al. (2015)</i>
Q14192	FHL2_HUMAN	Four and a half LIM domains protein	FHL2	<i>Wang et al. (2008)</i>
P35453	HXD13_HUMAN	Homeobox protein Hox-D13 (Homeobox protein Hox-4I)	HOXD13	<i>Wang et al. (2008)</i>
P51884	LUM_HUMAN	Lumican	LUM	<i>Ošťádal et al. (2015)</i>
Q16853	AOC3_HUMAN	Membrane primary amine oxidase	AOC3	<i>Ošťádal et al. (2015)</i>
P20774	MIME_HUMAN	Mimecan (Osteoglycin)	OGN	<i>Ošťádal et al. (2015)</i>
Q15063	POSTN_HUMAN	Periostin (Osteoblast-specific factor 2)	POSTN	<i>Ošťádal et al. (2015)</i>
P51888	PRELP_HUMAN	Prolargin	PRELP	<i>Ošťádal et al. (2015)</i>
Q7Z7G0	TARSH_HUMAN	Target of Nesh-SH3	ABI3BP	<i>Ošťádal et al. (2015)</i>
P24821	TENA_HUMAN	Tenascin, TN (Cytotactin)	TNC	<i>Eckhardt et al. (2019) and Ošťádal et al. (2015)</i>
Q15582	BGH3_HUMAN	TGF- β -induced protein ig-h3	TGFBI	<i>Eckhardt et al. (2019)</i>
P48436	SOX9_HUMAN	Transcription factor SOX-9	SOX9	<i>Wang et al. (2013a) and Wang et al. (2013b)</i>
P10071	GLI3_HUMAN	Transcriptional activator GLI3	GLI3	<i>Cao et al. (2009)</i>
P01137	TGFB1_HUMAN	TGF- β -1 proprotein	TGFB1	<i>Wang et al. (2013a) and Wang et al. (2013b)</i>
P13611	CSPG2_HUMAN	Versican core protein	VCAN	<i>Eckhardt et al. (2019) and Ošťádal et al. (2015)</i>

to the literature-curated comprehensive data from the Innate DB (*Breuer et al., 2013*). Gene-miRNA interactome analysis was carried out with comprehensive experimentally validated miRNA-gene interaction data collected from TarBase and miRTarBase (*Chou et al., 2018; Vlachos et al., 2015*). TF-gene interaction analysis was performed using the ENCODE database. Transcription factor and gene target data derived from the ENCODE ChIP-seq data. The peak intensity signal <500 and the predicted regulatory potential

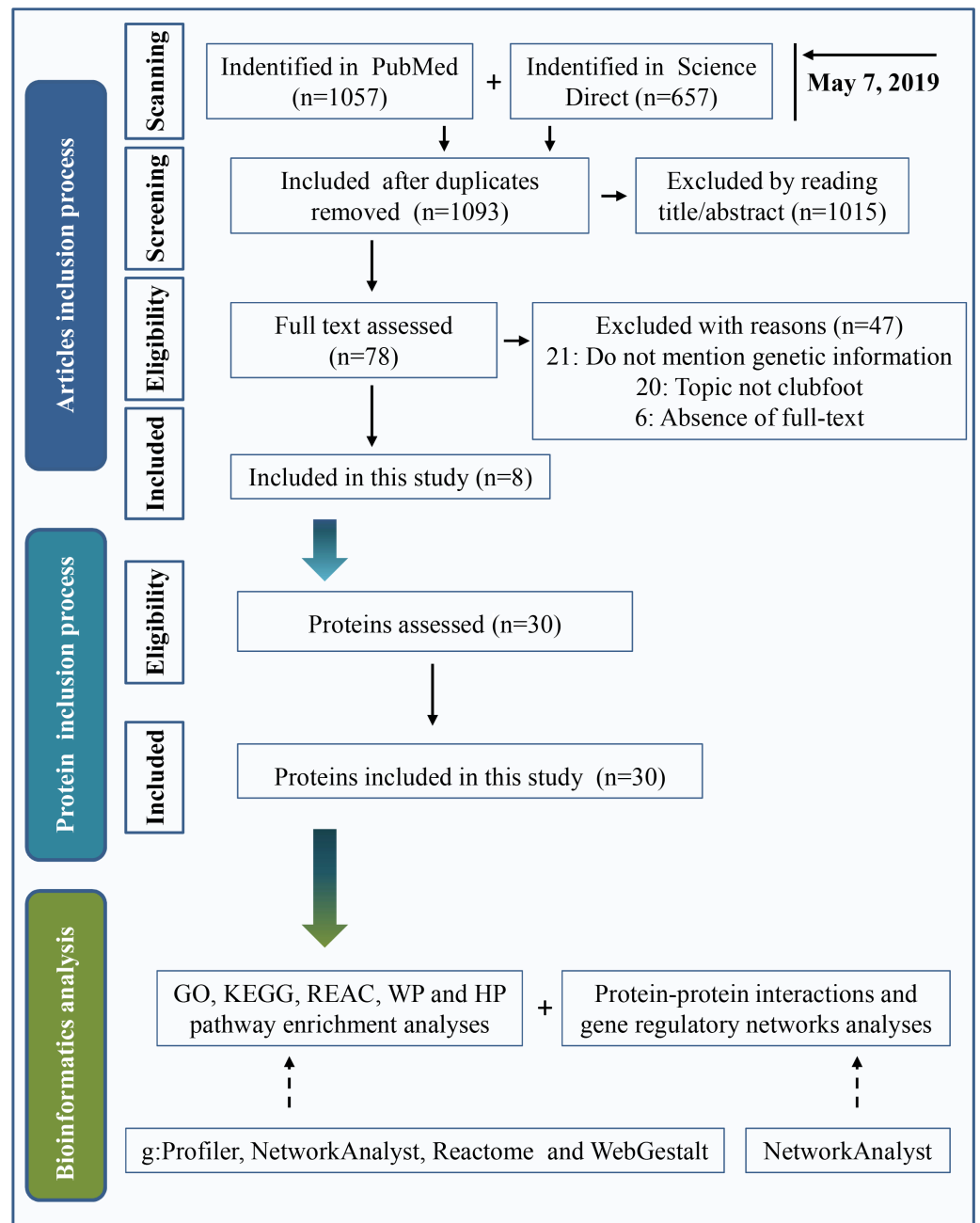


Figure 1 Inclusion criteria for abnormal protein candidates and the process of bioinformatics analysis.

Full-size  DOI: [10.7717/peerj.8422/fig-1](https://doi.org/10.7717/peerj.8422/fig-1)

score <1 used the BETA Minus algorithm (Wang et al., 2013a; Wang et al., 2013b). TF-miRNA coregulatory network analysis was performed using curated regulatory interaction information collected from the RegNetwork repository (Liu et al., 2015).

REAC enrichment analysis

The enrichment analysis was performed against Reactome version 66 on 04/05/2019 using UniProt identifiers for the mapping. The web link is as follows: <https://reactome.org/PathwayBrowser>. Information in the REAC database is authored by expert biologists and entered and maintained by Reactome's team of curators and editorial staff. Reactome content frequently cross-references other resources, e.g., NCBI, Ensembl, UniProt, KEGG (Gene and Compound), ChEBI, PubMed and GO. REAC analysis was performed as described previously (*Fabregat et al., 2016*). The parameters for the enrichment analysis were as follows. A specific organism was chosen *H. sapiens* (human). IntAct interactors were included to increase the analysis background. An overrepresentation analysis method was used for enrichment analysis. This test produces a probability score, which was corrected for the false discovery rate using the BH method. Twenty-seven out of 30 identifiers in the sample were found in Reactome, and 430 pathways were found by at least one of the identifiers.

Statistics

The enrichment analysis method in the WebGestalt analysis was used for ORA method. The significance threshold in the g:Profiler analysis was the g:SCS threshold (g:Profiler analysis soft version: e94_eg41_p11_9f195a1 (database updated on 01/24/2019)). The adjusted *p* value method used was the BH method and the adjusted *p* value was transformed to negative log₁₀ (-log₁₀(AdjP)). All significantly changed pathways and interactions had a *p* value <0.05.

RESULTS

Results of the enrichment analysis by g:Profiler and Reactome

To visually observe the enrichment information of these candidate proteins, g:Profiler and Reactome were performed for the bioinformatics analysis. A large number of terms were enriched by g:Profiler in the GO, KEGG, REAC, WP and HP databases (Fig. 2A). A large number of signaling pathways were enriched by Reactome (Fig. 2B). Additionally, there were extensive interactions among these signaling pathways (Fig. 2B).

To clarify the categories of these pathways, we summarized these pathways enriched by REAC. Signal transduction, disease, metabolism, gene expression (transcription), and immune system were the top 5 pathways and their percentages were as high as 65% (Fig. 3A). Among the 30 proteins, the top 10 proteins with the widest participation, such as P35222, P01137, P02751 and P08123, were mainly ECM proteins or proteins that interact with the ECM (Figs. 3B and 3C). We continued to summarize the top 10 pathways and found that these pathways were mainly focused on the ECM, metabolism and cell communication (Fig. 3D).

Results of the signaling pathway enrichment analysis

Through the Reactome enrichment analysis, a large number of pathways were enriched (Figs. 2 and 3). To further analyze these potential pathways, signaling pathway enrichment analysis was carried out using g:Profiler in the REAC, WP, KEGG and HP. A total of 32, 11,

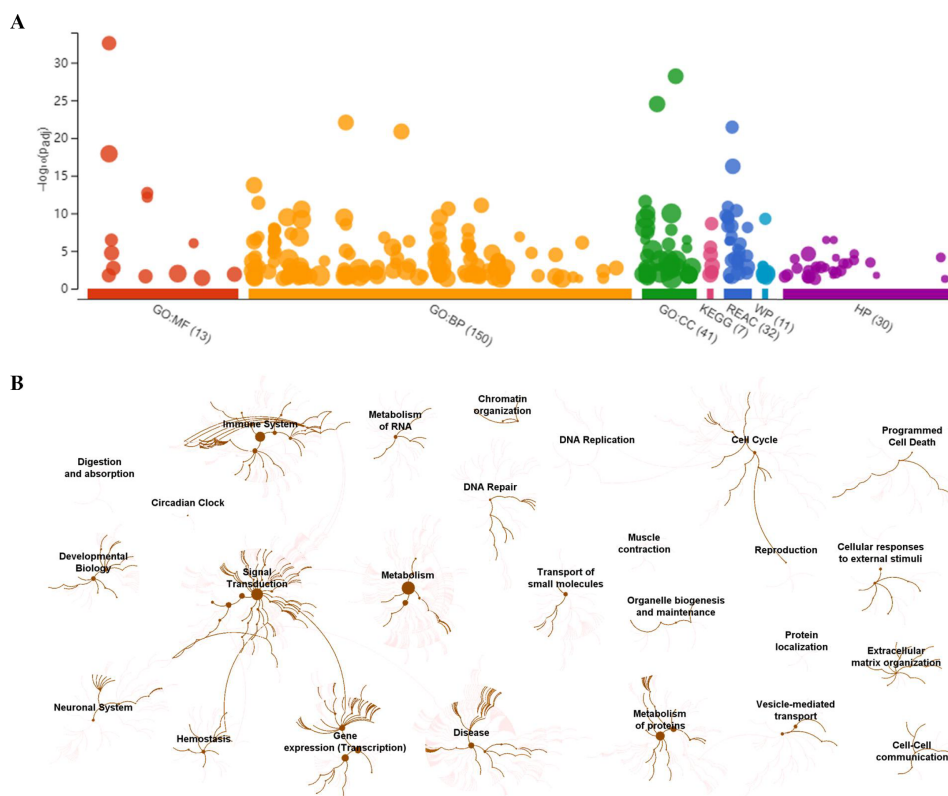


Figure 2 Overall results of bioinformatics analyses with candidate proteins using g:Profiler and Reactome. (A) The significantly changed terms enriched by GO, KEGG, Reactome (REAC), Wikpathways (WP) and Human phenotype ontology (HP) databases. (B) The enrichment pathways and their interactions enriched by REAC.

Full-size DOI: 10.7717/peerj.8422/fig-2

7 and 30 pathways were enriched by g:Profiler, respectively (Figs. 4A–4D). Classification analysis revealed that these pathways were mainly concentrated in ECM, disease and metabolic pathways (Fig. 4E). The proportion of the top three was as high as 78%. In addition, human phenotype ontology enrichment analysis results revealed that these abnormal proteins were mainly expressed in the lower appendages, such as muscle, ankle, foot, joint, skin and connective tissue, and their proportion was more than 50% (Fig. 4F). Similar, similar results were found in KEGG and REAC enrichment analyses using NetworkAnalyst and WebGestalt (Fig. S1).

Results of the GO enrichment analysis

GO enrichment analysis was further performed to explore the BP, MF and CC induced by these proteins. A total of 69, 9 and 24 GO terms were enriched by g:Profiler in BP, MF and CC, respectively (Figs. 5A–5C). Classification analysis revealed that GO: BP was mainly concentrated in embryo or organ development, and its combined percentage was over 80% (Fig. 5E). Among these pathways, skeletal development was dominant. ECM function, growth factor binding, cell adhesion, heparin binding and protein binding were the main pathways in MF, and their percentages were as high as 85% (Fig. 5F). Additionally, ECM

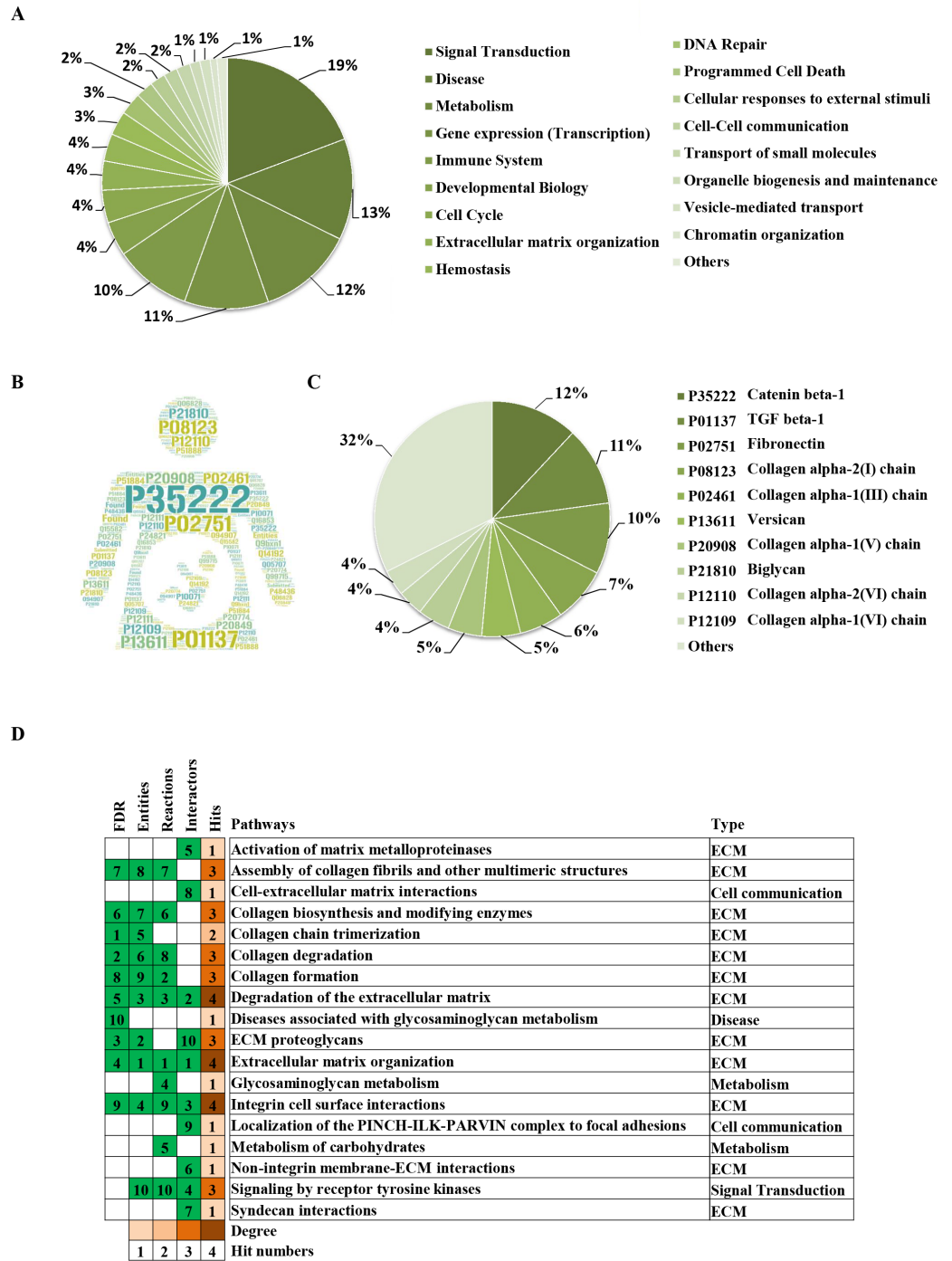


Figure 3 Classification statistics of significantly enriched pathways by REAC shown in Fig. 2B. Enrichment analysis methods are described in the Materials and Methods section. (A) Distribution of the enrichment pathways. (B) High frequency molecules in all of the significant signaling pathways were mapped by Wordart software (<https://wordart.com>). The larger the word frequency was, the larger the font size. (C) The top 10 proteins involved in enrichment pathways and their proportions. (D) The top 10 signaling pathways. Four columns on the left are the results of the top 10 FDR, entities, reactions and interactors. The numbers are their ranks. The hit numbers for each pathway are shown in column 5.

Full-size DOI: 10.7717/peerj.8422/fig-3

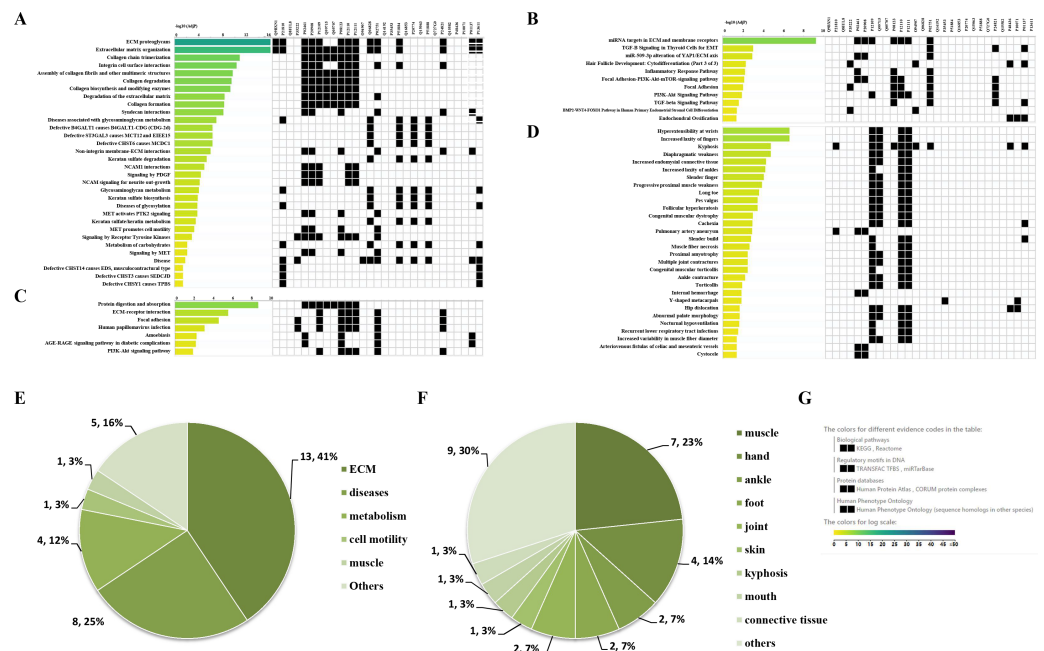


Figure 4 Signaling pathway enrichment analysis by g:Profiler in the REAC, WP, KEGG and HP databases. Signaling pathway enrichment analysis by g:Profiler in the (A) REAC, (B) WP, (C) KEGG and (D) HP databases. Statistical results of signaling pathways from REAC (E) and HP (F). (G) Graphical illustration.

Full-size [DOI: 10.7717/peerj.8422/fig-4](https://doi.org/10.7717/peerj.8422/fig-4)

and membrane were the main CC, and their percentages were almost 60% (Fig. 5G). In addition, GO enrichment analysis results analyzed by NetworkAnalyst were consistent with those from g:Profiler (Fig. S2).

Results of PPIs and gene regulatory networks analysis

The NetworkAnalyst analysis was performed to explore protein-protein interactions and gene regulatory networks for genes that code these candidate proteins. These genes interact extensively with predictive genes (Fig. 6A). A total of 2,630 PPIs were enriched by NetworkAnalyst. There are 1,263 interactions that were related to the top 10 genes and these interactions account for 48% of the total. The top genes were FN1, CTNNB1, FHL2, TGFB1 and COL1A2 (Fig. 6B). Among these genes, FN1 and CTNNB1 were dominant, and the percentage of PPIs induced by these two genes was almost 39%.

TF-gene interactions were also enriched by NetworkAnalyst. A total of 692 TF-gene interactions were enriched (Fig. 6C). The top genes were TGFB1, FN1, SOX9, AOC3 and HOXD13 (Fig. 6D). Additionally, the interactions among these 10 genes were as high as 43% of the total.

Gene-miRNA interactome analysis results revealed that 409 interactomes were identified between these genes and predictive miRNAs (Fig. 6E). The top genes were FHL2, PRELP, CTNNB1, SOX9 and COL12A1 (Fig. 6F). Top miRNAs were hsa-mir-124-3p, hsa-mir-26b-5p, hsa-mir-335-5p, hsa-mir-1-3p and hsa-mir-5698. The interactomes between the top 10 genes and the top 10 miRNAs accounted for 39% and 5% of the total, respectively.

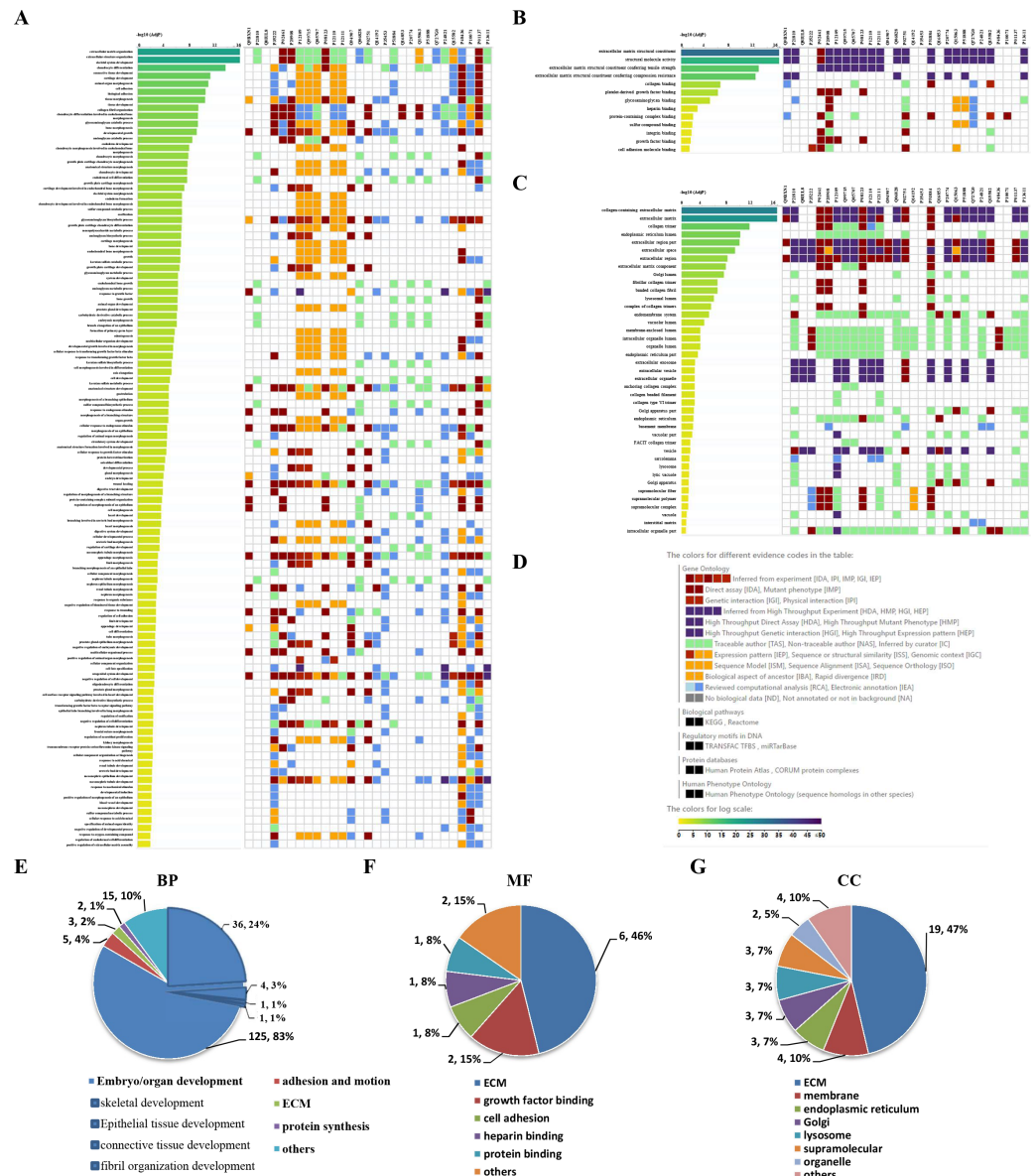


Figure 5 GO enrichment analysis by g:Profiler: BP, MF and CC terms. GO enrichment analysis by g:Profiler. (A) BP, (B) MF and (C) CC terms. (D) Graphical illustration. Statistical results from the GO BP (E), MF (F) and CC (G) enrichment analysis.

Full-size DOI: [10.7717/peerj.8422/fig-5](https://doi.org/10.7717/peerj.8422/fig-5)

A large number of TF-miRNA coregulatory interactions were enriched (Fig. 6G). Transcription factors, such as SOX9, CTNNB1, GLI3, FHL2, TGFBI and HOXD13, cooperated with hsa-miR-29a, hsa-miR-101, hsa-miR-520d-5p, hsa-miR-29b and hsa-miR-568 to regulate these candidate genes (Fig. 6G). The top 10 genes were SOX9, CTNNB1, GLI3, FHL2, TGFBI, TGFBI, VCAN, TNC, COL12A1 and COL5A1 (Fig. 6H).

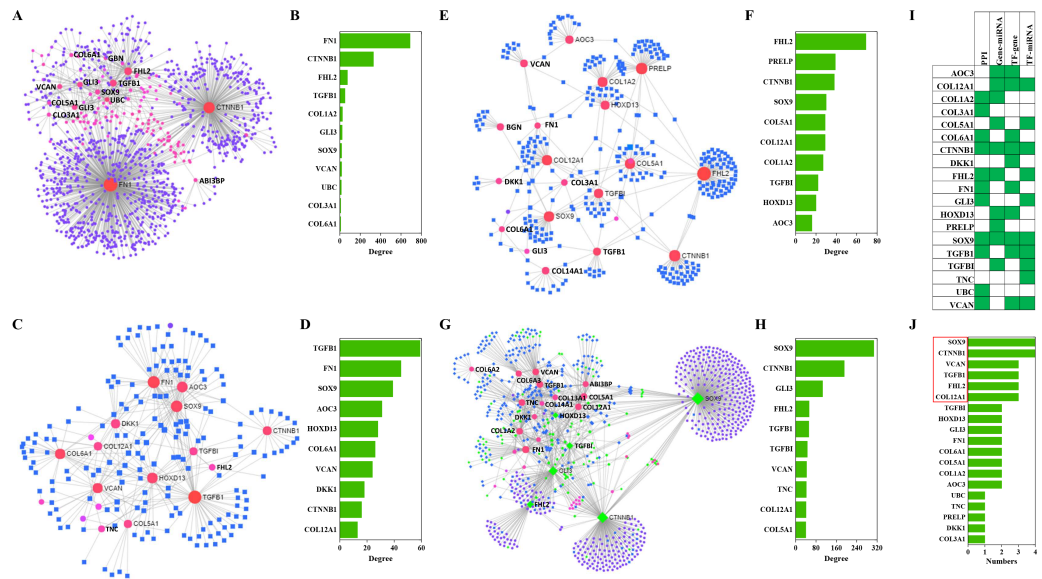


Figure 6 PPIs and GRNs analyses of these candidate protein-coding genes by NetworkAnalyst. Generic PPI (A) and the top 10 genes (B) involved in that PPI. TF-gene interactions (C) and the top 10 genes (D) involved in these interactions. Genes are shown as pink circles, and transcription factors are shown as blue squares. Gene-miRNA interactome (E) and the top 10 genes (F) involved in these interactions. Genes are shown as pink circles and miRNAs are shown as blue squares. TF-miRNA coregulatory interactions (G) and the top 10 genes (H) involved in these interactions. Genes are circles. miRNAs are shown as blue diamonds. TFs are shown as green diamonds. The distribution (I) and frequency (J) of the top 10 genes in these four interaction networks mentioned above are shown.

Full-size [DOI: 10.7717/peerj.8422/fig-6](https://doi.org/10.7717/peerj.8422/fig-6)

Combining the four interactions mentioned above, 19 genes were involved in the top interactions (Fig. 6I). SOX9, CCNNB1, VCAN, TGFB1, FHL2 and COL12A1 widely participated in the top-ranking interactions (Fig. 6J).

DISCUSSION

In the present investigation, a large number of signaling pathways were enriched in the GO, KEGG, REAC, WP and HP databases using g:Profiler, NetworkAnalyst and WebGestalt analyses of clubfoot. Among them, pathways in embryo or organ development, ECM, metabolism, immune system, cell cycle, cell responses to external stimuli, and apoptosis or programmed cell death were the top pathways. A wide range of interactions existed among these enriched signaling pathways. In addition, there were also extensive regulations between the upstream and downstream of genes encoding these proteins.

There were 452 enriched pathways identified by REAC enrichment analysis. Among them, signal transduction, disease, metabolism, gene expression (transcription), immune system, developmental biology, cell cycle, ECM and hemostasis were advantageous pathways. Additionally, signaling pathways in DNA repair, PCD, cell response to external stimuli, cell-cell communication, molecular transport and chromatin organization were enriched. In the process of cell biological activities, numerous changes occurred in cells, such as signal transduction, gene expression, cell cycle, DNA repair, molecular transport

and metabolism. Abnormal changes in these biological processes may alter cell fate, cell–cell communication, and cell response to external stimuli or even cause immune system changes.

Select soft tissues in clubfoot are contracted, resulting in stiffness. Extracellular matrix proteins, such as asporin, collagen types III, V, and VI, versican, tenascin-C, and TGF-beta induced protein, are highly expressed in contracted tissues clubfoot ([Eckhardt et al., 2019](#); [Ošťádal et al., 2015](#)). Additionally, the expression levels of growth factors TGF-beta and platelet-derived growth factor are high and a blockade of growth factors led to decreased collagen expression, proliferation, and chemotaxis ([Li et al., 2001](#)). These proteins seem to be promising targets for future investigations and treatments of this disease. Indeed, the muscle contraction strategy, such as botulinum toxin injection, can relieve muscle or soft tissue contracture and thus improve and alleviate clubfoot symptoms ([Howren, Jamieson & Alvarez, 2015](#); [Shrimpton et al., 2004](#)). ECM-associated pathways were enriched and included collagen chain trimerization, collagen formation and degradation, and ECM proteoglycans. TGF-beta receptor signaling was also enriched. These findings support the idea that these ECM proteins are promising targets for the treatment of clubfoot.

Abnormal biological development, such as muscle, neurological, skeletal and vascular abnormalities, has been previously identified and confirmed ([Basit & Khoshhal, 2018](#); [Eckhardt et al., 2019](#); [Herceg et al., 2006](#); [Hester et al., 2009](#); [Lovell & Morcuende, 2007](#); [Ošťádal et al., 2015](#)). The HOX and TBX families governed limb identity, and fibroblast growth factor participated in the formation of limb muscles ([Ohuchi & Noji, 1999](#); [Wang et al., 2008](#)). Additionally, gene-gene interactions between CASP SNPS and variants in HOXA, HOXD, and insulin-like growth factor binding protein affect muscle and limb development ([Ester et al., 2010](#)). Apoptosis and programmed cell death associated pathways and activation of HOX gene pathways, such as activation of HOX genes during differentiation, and activation of anterior HOX genes in hindbrain development during early embryogenesis, were enriched in the REAC database. A myogenesis pathway was also found.

ROBO family genes regulate axonal guidance and cell migration. ROBO1 and ROBO2 receptors regulate the proliferation and transition of primary to intermediate neuronal progenitors ([Borrell et al., 2012](#)), while the interaction of ROBO4 with SLIT3 is involved in the proliferation, motility and chemotaxis of endothelial cells, and accelerates the formation of blood vessels ([Zhang et al., 2009](#)). We found that the regulation of commissural axon pathfinding by SLIT, ROBO, netrin-1 signaling and pathways with signaling by ROBO receptors, semaphorin interactions, neurotransmitter release cycle were enriched. These results support that activation and inactivation of developmental signaling pathways initiate embryonic and organ development. Abnormalities in any link may cause deformities.

Epidemiological data confirmed that smoking by any parent or the presence of any household smoking increased the risk of clubfoot in Peru ([Palma et al., 2013](#)). Maternal diabetes also showed a significant association with clubfoot ([Parker et al., 2009](#)). In mice, maternal smoking or diabetes increased mitochondrial damage and oxidative stresses ([Stangenberg et al., 2015](#)). In addition, maternal diabetes also increased hypoxia-inducible factor 1 α expression in utero ([Moazzen et al., 2015](#)). [Robertson Jr & Corbett \(1997\)](#)

confirmed that clubfoot was caused by an intrauterine enterovirus through allowing anterior horn cell lesions. The adverse in utero environment caused by the external or internal environment may increase oxidative stress and thus induce cell damage in utero. We found that the signaling pathways involved in the regulation of gene expression by hypoxia-inducible factor, reactive oxygen species detoxification, cell response to hypoxia, cell response to heat stress, cell response to stress, oxidative stress-induced senescence, cellular senescence, and cellular responses to external stimuli were enriched in the REAC database. In addition, these external stimuli activate the immune system. We also found that interleukin family signaling, MAP kinase activation, toll-like receptor family cascade, adaptive immune system, and innate immune system were enriched. These data indicate that abnormal changes in the immune system and cell response to external stimuli may play key roles in clubfoot.

In the enrichment analysis, we also found that several molecular transport pathways were enriched by enrichment analysis, and induced plasma lipoprotein assembly, remodeling, and clearance, iron uptake and transport, small molecules transport, binding and uptake of ligands by scavenger receptors, lysosome vesicle biogenesis, vesicle-mediated transport, trans-Golgi network vesicle budding, and membrane trafficking. Molecular transport is essential for material transport, signal transduction and neurotransmitter transporters and modulates the cell response to an external force (*Hu & Papoian, 2013*). Disorders induced by abnormal molecular transport are serious and even fatal for cells. Based on these results, we proposed that molecular transport abnormalities may play a large role in clubfoot.

Metabolism of proteins, RNAs or other molecules plays a key role in the basic functions of cells (*Esteban-Martínez, Sierra-Filardi & Boya, 2017*). Pathways in these metabolic processes were found in the enrichment analyses, and induced regulation of IGF transport and uptake by IGFBPs, activation of genes by ATF4 in response to endoplasmic reticulum stress, synthesis, secretion, and inactivation of GLP-1, metabolism of nitric oxide, activation and regulation of eNOS, metabolism of carbohydrates, and regulation of lipid metabolism by PPAR α . A previous study indicated that lipid droplets and glycogen increased and the number of mitochondria decreased in chondrocytes from the biopsied iliac crest cartilage of joint contracture patients (*Nogami et al., 1983*). These data suggested that glucose metabolism, lipid metabolism and mitochondrial oxidative stress were associated with multiple joint contractures. In the present work, glucose metabolism regulated by IGF or GLP, lipid metabolism regulated by PPAR, and oxidative stress regulated by NOS were found in the REAC database. These data suggest that metabolic abnormalities may play a significant role in clubfoot.

All candidate papers were obtained from PubMed and Science Direct. Almost all of these differentially expressed proteins were carefully and strictly selected from clinical trials. Clubfoot is abnormal in both bone and muscle. Most of the samples are from bone, and few are from muscle, therefore selection bias cannot be avoided. Through strict screening criteria, candidate proteins from different tissues were adopted for subsequent bioinformatics analysis, and the selection bias was possibly minimized. Although the adjusted p value method was performed, there may be some false positives in using these candidate proteins for bioinformatics analysis because these proteins were not collected

from the same investigation. Although we proposed that cell or immune responses to external stimuli, abnormal molecular transport and metabolism are new potential etiopathogenetic mechanisms of clubfoot, direct experimental evidence is needed. We are carrying out preclinical and clinical studies to confirm these enrichment pathways and proposed hypotheses.

CONCLUSIONS

In summary, a large number of signaling pathways were enriched using REAC, KEGG and WP enrichment analyses by g:Profiler, NetworkAnalyst and WebGestalt. Among them, signal transduction, disease, metabolism, gene expression (transcription), immune system, developmental biology, cell cycle, and ECM were the top functions. GO enrichment analysis also revealed pathways in embryo or organ development, and ECM proteins were dominant. PPIs and GRNs analysis results indicated that extensive and complex interactions occur in these proteins, enrichment pathways, and TF-miRNA coregulatory networks. Transcription factors such as SOX9, CTNNB1, GLI3, FHL2, TGFBI and HOXD13, cooperated with hsa-miR-29a, hsa-miR-101, hsa-miR-520d-5p, hsa-miR-29b and hsa-miR-568 and regulated these candidate proteins. In addition to supporting the proposed hypotheses, such as ECM abnormalities, fetal movement reduction, genetic abnormalities, muscle abnormalities, neurological abnormalities, skeletal abnormalities, uterine compression and vascular abnormalities, we propose that cellular or immune responses to external stimuli, and abnormal molecular transport or metabolism are new potential etiological mechanisms of clubfoot.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Guiquan Cai and Xuan Yang performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ting Chen performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Fangchun Jin analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Jing Ding and Zhenkai Wu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability: Raw data is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8422#supplemental-information>.

REFERENCES

- Alvarez CM, De Vera MA, Chhina H, Williams L, Durlacher K, Kaga S. 2009. The use of botulinum type A toxin in the treatment of idiopathic clubfoot: 5-year follow-up. *Journal of Pediatric Orthopedics* 29(6):570–575 DOI 10.1097/bpo.0b013e3181b2b3d4.
- Basit S, Khoshhal KI. 2018. Genetics of clubfoot; recent progress and future perspectives. *European Journal of Medical Genetics* 61(2):107–113 DOI 10.1016/j.ejmg.2017.09.006.
- Borrell V, Cárdenas A, Ciceri G, Galcerán J, Flames N, Pla R, Nóbrega-Pereira S, García-Frigola C, Peregrín S, Zhao Z, Ma L, Tessier-Lavigne M, Marín O. 2012. Slit/Robo signaling modulates the proliferation of central nervous system progenitors. *Neuron* 76(2):338–352 DOI 10.1016/j.neuron.2012.08.003.
- Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Research* 41(Database issue):D1228–D1233 DOI 10.1093/nar/gks1147.
- Cao D, Jin C, Ren M, Lin C, Zhang X, Zhao N. 2009. The expression of Gli3, regulated by HOXD13, may play a role in idiopathic congenital talipes equinovarus. *BMC Musculoskeletal Disorders* 10:142 DOI 10.1186/1471-2474-10-142.
- Chesney D, Barker S, Maffulli N. 2007. Subjective and objective outcome in congenital clubfoot; a comparative study of 204 children. *BMC Musculoskeletal Disorders* 8:53 DOI 10.1186/1471-2474-8-53.
- Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai W, Huang PC, Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, Huang HD. 2018. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* 46(D1):D296–D302 DOI 10.1093/nar/gkx1067.
- Eckhardt A, Novotny T, Doubkova M, Hronkova L, Vajner L, Pataridis S, Hadraba D, Kulhava L, Plencner M, Knitlova J, Liskova J, Uhlik J, Zaloudikova M, Vondrasek D, Mikšik I, Ošťádal M. 2019. Novel contribution to clubfoot pathogenesis: The possible role of extracellular matrix proteins. *Journal of Orthopaedic Research* 37(3):769–778 DOI 10.1002/jor.24211.

- Esteban-Martínez L, Sierra-Filardi E, Boya P. 2017.** Mitophagy, metabolism, and cell fate. *Molecular & Cellular Oncology* 4(5):e1353854
DOI [10.1080/23723556.2017.1353854](https://doi.org/10.1080/23723556.2017.1353854).
- Ester AR, Weymouth KS, Burt A, Wise CA, Scott A, Gurnett CA, Dobbs MB, Blanton SH, Hecht JT. 2010.** Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. *American Journal of Medical Genetics Part A* 149A(12):2745–2752 DOI [10.1002/ajmg.a.33130](https://doi.org/10.1002/ajmg.a.33130).
- Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P. 2016.** The Reactome pathway knowledgebase. *Nucleic Acids Research* 44(D1):D481–D487 DOI [10.1093/nar/gkt1102](https://doi.org/10.1093/nar/gkt1102).
- Hecht JT, Ester A, Scott A, Wise CA, Iovannisci DM, Lammer EJ, Langlois PH, Blanton SH. 2007.** NAT2 variation and idiopathic talipes equinovarus (clubfoot). *American Journal of Medical Genetics Part A* 143A(19):2285–2291 DOI [10.1002/ajmg.a.31927](https://doi.org/10.1002/ajmg.a.31927).
- Herceg MB, Weiner DS, Agamanolis DP, Hawk D. 2006.** Histologic and histochemical analysis of muscle specimens in idiopathic talipes equinovarus. *Journal of Pediatric Orthopedics* 26(1):91–93 DOI [10.1097/01.bpo.0000188994.90931.e8](https://doi.org/10.1097/01.bpo.0000188994.90931.e8).
- Hester TW, Parkinson LC, Robson J, Misra S, Sangha H, Martin JE, Hester TW, Parkinson LC, Robson J, Misra S, Sangha H, Martin JE. 2009.** A hypothesis and model of reduced fetal movement as a common pathogenetic mechanism in clubfoot. *Medical Hypotheses* 73(6):986–988 DOI [10.1016/j.mehy.2009.04.056](https://doi.org/10.1016/j.mehy.2009.04.056).
- Howren AM, Jamieson DH, Alvarez CM. 2015.** Early ultrasonographic evaluation of idiopathic clubfeet treated with manipulations, casts, and Botox(®): a double-blind randomized control trial. *Journal of Children's Orthopaedics* 9(1):85–91 DOI [10.1007/s11832-015-0633-4](https://doi.org/10.1007/s11832-015-0633-4).
- Hu L, Pappoian GA. 2013.** Molecular transport modulates the adaptive response of branched actin networks to an external force. *The Journal of Physical Chemistry B* 117(42):13388–13396 DOI [10.1021/jp405179e](https://doi.org/10.1021/jp405179e).
- Li C, Nguyen Q, Cole WG, Alman BA. 2001.** Potential treatment for clubfeet based on growth factor blockade. *Journal of Pediatric Orthopedics* 21(3):372–377 DOI [10.1097/01241398-200105000-00021](https://doi.org/10.1097/01241398-200105000-00021).
- Liu ZP, Wu C, Miao H, Wu H. 2015.** RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* bav095 DOI [10.1093/database/bav095](https://doi.org/10.1093/database/bav095).
- Lovell ME, Morcuende JA. 2007.** Neuromuscular disease as the cause of late clubfoot relapses: report of 4 cases. *Iowa Orthopedic Journal* 27:82–84.
- Miedzybrodzka Z. 2003.** Congenital talipes equinovarus (clubfoot): a disorder of the foot but not the hand. *Journal of Anatomy* 202(1):37–42 DOI [10.1046/j.1469-7580.2003.00147.x](https://doi.org/10.1046/j.1469-7580.2003.00147.x).
- Miller NH, Carry PM, Mark BJ, Engelman GH, Georgopoulos G, Graham S, Dobbs MB. 2016.** Does strict adherence to the ponseti method improve isolated clubfoot

- treatment outcomes? A two-institution review. *Clinical Orthopaedics and Related Research* 474(1):237–243 DOI 10.1007/s11999-015-4559-4.
- Moazzen H, Lu X, Liu M, Feng Q. 2015.** Pregestational diabetes induces fetal coronary artery malformation via reactive oxygen species signaling. *Diabetes* 64(4):1431–1443 DOI 10.2337/db14-0190.
- Nogami H, Ogasawara N, Kasai T, Oki T, Murachi S. 1983.** Lipid storage myopathy associated with scoliosis and multiple joint contractures. *Acta Neuropathologica* 61(3-4):305–310 DOI 10.1007/bf00692002.
- Ohuchi H, Noji S. 1999.** Fibroblast-growth-factor-induced additional limbs in the study of initiation of limb formation, limb identity, myogenesis, and innervation. *Cell and Tissue Research* 296(1):45–56 DOI 10.1007/s004410051265.
- Ošťádal M, Eckhardt A, Herget J, Mikšík I, Dungl P, Chomiak J, Frydrychová M, Burian M. 2015.** Proteomic analysis of the extracellular matrix in idiopathic pes equinovarus. *Molecular and Cellular Biochemistry* 401(1-2):133–139 DOI 10.1007/s11010-014-2300-3.
- Palma M, Cook T, Segura J, Pecho A, Morcuende JA. 2013.** Descriptive epidemiology of clubfoot in Peru: a clinic-based study. *Iowa Orthopedic Journal* 33(2013):167–171.
- Parker SE, Mai CT, Strickland MJ, Olney RS, Rickard R, Marengo L, Wang Y, Hashmi SS. 2009.** Multistate study of the epidemiology of clubfoot. *Birth Defects Research Part A: Clinical and Molecular Teratology* 85(11):897–904 DOI 10.1002/bdra.20625.
- Pavone V, Chisari E, Vescio A, Lucenti L, Sessa G, Testa G. 2018.** The etiology of idiopathic congenital talipes equinovarus: a systematic review. *Journal of Orthopaedic Surgery and Research* 13(1):206 DOI 10.1186/s13018-018-0913-z.
- Poon R, Li C, Alman BA. 2009.** Beta-catenin mediates soft tissue contracture in clubfoot. *Clinical Orthopaedics and Related Research* 467(5):1180–1185 DOI 10.1007/s11999-008-0692-7.
- Robertson Jr WW, Corbett D. 1997.** Congenital clubfoot. Month of conception. *Clinical Orthopaedics and Related Research* 338:14–18 DOI 10.1097/00003086-199705000-00004.
- Shrimpton AE, Levinsohn EM, Yozawitz JM, Packard Jr DS, Cady RB, Middleton FA, Persico AM, Hootnick DR. 2004.** A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-Tooth disease. *American Journal of Human Genetics* 75(1):92–96 DOI 10.1086/422015.
- Shyy W, Dietz F, Dobbs MB, Sheffield VC, Morcuende JA. 2009.** Evaluation of CAND2 and WNT7a as candidate genes for congenital idiopathic clubfoot. *Clinical Orthopaedics and Related Research* 467(5):1201–1205 DOI 10.1007/s11999-008-0701-x.
- Smith PA, Kuo KN, Graf AN, Krzak J, Flanagan A, Hassani S, Caudill AK, Dietz FR, Morcuende J, Harris GF. 2014.** Long-term results of comprehensive clubfoot release versus the Ponseti method: which is better? *Clinical Orthopaedics and Related Research* 472(4):1281–1290 DOI 10.1007/s11999-013-3386-8.
- Sodre H, Bruschini S, Mestriner LA, Miranda Jr F, Levinsohn EM, Packard Jr DS, Crider Jr RJ, Schwartz R, Hootnick DR. 1990.** Arterial abnormalities in talipes

- equinovarus as assessed by angiography and the Doppler technique. *Journal of Pediatric Orthopedics* **10**(1):101–104 DOI [10.1097/01241398-199010010-00020](https://doi.org/10.1097/01241398-199010010-00020).
- Stangenberg S, Nguyen LT, Chen H, Al-Odat I, Killingsworth MC, Gosnell ME, Anwer AG, Goldys EM, Pollock CA, Saad S. 2015.** Oxidative stress, mitochondrial perturbations and fetal programming of renal disease induced by maternal smoking. *International Journal of Biochemistry and Cell Biology* **64**:81–90 DOI [10.1016/j.biocel.2015.03.017](https://doi.org/10.1016/j.biocel.2015.03.017).
- Thacker MM, Scher DM, Sala DA, Van Bosse HJ, Feldman DS, Lehman WB. 2005.** Use of the foot abduction orthosis following Ponseti casts: is it essential? *Journal of Pediatric Orthopedics* **25**(2):225–228 DOI [10.1097/01.bpo.0000150814.56790.f9](https://doi.org/10.1097/01.bpo.0000150814.56790.f9).
- Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, Anastasopoulos IL, Maniou S, Karathanou K, Kalfakakou D, Fevgas A, Dalamagas T, Hatzigeorgiou AG. 2015.** DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research* **43**(Database issue):D153–D159 DOI [10.1093/nar/gku1215](https://doi.org/10.1093/nar/gku1215).
- Wang H, Barisic I, Loane M, Addor MC, Bailey LM, Gatt M, Klungsoyr K, Mokoroa O, Nelen V, Neville AJ, O'Mahony M, Pierini A, Rissmann A, Verellen-Dumoulin C, De Walle HEK, Wiesel A, Wisniewska K, De Jong-van den Berg LTW, Dolk H, Khoshnood B, Garne E. 2019.** Congenital clubfoot in Europe: a population-based study. *American Journal of Medical Genetics Part A* **179**(4):595–601 DOI [10.1002/ajmg.a.61067](https://doi.org/10.1002/ajmg.a.61067).
- Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. 2017.** WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research* **45**(W1):W130–W137 DOI [10.1093/nar/gkx356](https://doi.org/10.1093/nar/gkx356).
- Wang LL, Fu WN, Li-Ling J, Li ZG, Li LY, Sun KL. 2008.** HOXD13 may play a role in idiopathic congenital clubfoot by regulating the expression of FHL1. *Cytogenetic and Genome Research* **121**(3-4):189–195 DOI [10.1159/000138884](https://doi.org/10.1159/000138884).
- Wang S, Sun H, Ma J, Zang C, Wang C, Wang J, Tang Q, Meyer CA, Zhang Y, Liu XS. 2013a.** Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nature Protocols* **8**(12):2502–2515 DOI [10.1038/nprot.2013.150](https://doi.org/10.1038/nprot.2013.150).
- Wang Z, Yan N, Liu L, Cao D, Gao M, Lin C, Jin C. 2013b.** SOX9 overexpression plays a potential role in idiopathic congenital talipes equinovarus. *Molecular Medicine Reports* **7**(3):821–825 DOI [10.3892/mmr.2012.1245](https://doi.org/10.3892/mmr.2012.1245).
- Weymouth KS, Blanton SH, Powell T, Patel CV, Savill SA, Hecht JT. 2016.** Functional assessment of clubfoot associated HOXA9, TPM1, and TPM2 variants suggests a potential gene regulation mechanism. *Clinical Orthopaedics and Related Research* **474**(7):1726–1735 DOI [10.1007/s11999-016-4788-1](https://doi.org/10.1007/s11999-016-4788-1).
- Zhang B, Dietrich UM, Geng JG, Bicknell R, Esko JD, Wang L. 2009.** Repulsive axon guidance molecule Slit3 is a novel angiogenic factor. *Blood* **114**(19):4300–4309 DOI [10.1182/blood-2008-12-193326](https://doi.org/10.1182/blood-2008-12-193326).

Zhang Z, Kong Z, Zhu M, Lu W, Ni L, Bai Y, Lou Y. 2016. Whole genome sequencing identifies ANXA3 and MTHFR mutations in a large family with an unknown equinus deformity associated genetic disorder. *Molecular Biology Reports* **43**:1147–1155 DOI [10.1007/s11033-016-4047-2](https://doi.org/10.1007/s11033-016-4047-2).

Zionts LE, Zhao G, Hitchcock K, Maewal J, Ebrahimzadeh E. 2010. Has the rate of extensive surgery to treat idiopathic clubfoot declined in the United States? *Journal of Bone and Joint Surgery. American Volume* **92(4)**:882–829 DOI [10.2106/JBJS.I.00819](https://doi.org/10.2106/JBJS.I.00819).