Rare and de novo duplications containing SHOX in clubfoot

Brooke Sadler, Gabe Haller, Lilian Antunes, Momchil Nikolov, Ina Amarillo, Bradley Coe, Matthew B. Dobbs, Christina A. Gurnett

ABSTRACT

Introduction  Congenital clubfoot is a common birth defect that affects at least 0.1% of all births. Nearly 25% of cases are familial and the remaining are sporadic in inheritance. Copy number variants (CNVs) involving transcriptional regulators of limb development, including PITX1 and TBX4, have previously been shown to cause familial clubfoot, but much of the heritability remains unexplained.

Methods Exome sequence data from 816 unrelated clubfoot cases and 2645 in-house controls were analysed using coverage data to identify rare CNVs. The precise size and location of duplications were then determined using high-density Affymetrix Cytoscan chromosomal microarray (CMA). Segregation in families and de novo status were determined using quantitative PCR.

Results Chromosome Xp22.33 duplications involving SHOX were identified in 1.1% of cases (9/816) compared with 0.07% of in-house controls (2/2645) (p = 7.98 × 10⁻⁵, OR = 14.57) and 0.27% (38/13592) of Atherosclerosis Risk in Communities (ARIC) control population. PITX1 and TBX4 haploinsufficiency was available for unaffected members of 23 (46%) families. The remaining cases were isolated and do not have known genetic evidence for clubfoot.

Conclusions Microduplications of the pseudoautosomal chromosome Xp22.33 region (PAR1) containing SHOX and downstream enhancer elements occur in ~1% of patients with clubfoot. SHOX and regulatory regions have previously been implicated in skeletal dysplasia as well as idiopathic short stature, but have not yet been linked to clubfoot. SHOX duplications likely contribute to clubfoot pathogenesis by altering early limb development.

BACKGROUND

Congenital clubfoot, also called talipes equinovarus, is a common and serious birth defect that affects an estimated 1 of every 1000 live births. Clubfoot is characterised by structural defects of several tissues of the foot and lower leg, which leads to abnormal positioning of foot and ankle joints. If left untreated, it can result in severe disability and deformity. Approximately 80% of clubfoot cases are isolated birth defects. The remaining 20% of cases are due to associated malformations, or known genetic syndromes, such as arthrogryposis and myleomeningocele. There is also strong evidence for a genetic basis for isolated clubfoot. There is a family history of clubfoot in 25% of all isolated cases. Data from twin studies show a higher concordance in monzygotic (33%) than dizygotic (3%) twins, and more recent data estimates heritability of isolated clubfoot at −30%.

Although few causative genes are known, progress has been made on understanding the genetics of isolated clubfoot. Thus far, the strongest genetic evidence for clubfoot is the PITX1-TBX4-HOXC pathway, the proper function of which is required for normal hindlimb development. Variation in this pathway has been linked to clubfoot phenotypes through a segregating dominant mutation in PITX1, inherited recurrent chromosome 17q23.1q23.2 microduplications containing TBX4, PITX1 haploinsufficiency, and copy number variants (CNVs) and point mutations in SHOX genes including HOXC12. However, these account for only ~1% of familial clubfoot and are rarely identified in non-familial cases.

Here, we report that approximately 1% of paediatric patients with clubfoot has microduplications of SHOX, a transcription factor involved in early limb development. These duplications include the SHOX gene as well as downstream regulatory regions. Four of these microduplications are de novo strengthening their role in clubfoot pathogenesis.

MATERIALS AND METHODS

Cohort description
All patients with clubfoot were recruited at St Louis Children’s Hospital or Shriners Hospital St Louis. Only those with a clinical diagnosis of clubfoot were included. The institutional review board approved this study and all patients and/or parents provided informed consent. We excluded patients with myelomeningocele, arthrogryposis or known genetic disorders. DNA was obtained from either blood or spit from probands, additional affected family members, and parents wherever possible. We include in the present study only those who self-reported their ethnicity as European-American. When family history data are available, it was obtained from the parents of the affected child.

Exome sequencing
Exome sequence data for 816 unrelated clubfoot probands were generated at the McDonnell Genome Institute using IDT xGen Exome Panel V1 capture on Illumina HiSeq 4000 paired-end reads.
Copy-number variation

These cases included 450 sporadic cases with no family history of clubfoot in either first-degree, second-degree or third-degree relatives. In-house control data consisted of 2645 samples, including 1197 with adolescent idiopathic scoliosis, 334 with Chiari 1 Malformation and 433 with male infertility, which were captured and sequenced using the identical IDT platform as the clubfoot cases. An additional 637 controls with amyotrophic lateral sclerosis were captured using Agilent reagents and sequenced on HiSeq 1500/2000.

Exome alignment and BAM file processing

The sequencing reads (ie, fastq files) from all clubfoot cases and control samples were aligned to the human genome reference (GRCh37) using BWA-MEM (V0.7.15). The resulting BAM files were sorted and duplicates were marked using Picard MarkDuplicates (V2.9.0). Re-alignment intervals for each BAM file were determined using GATK (V3.5) RealignerTargetCreator using a list of known indel sites (Mills and 1 kg indels data from the GATK resource bundle ftp://ftp.broadinstitute.org/bundle/b37/), local re-alignment and base quality recalibration were then performed by GATK (V3.5) IndelRealigner and BaseRecalibrator, respectively.

Exome CNV analysis

CNV analysis was performed using the software package Fish-ingCNV that is designed to identify rare CNVs from exome sequencing data without the need for a paired control. We previously used this programme to successfully identify CNVs in an adolescent idiopathic scoliosis cohort.17 This programme uses an algorithm to prioritise rare variants, and compares coverage depth in a test sample against the background distribution, as well as principal components analysis to remove batch effects. The programme extracts coverage information from GATK processed BAM files,18 then normalises read depth information at each exon19 into Reads Per Kilobase per Million mapped reads (RPKM) files. Following our previous protocol, to identify only the most confident CNV calls, we included only CNVs with an average RPKM of at least 10, the mean log-ratio coverage between background and sample of −1.2 to −0.8 (consistent with a deletion) or 0.4 to 0.8 (consistent with a duplication) and Holm–Bonferroni adjusted p value less than 10−5.

Sex chromosome aneuploidy calculations

We estimated sex chromosome aneuploidy by comparing the total read depth on chromosomes X and Y to the average total read depth for autosomes. Read depth was calculated for all samples using Depth of Coverage from the Genome Analysis Tool Kit V3.7 (GATKV3.7). For females, the average X:Autosome ratio was −1 average and Y:Autosome ratio was <0.5. For males, the average X:Autosome ratio was <0.5 and the average Y:Autosome ratio was −1.5.

External control samples

Control samples included the Atherosclerosis Risk in Communities (ARIC) and the Wellcome Trust Case Control Consortium 2 (WTCCC2) NBS and 8C cohorts, for a total of 13 692 samples. ARIC/WTCCC2 controls were profiled on the Affymetrix SNP6 array and processed as per Coe et al (2014). Manual inspection was performed to exclude low-quality CNVs. Due to inherent noise in array data in the pseudoautosomal regions (PARs), an additional check filter was performed on the samples based on their overall noise level as measured by SD of all genomic probes to look for evidence of lost sensitivity in noisier samples. This method detects CNVs from complete chromosomal deletions or duplications down to those 10 Kb in size.

CNV validation using Affymetrix Cytoscan

The Affymetrix Cytoscan HD chromosomal microarray (CMA) consists of more than 750000 SNPs and 1.9 million non-polymorphic probes, and results in 5–10 Kb resolution. Chromosome Analysis Suite was used to view and summarise chromosomal aberrations across the exomes. This array enables the performance of high-resolution, genome-wide CNV analysis and provides genotyping information for the detection of copy neutral loss/absence of heterozygosity, and was used to validate and determine the size of the SHOX containing CNVs detected in clubfoot probands. In some cases, CNVs were broken into multiple contiguous pieces by Cytoscan. We only considered pieces as part of the same CNV if they were <20000 bp apart.

Quantitative PCR of unaffected parents to determine de novo status

For validation of CNVs using quantitative PCR (qPCR), we designed three primer pairs to exonic sequence within the CNV. Primers were tested for quality and efficiency using PCR and real-time qPCR standard curves, respectively. QC was performed on the tested DNA using the Qubit Fluorimeter (Invitrogen). qPCR is performed on a 96 well high-sided low-profile PCR plates (Thermo Scientific) on the Applied Biosystems StepOne Real-Time PCR System. Positive and negative controls with known gains or losses as previously determined by Affymetrix Cytoscan in the gene of interest were included in each run. The RPL27 housekeeping gene was used as a reference.

De novo enrichment analysis and statistical analysis

Fisher’s exact tests were performed in R V3.6.1 on the number of SHOX-containing CNVs in cases versus controls. The R package ‘denovolyze’ was used to assess the enrichment of de novo coding variants on a per gene basis using the estimated de novo mutation rate per gene derived previously from a large exome-sequenced control cohort.21 We estimated the proportion of de novo variants contributing to clubfoot based on the observed number of protein-altering de novo mutations observed in our cohort of clubfoot sporadic uniplex trios (n=450) compared with expected (Nx((M1–M2)/M1), where N is the total number of trios and M, and M2 are the observed and expected count of protein-altering de novo mutations per trio, respectively).

Comparative genomic analysis

The human genomic sequence was retrieved from the UCSC genome browser (www.ucsc.edu/genome). The orthologous sequences from vertebrate species including dog, opossum, chicken, frog and multiple fish species were obtained from the Evolutionarily Conserved Region (ECR) Browser (https://ecrbrowser.dcode.org). As per previous studies, the settings of the ECR browser were adjusted to identify sequences with a minimal length of 100 bp and a 70% homology between species.22 23

RESULTS

SHOX duplications are enriched in patients with clubfoot

SHOX-containing duplications were identified in nine out of 816 (1.1%) clubfoot probands and were the only CNVs significantly enriched in cases (table 1). Identical analysis of our in-house control samples revealed only two duplications out of 2645 controls (0.07%) (p=7.98×10−5, OR=14.57). This
Table 1  Frequency of X chromosome duplications and deletions in cases and controls

<table>
<thead>
<tr>
<th>CNV region</th>
<th>CNV type</th>
<th>Chromosomal coordinates</th>
<th>Cases (n=816)</th>
<th>Controls (n=2645)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>47, XXY</td>
<td>Duplication</td>
<td>Entire chromosome</td>
<td>5</td>
<td>0</td>
<td>7.4×10^{-4}</td>
</tr>
<tr>
<td>Xp22.33</td>
<td>Duplication</td>
<td>2 008 854–2 729 493</td>
<td>9</td>
<td>2</td>
<td>7.9×10^{-5}</td>
</tr>
<tr>
<td>Xq22.33</td>
<td>Duplication</td>
<td>1 401 596–3 631 294</td>
<td>11</td>
<td>31</td>
<td>0.4</td>
</tr>
<tr>
<td>Xp11.4</td>
<td>Duplication</td>
<td>37 701 034–38 009 147</td>
<td>1</td>
<td>2</td>
<td>0.6</td>
</tr>
<tr>
<td>Xp11.23</td>
<td>Duplication</td>
<td>48 265 385–48 270 814</td>
<td>1</td>
<td>3</td>
<td>0.7</td>
</tr>
<tr>
<td>Xq22.2-q23</td>
<td>Duplication</td>
<td>102 973 984–109 310 624</td>
<td>2</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Xq26.3</td>
<td>Duplication</td>
<td>134 978 401–135 586 622</td>
<td>2</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Xq28</td>
<td>Duplication</td>
<td>152 241 445–155 235 144</td>
<td>9</td>
<td>14</td>
<td>0.07</td>
</tr>
<tr>
<td>Xp11.23</td>
<td>Deletion</td>
<td>41 307 142–46 508 232</td>
<td>2</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Xq28</td>
<td>Deletion</td>
<td>155 003 533–155 235 853</td>
<td>1</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

enrichment was also significant when compared with 13,692 ARIC/WTCCC2 controls, where 38 individuals were found to have high-confidence SHOX duplications (p=0.001, OR=3.97).

All nine chromosome Xp22.33 duplications detected using FishingCNV on exome data were validated using Affymetrix Cytoscan arrays and the size of online supplementary tables S1 and S2. The duplications were more precisely determined, and varied from 235 kb to 2.528 Mb online supplementary tables S1 and S2. No microdeletions of SHOX were detected in the clubfoot probands.

Aneuploidy in the sample
We evaluated 816 clubfoot probands for sex chromosome aneuploidy and sex chromosome CNVs using exome read depth data. Five out of 564 male probands (0.9%) were discovered to have 47, XXY or Klinefelter Syndrome based on X:Autosome ratios consistent with female sex (range: 0.85–1.1) but also had Y:Autosome ratios consistent with male sex (range: 1.4–1.7) and were reported as biological male. We removed these aneuploid samples prior to further CNV analysis. There were no clubfoot probands with 45, X (Turner Syndrome) or 47, XXX. Our in-house controls and ARIC/WTCCC2 controls could not be used for comparison due to the exclusion of patients with aneuploidy from both datasets prior to generation of genetic data.

Duplication of SHOX and downstream regions containing known evolutionary conserved enhancers
All nine clubfoot probands share an overlapping 180.28 kb duplication of SHOX and portions of the 200 kb downstream position effect (PE) region, containing known ECRs and enhancers (figure 1). An affected great uncle of 7,470,001 (individual 7,470,004) with clubfoot was also evaluated and found to have a small 1.41 kb chromosome Xp22.33 duplication on Affymetrix Cytoscan array, although his duplication was much smaller than the proband. When he is included, the overlapping duplicated region shared by all 10 clubfoot cases is narrowed down to a 1.41 kb portion of the downstream PE region (figure 1). This 1.41 kb region contains a 131 bp ECR composed of transposable elements and simple repeats.

Clinical phenotypes associated with chromosome Xp22.33 duplications
Clinical evaluations revealed that seven out of nine probands with the duplication were men (77.7%), whereas the entire clubfoot cohort evaluated was 68% men. Height data were available for seven out of nine clubfoot probands. Six patients were <50th percentile of height for their age, and five of the seven were <10th percentile (table 2). Three probands were...
found on follow-up to have mild global developmental delay, and two of these probands were also found to have additional known recurrent CNVs associated with developmental delay phenotypes along with the SHOX duplication. Eight of the nine probands had bilateral clubfoot. Two probands presented with hand abnormalities, including symbrachydactyly and camptodactyly. One proband had adolescent idiopathic scoliosis. Only one proband (7470001) had a family history of clubfoot, and his affected great uncle also had a chromosome Xp22.33 microdeletion, although much smaller than the proband.

Four de novo SHOX duplications among those with available DNA

Of the nine probands with SHOX duplications, DNA was available for both unaffected parents of six probands. DNA of one set of parents was analysed with Affymetrix Cytoscan and the other five sets were analysed with qPCR. Four out of six SHOX duplications were not present in the unaffected parents but were confirmed in the proband, resulting in a de novo rate of 66% (4/6) for chromosome Xp22.33 duplications. When we consider that only 450 out of the 816 clubfoot probands evaluated were considered sporadic cases because they did not have a family history of clubfoot, the probability of detecting four de novo mutations in the SHOX gene by chance in a cohort of this size (n=450) is 5.4×10^{-10}.

DISCUSSION

We report evidence supporting a new role for SHOX gene duplications in clubfoot pathogenesis. First, we have shown that approximately 1% of clubfoot probands have a microduplication of SHOX and that there is a significant enrichment of de novo duplications. Short stature homeobox (SHOX) is located in the PAR1 on the distal end of the short arms of the X and Y chromosomes, and encodes a homeodomain transcription factor expressed in the developing forelimb and lower leg, as well as on the first and second pharyngeal arches.24 SHOX escapes X-inactivation, hence XX and XY individuals both have two copies of SHOX.25 SHOX-related short stature phenotypes, including growth failure and limb deformities, are observed in patients with Turner syndrome who have only one copy of SHOX.26 Point mutations or deletions resulting in haploinsufficiency of SHOX are also associated with idiopathic growth retardation,27 and idiopathic short stature (ISS) and are causative of Leri–Weill dyschondrosteosis (LWD), a dominant skeletal dysplasia presenting with mesomelic limb shortening and curving of the radius known as Madelung deformity.28 29 Homozygous loss of SHOX or compound heterozygous SHOX defects result in the more severe phenotype known as Langer mesomelic dysplasia (LMD).30 31 32 Mesomorphic involvement of the lower legs observed in patients with LWD and LMD with SHOX defects is consistent with the predominant anatomic abnormalities previously described in patients with idiopathic clubfoot, in which skeletal muscle hypoplasia is most evident in the lower legs but not the calf.32 33 Duplications of the SHOX region are less often described, though they have been associated with tall stature in one study.34 However, positive effects on height appear to require duplication of the entire SHOX coding and regulatory region, as partial duplications have been associated with short stature.32 33 For instance, a girl with severe short stature and agenesis of the right tibia and fibula was found to have large duplications flanking either side of the SHOX gene, but a normal copy number of the gene itself.35 In fact, SHOX microduplications represent a small fraction of the aetiology of ISS because SHOX enhancer duplications may impair gene expression as extra copies of the regulatory region can dysregulate SHOX.37 Many of our clubfoot probands with SHOX duplications also had short stature and none of them were tall. While variable hand anomalies were observed in two of our clubfoot probands with the SHOX duplication, we do not know whether the preferential involvement of the lower leg is a consequence of selection bias due to our recruitment from clubfoot clinic or reflects enhanced susceptibility of the lower leg. It is also important to note that analysis of CNVs using exome data has some limitations, most notably that potential intronic or intergenic regions CNVs acting through regulatory effects may be missed. As in CNV analysis of whole genome data, detection is only as good as the sequence coverage, and more even coverage will always lead to cleaner results. The non-coding region downstream of SHOX that is located within the 180 kb common clubfoot deletion interval is known as a position effect (PE) of SHOX31 because alterations can affect gene expression even when the gene is intact.32 33 Because the SHOX duplication common region shared among all patients with clubfoot also contains long-range enhancers of the gene,10 39 we cannot exclude the possibility that these non-coding regions also play a role in clubfoot pathogenesis. Deletions of these

<table>
<thead>
<tr>
<th>ID</th>
<th>Current age</th>
<th>Sex</th>
<th>Type of clubfoot</th>
<th>CDC height percentile</th>
<th>Duplication size (kb)</th>
<th>De novo</th>
<th>Additional history</th>
</tr>
</thead>
<tbody>
<tr>
<td>5162001</td>
<td>13.8</td>
<td>F</td>
<td>LTEV</td>
<td>48</td>
<td>301.166</td>
<td>Unknown</td>
<td>Adolescent idiopathic scoliosis</td>
</tr>
<tr>
<td>5310001</td>
<td>14.2</td>
<td>M</td>
<td>BTEV</td>
<td>Unknown</td>
<td>277.152</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>5333001</td>
<td>14.2</td>
<td>M</td>
<td>BTEV</td>
<td>Unknown</td>
<td>241.258</td>
<td>Y</td>
<td>Developmental delay, chr 16p13.11 duplication</td>
</tr>
<tr>
<td>7436001</td>
<td>7.9</td>
<td>M</td>
<td>BTEV</td>
<td>56</td>
<td>2528.216</td>
<td>Y</td>
<td>Bilateral camptodactyly, great uncle with clubfoot</td>
</tr>
<tr>
<td>7470001</td>
<td>1.9</td>
<td>M</td>
<td>BTEV</td>
<td>4.7</td>
<td>847.043</td>
<td>Y</td>
<td>Developmental delay</td>
</tr>
<tr>
<td>7282001</td>
<td>3.0</td>
<td>M</td>
<td>BTEV</td>
<td>3.4</td>
<td>659.028</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>7506001</td>
<td>2.4</td>
<td>M</td>
<td>BTEV</td>
<td>8.4</td>
<td>235.578</td>
<td>N</td>
<td>Amniotic band syndrome, bilateral symbrachydactyly, sixth nerve palsy</td>
</tr>
<tr>
<td>7482001</td>
<td>2.1</td>
<td>F</td>
<td>BTEV</td>
<td>0.1</td>
<td>847.043</td>
<td>Y</td>
<td>Torticollis, chr 2q37.3 deletion, developmental delay</td>
</tr>
<tr>
<td>7096001</td>
<td>16.8</td>
<td>M</td>
<td>BTEV</td>
<td>3.7</td>
<td>847.043</td>
<td>unknown</td>
<td>Developmental delay</td>
</tr>
<tr>
<td>7470004</td>
<td>45.3</td>
<td>M</td>
<td>BTEV</td>
<td>67</td>
<td>1.415</td>
<td>unknown</td>
<td>Great uncle of 7470001</td>
</tr>
</tbody>
</table>

BTEV, bilateral talipes equinovarus; LTEV, left talipes equinovarus.
enhancers, without involvement of the SHOX transcript, have previously been shown to be enriched in patients with ISS and LWS and absent in controls, although there is evidence of incomplete penetrance in families.

Evolutionary sequence conservation has previously been used as a guide to locate potential enhancers downstream of the SHOX gene, and is a good predictor for the identification of cis-regulatory elements, particularly in developmental genes. Comparative genomic analysis identified eight highly conserved non-genic elements (CNEs) located in the SHOX downstream region, and three have previously been shown to have cis-regulatory activity in the developing limb bud in chicken embryos. Duplications of these PEs and CNEs are not as well characterised as deletions. While we narrowed the causative clubfoot genetic region to a 180 kb interval that contains both the SHOX gene and a large downstream regulatory region, the interval may be as small as 1.41 kb if we include the duplication in an affected great uncle that was much smaller than that present in the proband (online supplementary tables S1 and S2). It is important to note that the variant in the great uncle and corresponding proband are not the same variant; therefore, they may not be inherited from a common ancestor, so this does not impact the number of de novo variants found here. A single ECR is located in the very smallest 1.41 kb region but is only conserved back to the dog lineage. However, previous research of SHOX enhancers has shown that some less highly evolutionarily conserved elements also have enhancer activity.

In a prior analysis of autosomal CNVs, our group showed that CNVs involving transcription factors or transcriptional regulators of hindlimb development segregate with clubfoot in several families. These include specific regulators of hindlimb development, including PITX1, TBX4 and SHOX genes. Therefore, it is not surprising that we find an enrichment of CNVs in yet another transcription factor, this time one that is expressed on the X-chromosome. Our identification of four de novo duplications of the SHOX gene strongly suggests a causative role in clubfoot pathogenesis. Further evidence for a role of SHOX in clubfoot pathogenesis is provided by studies showing that recurrent chromosome 17q23.1-q23.2 microduplications involving Tbx4, a transcriptional regulator of the closely related gene Sho2 during limb development in mice, are among the most common known causes of familial clubfoot.

Although the vast majority of the patients in this study had isolated clubfoot, at least when recruited for the study during infancy, three out of nine duplication carriers were noted to have mild developmental disabilities on long-term follow-up. Two of these individuals had other CNVs, in addition to the SHOX duplication. The chromosome 16p13.11 duplication and chromosome 2q37.3 microdeletion that we identified in these patients are both known to cause neurodevelopmental abnormalities, but not clubfoot, and are therefore likely explain their developmental disabilities since most of our SHOX duplication carriers are developmentally normal. Other investigators have also noted a tendency for individuals to have multiple CNVs. Recent studies have shown that there is a slightly higher prevalence of neurodevelopmental difficulties in children with isolated clubfoot which is also consistent with this observation.

We also identified five individuals in our clubfoot cohort with 47,XXX karyotypes, which is approximately six times greater than the population average of 0.15% or ~1/660 males. Although there are nearly 800 genes duplicated in patients with 47,XXX karyotype, we hypothesise that duplication of SHOX may be responsible for the clubfoot phenotype in these males. We did not identify any females with clubfoot who had an 47,XXX karyotype, suggesting that males may be particularly susceptible to the effects of SHOX duplications. Overall, the over-representation of males with SHOX duplications in our study is consistent with the 3:1 male:female ratio for clubfoot. Interestingly, the opposite is seen in patients with LWD and SHOX deficiency, as Madelung deformity is both more common and more severe in females. Our identification of SHOX duplications and enrichment of 47,XXX karyotype abnormalities in patients with clubfoot, along with our prior work identifying CNVs involving other regulators of early limb development, including PITX1, TBX4 and SHOX genes, suggests that it may be useful to screen patients with clubfoot for CNVs, particularly when there is any concern about development.

Our results show that SHOX duplications are present in ~1% of all patients with clubfoot. This study will have immediate and direct impact on patient care because these genetic abnormalities can easily be detected with clinically available CMA studies. In addition, our results provide important insight into the pathogenetic mechanisms of clubfoot now that we have shown a strong link to abnormalities in multiple early transcriptional regulators of limb development. Furthermore, the enrichment of SHOX duplications in males may partially explain the increased prevalence of clubfoot among males. Future research employing high-resolution methods to detect even smaller X chromosome CNVs may help us to understand the clubfoot pathogenesis in the remaining unexplained cases, and may also provide insight into the role of non-coding regulatory regions in human development.

Acknowledgements We thank patients and their families for their participation, as well as Amy Young and Tim Kuensting. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk/. We also thank Dr Evan Eichler for providing access to the external control samples.

Contributors BS, GH, MBD, CAG conceived of the study; BS and GH carried out statistical analyses, LA compiled exome data; MN performed all qPCR experiments and helped with methods writing, IA provided additional advice on CNV location methods; BC provided and analysed external control datasets; all authors edited the manuscript.

Funding Research reported in this publication was supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases under Award Number R01AR067715, Eunice Kennedy Shriver National Institutes of Child Health and Human Development of the National Institutes of Health under the award number P01HD083871, the Marfan Foundation Faculty Grant award #81831, Washington University Institute of Clinical and Translational Sciences grant UL1 TR002345 from the National Center for Advancing Translational Sciences of the National Institutes of Health, Washington University Musculoskeletal Research Center (NIH/Niams P30 AR057235), the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number U54 HD070111 to the Intellectual and Developmental Disabilities Research Center at Washington University. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICS/TCTSA Grant# U10RR24992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially funded by NIH grants 1S10RR022984-01A1, and 1S100018091-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Hope Center DNA/RNA Purification Core at Washington University School of Medicine subsidised the cost of extracting DNA for this study.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Deidentified exome data from this study are available upon request. Please contact the corresponding author.

Copy-number variation

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID id
Brooke Sadler http://orcid.org/0000-0002-3268-3809

REFERENCES
Copy-number variation


