ORIGINAL ARTICLE

Genome-wide association study identifies new disease loci for isolated clubfoot

Tian-Xiao Zhang,[1] Gabe Haller,[2] Peng Lin,[1] David M Alvarado,[2] Jacqueline T Hecht,[3] Susan H Blanton,[4] B Stephens Richards,[5] John P Rice,[6] Matthew B Dobbs,[2,7] Christina A Gurnett[2,8,9]

▶ Additional material is published online. To view please visit the journal (http://dx.doi.org/10.1136/jmedgenet-2014-102303).

For numbered affiliations see end of article.

Correspondence to
Dr Christina A Gurnett, Department of Neurology, Pediatrics, and Orthopaedic Surgery, Washington University School of Medicine, 660 S Euclid Ave, St Louis, MO 63110, USA; gurnettc@neuro.wustl.edu

Received 21 January 2014
Revised 25 February 2014
Accepted 26 February 2014
Published Online First 25 March 2014

ABSTRACT

Background Clubfoot is a common congenital birth defect with complex inheritance patterns. Currently, the genetic and morphological basis of clubfoot is poorly understood. To identify genetic risk factors associated with clubfoot, we performed a genome-wide association study of common genetic variants.

Methods The DNA of 396 isolated clubfoot patients and 1000 controls of European descent was genotyped for >600 000 single nucleotide polymorphisms (SNP) using the Affymetrix 6.0 array. Replication was performed with an independent cohort of 370 isolated clubfoot cases and 363 controls of European descent.

Results Strongest evidence for an association of clubfoot was found with an intergenic SNP on chromosome 12q24.31 between *NCOR2* and *ZNF664* (rs7969148, OR=0.58, $p=1.25\times10^{-5}$) that was significant on replication (combined OR=0.63, $p=1.90\times10^{-7}$). Additional suggestive SNPs were identified near *FOXN3*, *SORCS1* and *MMP7/TMEM123* that also confirmed on replication.

Conclusions Our study suggests a potential role for common genetic variation in several genes that have not previously been implicated in clubfoot pathogenesis.

INTRODUCTION

Clubfoot is a congenital malformation of the lower limb that occurs in 1 in 1000 infants. An important role for genetic factors in clubfoot aetiology is supported by high concordance rates in identical compared to fraternal twins (33% vs 3%),[1] and an increased risk to first-degree relatives.[2] Clubfoot prevalence varies across ethnic populations, with the lowest prevalence in Chinese (0.39 cases per 1000 live births) and the highest in Hawaiians and Maoris (7 per 1000).[3 4] The males to females ratio of idiopathic clubfoot is 2:1 and is consistent across ethnic groups.[3]

Genome-wide association studies using SNPs have not previously been reported for clubfoot. A role for common genetic variants in clubfoot susceptibility has been evaluated for candidate genes, including *HOX* homeodomain genes and caspase genes that both show modest association.[5 6] Clubfoot has also been associated with maternal methylenetetrahydrofolate reductase gene (*MTHFR*) polymorphisms.[7] However, clubfoot remains as one of a handful of common human birth defects for which an unbiased genome-wide association study has not yet been published.

An important role for hindlimb-specific transcription factors in the aetiology of clubfoot has previously been shown through genome-wide studies of copy number variants associated with clubfoot susceptibility. Data from our laboratory support a role for the *PITX1-TBX4* developmental pathway in clubfoot aetiology, including the presence of *PITX1* mutations and deletions in clubfoot families.[8–10] Recurrent chromosome 17q23 copy number variants containing *TBX4*, a downstream transcriptional target of *PITX1*,[11 12] were found in ~5% of familial isolated clubfoot patients, representing the most common cause of isolated familial clubfoot.[13 14] However, the absence of these genetic abnormalities in most clubfoot patients,[9 13 14] the presence of multiple chromosomal loci previously associated with clubfoot,[15] and the lack of association of clubfoot with common SNPs within *TBX4*[14] suggests the possibility that clubfoot is due to other genetic factors. Because clubfoot inheritance is most often considered complex with more than 75% of all cases reporting no family history,[2 16] we decided to test the common disease-common variant hypothesis by performing a genome-wide association study of isolated clubfoot.

METHODS

Ascertainment of cases for discovery

Isolated clubfoot patients of European descent were recruited from St Louis Children's Hospital and Shriners Hospital in St Louis for the discovery phase of the study. Individuals recruited from St Louis are representative of the US population of European descent as a whole, with no particular overrepresentation of any immigrant group. The institutional review boards of all centres approved the study, and informed consent was obtained. All patients were evaluated by an experienced orthopaedic surgeon (MBD) and diagnosed with clubfoot that requires the presence of rigid, not passively correctable hindfoot equinus, hindfoot varus, forefoot supination, and midfoot cavus deformities. Exclusion criteria include the presence of additional birth defects, known genetic syndromes, developmental delay or mental retardation, or inclusion of any other family member in the study.

Microarray genotyping and quality control for cases for discovery

Affymetrix 6.0 microarray genotyping of 420 clubfoot cases was performed by the Washington University DNA Microarray Core using protocols

To cite: Zhang T-X, Haller G, Lin P, *et al*. *J Med Genet* 2014;**51**:334–339.

provided by Affymetrix. Genotype data was acquired by genotype calling of batches of 48 samples using the default algorithm of the Affymetrix Genotyping Console (Birdseed v2). Twenty-four samples were excluded for contrast quality control scores <1.7. Quality control consisted of assessing genotyping rates, genomic sex, unexpected duplicates and cryptic relatedness. After quality control, 396 persons were kept for the association analysis.

Ascertainment of controls and quality control for discovery

A control genotype dataset consisting of 1031 controls recruited from Queensland Institute of Medical Research (QIMR) was independently genotyped on the same Affymetrix 6.0 array.[17] The controls were individuals of European descent who were unaffected with psychiatric disorders.[17] Quality control consisted of assessing genomic sex, unexpected duplicates and cryptic relatedness. Additionally, 12 persons were removed due to low genotyping rates. This resulted in 1000 controls who were used in the analysis.

SNP genotyping quality

Deviations from Hardy–Weinberg equilibrium (HWE) were assessed with 1° of freedom in controls. SNPs with significant deviations from HWE ($p<1\times10^{-3}$) were excluded from our association analysis. We also excluded SNPs with call rates <95% and SNPs with minor allele frequency <0.01. We also assessed the difference of missing call rates between cases and controls. If the difference of the missing call rates between cases and controls is greater than 0.02 as described in Shyn *et al*, these SNPs were dropped before association analysis.[18] The final analysis was carried out with 620 820 SNPs.

Assessment of population stratification

To assess substructure, the principal component analysis method of EIGENSTRAT was used for the data analysis reported here.[19] Seven cases and four controls were removed because of questionable European descent (see online supplementary figures S1 and S2).

Pathway analysis

Pathway analysis was performed using the method of Wang *et al*.[20] To reduce the multiple comparison problem, we implemented a candidate-based strategy[21] through which we identified 63 clubfoot-related genes through the GeneCards database (http://www.genecards.org/),[22] and extracted pathways that have at least one of these 63 genes from Gene Ontology database (http://www.geneontology.org/).[23]

Imputation analysis

Imputation using BEAGLE V.3.3.2[24] analysis was performed to increase the coverage of markers in genome wide association study (GWAS). We used 1092 individuals from phase one of the 1000 Genomes Project as our reference panel.[25]

Statistical analysis

Association analysis was carried out using PLINK, with case versus control status as the binary variable. It consisted of 396 cases and 1000 QIMR controls. Logistic regression analysis was performed with sex as a covariate. The conventional criteria for genome-wide significance $p<5\times10^{-8}$ and a significance threshold for multiple testing after adjusting for linkage disequilibrium was applied $p<1.6\times10^{-6}$ (Bonferroni method: p<(p=0.05/total number of markers)).

Replication cohort genotyping

Cases for the replication study were recruited from St Louis Children's Hospital/Shriners Hospital, Carolinas Medical Center, and Sinai Hospital (Baltimore), and met the same inclusion criteria as for the discovery phase. Controls for the replication study consisted of clubfoot-unaffected individuals recruited from St Louis Children's Hospital and Shriners Hospital in St Louis. Both were restricted to individuals of European descent. Markers were selected for replication if they were: (1) among the top SNPs from analyses ($p<2\times10^{-5}$), (2) located near other high-scoring SNPs, or (3) in a gene that is biologically plausible (expressed in developing limb, or related to other genes previously associated with limb contractures). Thirty-one SNPs were selected for replication. SNPs chosen for replication were genotyped using the Sequenom Mass ARRAY iPlex technology in an independent cohort of 374 cases and 310 controls of European descent. All SNPs were in HWE.

RESULTS

GWAS results

Quantile-quantile plot showed that most of the observed associations lie along the expected distribution conforming to the null hypothesis of no association for the majority of SNPs (see online supplementary figure S3). The lambda value (inflation factor) was 1.07. Three SNPs analysed met conventional criteria for genome-wide significance $p<5\times10^{-8}$. Four SNPs in total reached the significance threshold determined after correction for multiple testing (figure 1). SNPs with $p<1\times10^{-5}$ in the discovery phase are shown in table 1, and the results of the top 200 SNPs are listed in online supplementary table S1. The most significant SNP in the discovery phase (rs6705159; $p=1.35\times10^{-9}$) is located 700 kb proximal to the *HOXD* cluster, near the 5′ gene *HOXD10* that was previously implicated in a related musculoskeletal disorder, congenital vertical talus.[26 27]

Replication

Replication studies were performed on selected SNPs in an independent cohort of 370 cases of European descent that were also recruited from our research centres. None of the SNPs that met the genome-wide significance threshold on the GWAS discovery phase were significant on subsequent replication testing. However, rs7969148, located on chromosome 12q24.31, was associated with clubfoot in discovery ($p=1.25\times10^{-5}$) and replication studies (p=0.0022), resulting in a combined $p=1.91\times10^{-7}$; OR=0.63 (770 cases, 1310 controls). Several flanking SNPs near rs7969148 also showed an association with clubfoot although to a lesser degree (figure 2). The SNP rs7969148 is intergenic between transcriptional regulators *ZNF664* and *NCOR2*.

We attempted to replicate the association rs7969148 in a third clubfoot dataset by comparing our data with results of an unpublished GWAS composed of 603 clubfoot families including a total of 1716 study persons of Hispanic (344 affecteds/495 unaffecteds) and European (382 affecteds/495 unaffecteds) ancestry (Hecht and Blanton, unpublished results). These families were genotyped on the Illumina HumanOmniExpress_12v1_C array and were described previously.[28] Evidence was found for an association of rs7969148 (near *ZNF664* and *NCOR2*) with clubfoot in the Hispanic subgroup with a p=0.011871, but not in the patients with European ancestry.

Additional SNPs for replication were selected based on consideration of p value from the discovery GWAS and hindlimb expression.[10] Table 2 lists all SNPs that were successfully

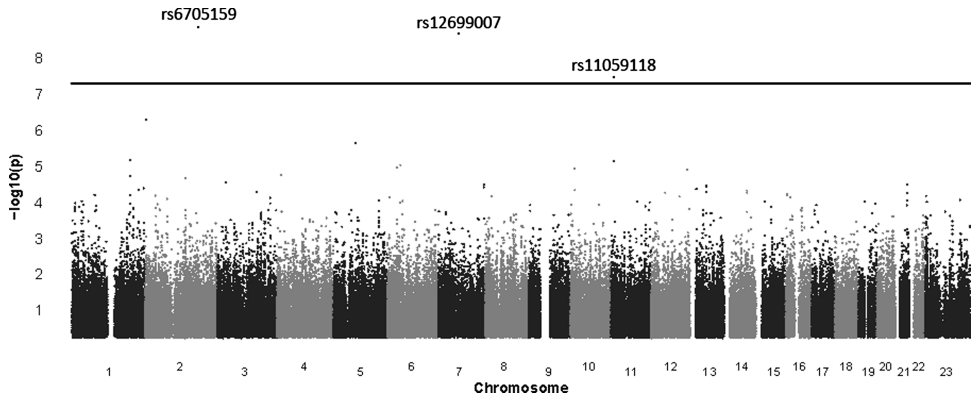

Figure 1 Manhattan Plot of single nucleotide polymorphism (SNP) −log10 p values (Y-axis) obtained by genome-wide association analysis of 396 clubfoot cases versus 1000 controls of European descent. Chromosomes are along the X-axis. The upper line indicates conventional threshold for genome-wide significance ($p<5\times10^{-8}$), and the lower line indicates threshold for significance with multiple testing after adjusting for linkage disequilibrium (LD) was applied $p<1.6\times10^{-6}$ (Bonferroni method: p<(p=0.05/total number of markers)).

replicated. These include rs12885505 that is located within the first intron of the forkhead/winged helix transcription factor gene *FOXN3*. Association analysis of rs12885505 resulted in a combined OR=0.64, $p=2.953\times10^{-5}$. Also significant in the replication phase was rs4918273 located within the *SORCS1* gene, a member of the vacuolar protein sorting (VPS10) domain-containing receptor gene family. Combined analysis yielded an OR=0.77, p=0.0001057. Finally, rs11225266, located on chromosome 11 between *TMEM123*, a predicted transmembrane protein of unknown function, and matrix metalloproteinase *MMP7*, was associated with clubfoot in discovery and replication studies. Online supplementary table S2 lists all SNPs that were tested and not replicated in the study.

Clubfoot candidate regions

We also evaluated SNPs near candidate regions of interest including *HOX* genes, *PITX1*, *TBX4*, as well as skeletal muscle sarcomeric genes (*TPM1*, *MYH3* and *TPM2*) because of their previous association with clubfoot in candidate gene studies or copy number analyses.[10 13 28 29] None except the chromosome 17q23 region containing *TBX4* gave evidence for association with isolated clubfoot. A cluster of SNPs showed weak evidence for association, including rs11079429 (p=0.003) that is located ~5 kb upstream of *TBX2*, and two other SNPs that are located within or near *TBX4* (rs7213643, p=0.009; rs744651, p=0.009) (see online supplementary figure S4). However, these SNPs were not significant in the replication phase.

Pathway analysis

Of the 13 301 pathways that were extracted from Gene Ontology, 796 were associated with the 63 clubfoot candidate genes identified by GeneCards (see online supplementary table S3). To limit the potential bias introduced by differences of the length of pathway, we only analysed 312 pathways containing a range of 20–200 genes. Analysis of clubfoot-associated SNPs revealed 13 pathways that achieved nominal significance (p<0.05). However, after multiple comparison correction (FDR and FWER), none of these pathways were significant (see online supplementary table S4).

Imputation analysis

By applying $r^2>0.9$ as criteria for analysis, we successfully imputed 5 190 243 markers. The association analysis failed to identify any significant markers that pass the genome-wide

Table 1 Top results of genome-wide association study of clubfoot with replication results

					Discovery (N=396 cases and N=1000 controls)				Replication (N=370 cases and N=363 controls)			
					Allele frequency				Allele frequency			
SNP	Chr	Position	Locus	Test allele	Cases	Controls	OR	p Value	Cases	Controls	OR	p Value
rs6705159	2	176 260 190	*ATP5G3/HOXD13*	G	0.1308	0.0460	2.477	1.35E–09	0.048	0.035	1.41	0.269
rs12699007	7	70 389 161	*WBSCR17/AUTS2*	G	0.1649	0.0587	2.184	2.00E–09	0.100	0.128	0.76	0.219
rs274503	11	11 059 118	*GALNTL4/ZBED5*	C	0.0893	0.0255	2.736	3.16E–08	0.022	0.023	1.10	0.803
rs12714318	2	2 895 962	*MYT1L*	G	0.1628	0.0743	1.862	5.17E–07	0.079	0.078	1.07	0.922
rs6861281	5	73 389 024	*ENC1*	C	0.1390	0.0651	1.875	2.20E–06	0.071	0.066	1.05	0.721
rs2648772	1	197 381 370	*PTPRC*	T	0.4255	0.4700	1.529	6.52E–06	0.508	0.491	1.07	0.628
rs6484839	11	11 335 231	*GALNTL4*	T	0.4690	0.3790	1.535	7.08E–06	0.384	0.397	1.07	0.629
rs881934	6	44 158 247	*LOC652990*	G	0.4630	0.4465	1.523	9.20E–06	0.4796	0.4858	0.9747	0.8338
rs3892710	6	32 790 840	*HLA-DQB1*	A	0.2051	0.1436	1.74	1.09E–05	0.184	0.176	1.21	0.340
rs10906679	10	14 524 480	*FAM107B*	T	0.2828	0.3555	0.6439	1.15E–05	0.3815	0.3608	1.093	0.4741
rs7969148	12	123 180 491	*NCOR2/ZNF664*	G	0.1474	0.2195	0.5803	1.25E–05	0.164	0.2302	0.6722	0.008

SNP, single nucleotide polymorphism.

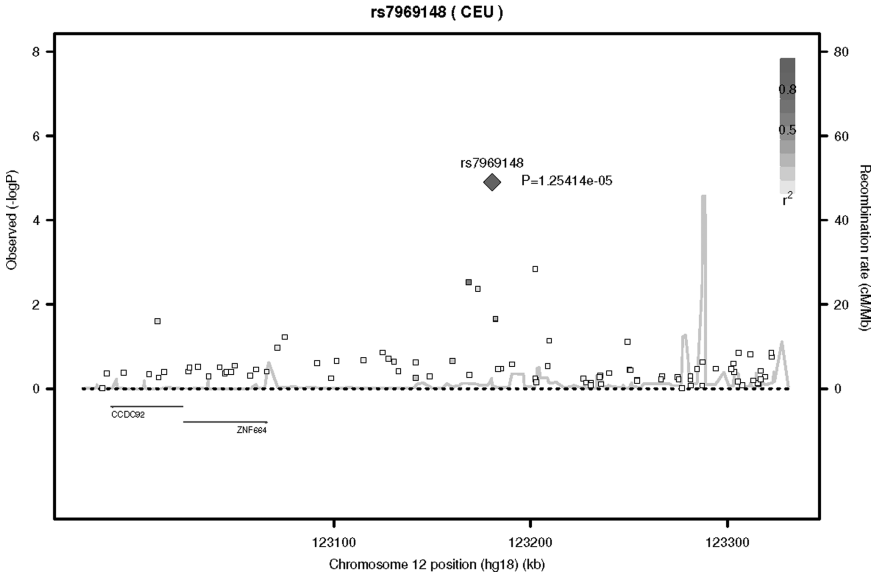

Figure 2 Clubfoot genetic association results for single nucleotide polymorphisms (SNP) on chromosome 12q24.31 including rs7969148. Chromosomal location along the X-axis is given in kb as denoted in assembly hg18.

threshold level of significance. We summarised the top 50 significant imputed SNPs in online supplementary table S5.

DISCUSSION

Although clubfoot may occasionally be familal, the majority of clubfoot cases are sporadic. Thus, clubfoot may be considered a complex trait. To test the common disease-common variant model of disease susceptibility, we completed a genome-wide association study of isolated clubfoot. The strongest evidence for an association of a SNP with clubfoot in this study occurred for a variant (rs7969148) on chromosome 12q24.31 that is intergenic to two transcriptional regulators, *ZNF664* and *NCOR2*. Common variants in *NCOR2*, also known as silencing mediator of retinoic acid and thyroid hormone receptor (SMRT), have been associated with knee osteoarthritis in one study.[30] *NCOR2* is widely expressed in the embryo[31] and knockout of *NCOR2* in mice is embryonic lethal, resulting in forebrain defects, cardiac malformations and cleft palate.[31] *ZNF664* is a krueppel C2H2-type zinc-finger protein of unknown function. Neither gene has been associated with limb development. Although the association of rs7969148 with clubfoot replicated in our replication cohort, further evaluation of this SNP in an independent, unpublished Texas clubfoot dataset showed only a weak association within a Hispanic subgroup. Failure to replicate this association may reflect either genetic heterogeneity, or lack of power due to relatively small sample size.

Several single nucleotide variants in other genes replicated on further study, though they are less strongly associated with clubfoot than rs769148. These genes include *FOXN3*, a forkhead/winged helix transcription factor that is highly expressed in the developing limb.[32] *FOXN3* is essential for craniofacial development in mice, and its loss is associated with embryonic lethality, growth retardation, and scoliosis-like deformity.[32] Limb abnormalities, including congenital vertical talus (DECIPHER Pt 263695),[33] over-riding toes (DECIPHER Pt 251094)[33] and syndactyly,[34] have been described in patients with large chromosomal deletions including *FOXN3*, although a specific role in limb development has not been shown. Clubfoot is also associated with intronic variants of *SORCS1*, a member of the vacuolar protein sorting 10 (VPS10) domain-containing receptor protein that has been repeatedly linked to type 2 diabetes.[35–37] The mechanism by which *SORCS1* may contribute to clubfoot susceptibility is unknown, although maternal diabetes is a known risk factor for human clubfoot.[38] Finally, intergenic SNPs between *TMEM123* and *MMP7* are also associated with clubfoot. *TMEM123*, also known as porimin, is a transmembrane protein that plays a role in oncotic cell death and cell adhesion,[39] and *MMP7* is a ubiquitously expressed matrix metalloproteinase; neither gene has a known role in limb development.

To optimise the number of cases that could be cost-effectively genotyped for this study, we used controls that were collected

Table 2 Clubfoot genetic association results for most significant SNPs in discovery and replication studies

					Discovery (N=396 cases and N=1000 controls)				Replication (N=370 cases and N=363 controls)				Combined	
					Allele frequency				Allele frequency					
SNP	Chr	Position	Locus	Test allele	Cases	Controls	OR	p Value	Cases	Controls	OR	p Value	OR	p Value
rs7969148	12	123 180 491	*NCOR2/ ZNF664*	G	0.1474	0.2195	0.5803	1.25E–05	0.1620	0.2270	0.6607	0.0022	0.6345	1.901E–07
rs12885505	14	88 971 715	*FOXN3*	A	0.0947	0.1405	0.5781	0.000261	0.0990	0.1360	0.6991	0.0302	0.644	2.953E–05
rs4918273	10	108 715 485	*SORCS1*	C	0.3472	0.417	0.7148	0.000442	0.3676	0.4298	0.7813	0.0216	0.7693	0.0001057
rs11225266	11	101 851 796	*TMEM123/ MMP7*	C	0.05177	0.094	0.5305	0.000829	0.0622	0.1088	0.5445	0.0019	0.565	1.152E–05

SNP, single nucleotide polymorphism.

and genotyped independently on the same platform for a different study.[17] Common controls have successfully been used in previous studies as a cost-saving measure, most notably in the Wellcome Trust Case Control Consortium that studied seven common diseases with shared controls.[40] To reduce the possibility that differences in genotyping between the two centres would result in false positive results we conservatively applied data quality control measures in this study.[41] Despite our rigorous application of principal components analysis to cases and controls, we also recognise that residual differences in population substructure may exist between our cases and controls resulting in spurious claims of associations, particularly as our cases are drawn from a different continent than our controls. However, when confronted with similar issues, the Wellcome Trust Case Control study found only a modest effect of residual substructure on type I error.[40] Notably, however, the study we report here is of much smaller scale, and therefore, may be more susceptible to these artefacts.

Although the most significant SNP in the discovery phase, rs6705159, did not replicate in further studies, we find it intriguing that this SNP is located near the *HOXD* cluster, near the 5′ gene *HOXD10* that was previously implicated in the related condition, congenital vertical talus.[26 27] *HOXD10* coding mutations were previously excluded in clubfoot,[42] and association of clubfoot with common SNPs in *HOXD10* was not found.[28] However, regulatory variants were not assessed. One possible explanation for lack of replication of this SNP, and perhaps others in this study, is that the patients in the discovery phase included more familial cases than those in the replication cohort (35% vs 22%, data not shown). Many of the top hits from this GWAS are in SNPs with low minor allele frequencies that are more likely to be affected by small biases in the patient populations. Exome sequencing of familial cases is now being performed to further evaluate a role for rare variants, although significantly larger numbers of cases will clearly be needed for meaningful analysis.

To extract more information from our resultant GWAS dataset, we conducted pathway analysis and imputation analysis. Although we failed to identify pathways that were significantly associated with clubfoot after multiple corrections, 13 pathways were nominally significant, including a group of genes involved in osteoblast differentiation. Imputation analysis did not reveal any other significant signals.

Overall, this study suggests a potential role for common genetic variation in or near several genes that have not previously been implicated in clubfoot pathogenesis including two transcriptional regulators (*NCOR2* and *ZNF664*) on chromosome 12q24.31. Despite evidence linking rare copy number variants involving hindlimb-specific transcription factors *PITX1*, *TBX4* and *HOXC* genes to clubfoot pathogenesis,[8] only weak, non-replicating association was seen with common variants around *TBX4*, similar to a previous study.[14] These results provide further evidence for the genetic heterogeneity of clubfoot. As sequencing studies now seek to identify novel rare variants associated with clubfoot in families and in large patient cohorts, it will be interesting to determine whether rare variants within any of the genes that we identified here will play a role in clubfoot susceptibility.

Author affiliations
[1]Division of Biology and Biomedical Sciences, Washington University School of Medicine, St Louis, Missouri, USA
[2]Department of Orthopaedic Surgery, Washington University School of Medicine, St Louis, Missouri, USA
[3]Department of Pediatrics, University of Texas Medical School at Houston, Houston, Texas, USA
[4]Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, Florida, USA
[5]Texas Scottish Rite Hospital for Children, Dallas, Texas, USA
[6]Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri, USA
[7]St Louis Shriners Hospital for Children, St Louis, Missouri, USA
[8]Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri, USA
[9]Department of Neurology, Washington University School of Medicine, St Louis, Missouri, USA

Acknowledgements We thank Carol Wise and Jonathan Rios for helpful comments on the manuscript. We thank Steven Frick and John Herzenberg for contributing clubfoot patient samples for the replication phase. We acknowledge the generous sharing of control samples from QIMR (Queensland Institute of Medical Research) study of Wray *et al*.[17] This study makes use of data generated by the DECIPHER Consortium. A full list of centres that contributed to the generation of the data is available from (http://decipher.sanger.ac.uk) and via email from (decipher@sanger.ac.uk). Funding for the project was provided by the Wellcome Trust.

Contributors T-XZ, GH, PL, JR and DMA performed the genome-wide association analysis and replication studies. Replication study was performed by Hecht, Blanton and Richards. The study was designed by MBD and CAG. The manuscript was written with contributions from all the authors.

Funding This work was supported by grants from Shriners Hospital for Children (MBD and JTH), The Children's Discovery Institute of Washington University and St Louis Children's Hospital, The Orthopaedic Research and Education Foundation, Paediatric Orthopaedic Society of North America, St Louis Children's Hospital Foundation, and a W.M. Keck Postdoctoral Fellowship in Molecular Medicine (DMA), and NIH R01HD043342 (JTH).

Competing interests None.

Ethics approval Washington University Human Subjects Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The Affymetrix 6.0 data from this study will be released to interested scientific researchers upon request.

REFERENCES

1 Wynne-Davies R. Family studies and the cause of congenital club foot: talipes equinovarus, talipes calcaneo-valgus and metatarsus varus. *J Bone Joint Surg Br* 1964;46:445–63.
2 Lochmiller C, Johnston D, Scott A, Risman M, Hecht JT. Genetic epidemiology study of idiopathic talipes equinovarus. *Am J Med Genet* 1998;79:90–6.
3 Chung CS, Nemechek RW, Larsen IJ, Ching GH. Genetic and epidemiological studies of clubfoot in Hawaii. General and medical considerations. *Hum Hered* 1969;19:321–42.
4 Beals RK. Club foot in the Maori: a genetic study of 50 kindreds. *NZ Med J* 1978;88:144–6.
5 Heck AL, Bray MS, Scott A, Blanton SH, Hecht JT. Variation in CASP10 gene is associated with idiopathic talipes equinovarus. *J Pediatr Orthop* 2005;25:598–602.
6 Sun M, Ma F, Zeng X, Liu Q, Zhao XL, Wu FX, Wu GP, Zhang ZF, Gu B, Zhao YF, Tian SH, Lin B, Kong XY, Zhang XL, Yang W, Lo W, Zhang X. Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long-range, limb-specific SHH enhancer. *J Med Genet* 2008;45:589–95.
7 Sharp L, Miedzybrodzka Z, Cardy AH, Inglis J, Madrigal L, Barker S, Chesney D, Clark C, Maffulli N. The C677T polymorphism in the methylenetetrahydrofolate reductase gene (MTHFR), maternal use of folic acid supplements, and risk of isolated clubfoot: A case-parent-triad analysis. *Am J Epidemiol* 2006;164:852–61.
8 Alvarado DM, Buchan JG, Frick SL, Herzenberg JE, Dobbs MB, Gurnett CA. Copy number analysis of 413 isolated talipes equinovarus patients suggests role for transcriptional regulators of early limb development. *Eur J Hum Genet* 2013;21:373–80.
9 Gurnett CA, Alaee F, Kruse LM, Desruisseau DM, Hecht JT, Wise CA, Bowcock AM, Dobbs MB. Asymmetric lower-limb malformations in individuals with homeobox PITX1 gene mutation. *Am J Hum Genet* 2008;83:616–22.
10 Alvarado DM, McCall K, Aferol H, Silva MJ, Garbow JR, Spees WM, Patel T, Siegel M, Dobbs MB, Gurnett CA. Pitx1 Haploinsufficiency Causes Clubfoot in Humans and a Clubfoot-like Phenotype in Mice. *Hum Mol Genet* 2011;20:3943–52.
11 Logan M, Tabin CJ. Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* 1999;283:1736–9.

12 Menke DB, Guenther C, Kingsley DM. Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. *Development* 2008;135:2543–53.
13 Alvarado DM, Aferol H, McCall K, Huang JB, Techy M, Buchan J, Cady J, Gonzales PR, Dobbs MB, Gurnett CA. Familial isolated clubfoot is associated with recurrent chromosome 17q23.1q23.2 microduplications containing TBX4. *Am J Hum Genet* 2010;87:154–60.
14 Lu W, Bacino CA, Richards BS, Alvarez C, Vandermeer JE, Vella M, Ahituv N, Sikka N, Dietz FR, Blanton SH, Hecht JT. Studies of TBX4 and chromosome 17q23.1q23.2: An uncommon cause of nonsyndromic clubfoot. *Am J Med Genet A* 2012;158A:1620–7.
15 Dietz FR, Cole WG, Tosi LL, Carroll NC, Werner RD, Comstock D, Murray JC. A search for the gene(s) predisposing to idiopathic clubfoot. *Clin Genet* 2005;67:361–2.
16 Gurnett CA, Boehm S, Connolly A, Reimschisel T, Dobbs MB. Impact of congenital talipes equinovarus etiology on treatment outcomes. *Dev Med Child Neurol* 2008;50:498–502.
17 Wray NR, Pergadia ML, Blackwood DH, Penninx BQ, Gordon SD, Nyholt DR, Ripke S, Macintyre DJ, McGhee KA, Maclean AW, Smit JH, Hottenga JJ, Willemsen G, Middelldorp CM, de Geus EJ, Lewis CM, McGuffin P, Hickie IB, van den Oord EJ, Liu JZ, Macgregor S, McEvoy BP, Byrne EM, Medland SE, Statham DJ, Henders AK, Heath AC, Montgomery GW, Martin NG, Boomsma DI, Madden PA, Sullivan PF. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* 2010;17:36–48.
18 Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, Garriock HA, Yokoyama JS, McGrath PJ, Peters EJ, Scheftner WA, Coryell W, Lawson WB, Jancic D, Gejman PV, Sanders AR, Holmans P, Slager SL, Levinson DF, Hamilton SP. Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry* 2011;16:202–15.
19 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
20 Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278–83.
21 Zhang TX, Beaty TH, Ruczinski I. Candidate pathway based analysis for cleft lip with or without cleft palate. *Stat Appl Genet Mol Biol* 2012;11–2.
22 Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet* 1997;13:163.
23 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
24 Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009;84:210–23.
25 Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
26 Dobbs MB, Gurnett CA, Pierce B, Exner GU, Robarge J, Morcuende JA, Cole WG, Templeton PA, Foster B, Bowcock AM. HOXD10 M319K mutation in a family with isolated congenital vertical talus. *J Orthop Res* 2006;24:448–53.
27 Shrimpton AE, Levinsohn EM, Yozawitz JM, Packard DS Jr, Cady RB, Middleton FA, Persico AM, Hootnick DR. A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-Tooth disease. *Am J Hum Genet* 2004;75:92–6.
28 Ester AR, Weymouth KS, Burt A, Wise CA, Scott A, Gurnett CA, Dobbs MB, Blanton SH, Hecht JT. Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. *Am J Med Genet A* 2009;149A:2745–52.
29 Weymouth KS, Blanton SH, Bamshad MJ, Beck AE, Alvarez C, Richards S, Gurnett CA, Dobbs MB, Barnes D, Mitchell LE, Hecht JT. Variants in genes that encode muscle contractile proteins influence risk for isolated clubfoot. *Am J Med Genet A* 2011;155A:2170–9.
30 Valdes AM, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, Schafer AJ, Spector TD. Association study of candidate genes for the prevalence and progression of knee osteoarthritis. *Arthritis Rheum* 2004;50:2497–507.
31 Jepsen K, Solum D, Zhou T, McEvilly RJ, Kim HJ, Glass CK, Hermanson O, Rosenfeld MG. SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature* 2007;450:415–19.
32 Samaan G, Yugo D, Rajagopalan S, Wall J, Donnell R, Goldowitz D, Gopalakrishnan R, Venkatachalam S. Foxn3 is essential for craniofacial development in mice and a putative candidate involved in human congenital craniofacial defects. *Biochem Biophys Res Commun* 2010;400:60–5.
33 https://decipher.sanger.ac.uk/patient/263695
34 Schlade-Bartusiak K, Macintyre G, Zunich J, Cox DW. A child with deletion (14)(q24.3q32.13) and auditory neuropathy. *Am J Med Genet A* 2008;146A:117–23.
35 Clee SM, Yandell BS, Schueler KM, Rabaglia ME, Richards OC, Raines SM, Kabara EA, Klass DM, Mui ET, Stapleton DS, Gray-Keller MP, Young MB, Stoehr JP, Lan H, Boronenkov I, Raess PW, Flowers MT, Attie AD. Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat Genet* 2006;38:688–93.
36 Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, Frayling TM, Bell JI, Lathrop GM, Bennett A, Dhillon R, Fletcher C, Groves CJ, Jones E, Prestwich P, Simecek N, Rao PV, Wishart M, Bottazzo GF, Foxon R, Howell S, Smedley D, Cardon LR, Menzel S, McCarthy MI. A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 2001;69:553–69.
37 Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 1999;64:1127–40.
38 Parker SE, Mai CT, Strickland MJ, Olney RS, Rickard R, Marengo L, Wang Y, Hasmi SS, Meyer RE. Multistate study of the epidemiology of clubfoot. *Birth Defects Res A Clin Mol Teratol* 2009;85:897–904.
39 Ma F, Zhang C, Prasad KV, Freeman GJ, Schlossman SF. Molecular cloning of Porimin, a novel cell surface receptor mediating oncotic cell death. *Proc Natl Acad Sci USA* 2001;98:9778–83.
40 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–78.
41 Plagnol V, Cooper JD, Todd JA, Clayton DG. A method to address differential bias in genotyping in large-scale association studies. *PLoS Genet* 2007;3:e74.
42 Gurnett CA, Keppel C, Bick J, Bowcock AM, Dobbs MB. Absence of HOXD10 Mutations in Idiopathic Clubfoot and Sporadic Vertical Talus. *Clin Orthop Relat Res* 2007;462:27–31.

Genome-wide association study identifies new disease loci for isolated clubfoot

Tian-Xiao Zhang, Gabe Haller, Peng Lin, David M Alvarado, Jacqueline T Hecht, Susan H Blanton, B Stephens Richards, John P Rice, Matthew B Dobbs and Christina A Gurnett

J Med Genet 2014 51: 334-339 originally published online March 25, 2014
doi: 10.1136/jmedgenet-2014-102303