Cytogenetic and Genome Research

Cytogenet Genome Res , DOI: 10.1159/000530171
Received: December 19, 2022
Accepted: March 13, 2023
Published online: March 16, 2023

SHOX whole gene duplications are over-represented in SHOX haploinsufficiency phenotype cohorts

Bunyan DJ, Hobbs JI, Duncan-Flavell PJ, Howarth RJ, Beal S, Baralle D, Thomas NS

ISSN: 1424-8581 (Print), eISSN: 1424-859X (Online)
https://www.karger.com/CGR
Cytogenetic and Genome Research

SHOX whole gene duplications are over-represented in _SHOX_ haploinsufficiency phenotype cohorts

David J Bunyan[a,b], James I. Hobbs[a], Philippa J. Duncan-Flavell[a], Rachel J Howarth[a], Sarah Beal[a], Diana Baralle[b,c] and Nicholas Simon Thomas[a,b]
[a] Wessex Genomics Laboratory Service, Salisbury District Hospital, Salisbury, Wiltshire, UK; [b] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, Hampshire, UK; [c] Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, UK

Short Title: *SHOX* whole gene duplications and haploinsufficiency syndromes

Corresponding author:
Dr David J Bunyan, Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, Wiltshire, SP2 8BJ, UK
Tel . +44 1722 429012
Email: dbunyan@nhs.net

Number of Tables: 2
Number of Figures: 1
Word count: 4603

Keywords: SHOX, Regulatory element, Duplication, Haploinsufficiency, Transcription

Abstract

Transcription of *SHOX* is dependent upon the interaction of the gene with a complex array of flanking regulatory elements. Duplications that contain flanking regulatory elements but not the *SHOX* gene have been reported in individuals with *SHOX* haploinsufficiency syndromes, suggesting that alterations to the physical organisation or genomic architecture may affect *SHOX* transcription. Individuals with tall stature and an additional X or Y chromosome have an extra copy of both the *SHOX* gene and the entire *SHOX* regulatory region, so all three copies of *SHOX* can be expressed fully. However, for a duplication of the *SHOX* gene that does not include all of the flanking regulatory elements, the potential effect on *SHOX* expression is difficult to predict. We present nine unpublished individuals with a *SHOX* whole gene duplication in whom the duplication contains variable amounts of the *SHOX* regulatory region, and we review 29 similar cases from the literature where phenotypic data were clearly stated. While tall stature was present in a proportion of these cases, we present evidence that *SHOX* whole gene duplications can also result in a phenotype more typically associated with *SHOX* haploinsufficiency and are significantly over-represented in Leri-Weill Dyschondrosteosis and idiopathic short stature probands compared to population controls. Although similar-looking duplications do not always produce a consistent phenotype, there may be potential genotype-phenotype correlations regarding the duplication size, regulatory element content and the breakpoint proximity to the *SHOX* gene.

Although ClinGen (clinicalgenome.org) do not currently consider *SHOX* whole gene duplications to be clinically significant, the ClinGen triplosensitivity score does not take into account the context of the duplication, and more is now known about *SHOX* duplications and the role of flanking elements in *SHOX* regulation. The evidence presented here suggest that these duplications should not be discounted without considering the extent of the duplication and the patient phenotype, and should be included on diagnostic laboratory reports as variants of uncertain significance. Given the uncertain pathogenicity of these duplications, any reports should encourage the exclusion of all other causes of short stature where possible.

Introduction

Haploinsufficiency of *SHOX* results in phenotypes ranging from Leri-Weill Dyschondrosteosis (LWD; MIM ID#127300) [Leri and Weill 1929] to short stature [Marchini et al., 2016]. Idiopathic short stature (ISS; MIM ID#300582) is defined as a height below the 3rd centile in the absence of a known specific causative disorder [Wit et al., 2008], and although *SHOX* loss-of-function variants are highly penetrant, clinical expression is extremely variable, even within the same family [Binder et al., 2003; Stuppia et al., 2003; Huber et al., 2006; Jorge et al., 2007; Rappold et al., 2007].
SHOX regulation is highly complex, with long-range enhancers both 5' and 3' of the gene. Comparative genomic studies identified multiple Conserved Non-coding DNA Elements (CNEs) downstream of *SHOX*, four of which have transcriptional activity: CNE4 (X:714,085-714,740 (hg19)), CNE5 (X:750,825-751,850), CNE7 (ECR1; X:780,700-781,220) and CNE9 (ECS4; X:834,746-835,567) [Fukami et al., 2006; Sabherwal et al., 2007; Chen et al., 2009; Benito-Sanz et al., 2012]. The ZED element (Zeugopodal Enhancer Downstream of *SHOX*) [Skuplik et al., 2018] was shown to be the critical functional element within the recurrent 47.5 kb X:780,550-828,092 downstream deletion [Benito-Sanz et al., 2012; Bunyan et al., 2013] and an additional cis-regulatory element was proposed around X:970,000 [Bunyan et al., 2014] that would potentially further extend the downstream regulatory region. Three active *SHOX* upstream CNEs have also been demonstrated, namely CNE-5 (X:398,357-398,906), CNE-3 (X:460,279-460,664) and CNE-2 (X:516,610-517,229) [Durand et al., 2010]. Duplications that contain only flanking *SHOX* regulatory elements have been reported in individuals with *SHOX* haploinsufficiency syndromes [Benito-Sanz et al., 2011; Fukami et al., 2015; Bunyan et al., 2016; Hirschfeldova and Solc 2017; Sadler et al., 2020; Bunyan et al., 2021], suggesting that the physical organisation of the *SHOX* regulatory region is critical, so alterations to the physical separation or to the intervening genomic architecture may affect *SHOX* transcription.
An additional copy of the entire *SHOX* gene and all associated regulatory elements would be expected to cause overexpression, and tall stature is typically observed in individuals with an additional sex chromosome (47,XXX, 47,XXY or 47,XYY) [Ottesen et al., 2010]. However, the consequences of duplications which include the entire *SHOX* gene but only part of the regulatory "cassette" are hard to predict, even though many cases have been described in the literature [Thomas et al., 2009; Benito-Sanz et al., 2011; Brosens et al., 2014; Fukami et al., 2015; Donze et al., 2015; Tropeano et al., 2016; Hirschfeldova and Solc 2017; Upners et al., 2017; Sadler et al., 2020]. *SHOX* whole gene duplications are associated with a wide range of phenotypes with no established genotype-phenotype correlations, making genetic counselling and clinical management difficult. One hypothesis is that the phenotypic variability may be directly related to the regulatory element content of the duplicated interval, and/or to the location of the duplicated fragment. We can test this hypothesis by studying individuals with *SHOX* whole gene duplications that contain a variable number of enhancers, i.e. where at least one breakpoint of the duplication maps within the X:398,357-970,000 interval which contains all currently known or proposed cis-regulatory elements. We describe nine previously-unreported individuals with a *SHOX* whole gene duplication that includes only part of the X:398,357-970,000 region and have identified 29 further cases from the literature with a similar duplication where the individual's phenotype was clearly stated. We have determined the frequencies of these duplications within various study cohorts and compared them to control cohorts. To limit ascertainment bias, our results include data from in-house and published array comparative genome hybridisation (aCGH) cohorts to determine how frequently probands with a *SHOX* whole gene duplication, detected incidentally, have an ISS or LWD phenotype.

Materials and Methods

The novel probands presented in this study were tested at the Wessex Regional Genetics Laboratory as part of the National Health Service (NHS). Purified genomic DNA obtained from an EDTA blood sample was extracted according to standard protocols.

This is a retrospective study. Within the NHS in England, probands with isolated short stature are only eligible for *SHOX* analysis and have no access to Whole Genome Sequencing (WGS) and are not consented for research. However, where patients have had additional testing, we have listed this information in Table 1.

Novel duplications in this study

Patients 1-2 were identified from a cohort of 1,959 referrals from local, national and international referrers between June 2003 and March 2020 for *SHOX* testing only. Analysis of *SHOX* and its flanking regions was carried out using Multiplex Ligation-dependent Probe Amplification (MLPA) and Sanger sequencing. For Patients 1-2, the duplications sizes were further defined using aCGH. Patients 3-9 were identified from 22,018 individuals referred for aCGH (mostly investigated for developmental delay) from local, national and international referrers between March 2009 and March 2020. Individuals in the aCGH cohort were not specifically referred for *SHOX* analysis. Probands with *SHOX* whole gene duplications from this cohort were identified solely to allow a comparison of their frequency versus the *SHOX* cohort, and also to determine the presence of *SHOX*-related phenotypes in independently-ascertained individuals with *SHOX* whole gene duplications.

Methods

a) MLPA [Schouten et al., 2002] was performed using the current *SHOX* kit at the time of testing according to the manufacturer's protocol (P018; MRC-Holland, Amsterdam, The Netherlands). The current kit version (P018-G2) contains probes for every exon of the *SHOX* gene and every CNE shown in Figure 1. The proposed X:970,000 regulatory element does not contain an MLPA probe but is flanked by probes at approximately X:963,700 and X:1,029,700.
b) Direct sequencing of all coding exons (isoform A, NM_000451.3, exons 2 to 6a) was used to exclude the presence of single nucleotide variants and small deletions/insertions in the *SHOX* coding sequence and intron/exon boundaries (primer sequences available upon request).
c) aCGH was performed using Oxford Gene Technologies (OGT, Oxford, UK) 60-mer oligo-array printed in 8x60 K International Standard Cytogenomic Array (ISCA) Consortium configuration, according to manufacturer's instructions, using Kreatech's pooled control DNA as a reference (Kreatech Diagnostics, Amsterdam, Holland). Slides were scanned using a G2539A Agilent microarray scanner (Agilent Technologies, Wokingham, UK) and analysed using OGT's CytoSure Interpret (v3.6) microarray software.

Phenotypes

The phenotypes of the nine novel individuals with a *SHOX* duplication are given in Table 1, together with the phenotypes of the 29 probands from the literature with similar duplications. Parental samples were received for five of the nine novel probands identified in our laboratory. We do not have accurate heights for four of the parents who carry the same duplication as their offspring and we have been unable to retrospectively obtain this information, although the original referral forms stated that the father of Patient 3 is "not particularly short" and the father of Patient 8 is "phenotypically normal". The mother of Patient 9 is 155cm tall, putting her on the 17th centile.

Previously-reported duplications

The first published collection of *SHOX* whole gene duplications contained four cases [Thomas et al., 2009]. One case was ascertained through screening a cohort of patients with Madelung deformity and had height on the 11th centile, suggestive of a possible diagnosis of LWD. The other two cases were originally referred for aCGH analysis (because of Asperger syndrome and familial cleft palate respectively). As the cohort sizes were not given, these three cases have not been included in any detection rate calculations. The fourth proband from this publication has been excluded from the

genotype/phenotype component of this study as they also have a *SHOX* whole gene deletion, considered the explanation for their diagnosis of LWD. All individuals were negative for pathogenic *SHOX* sequence variants.
In a subsequenct study, MLPA analysis of 122 LWD and 613 ISS referrals identified *SHOX* whole gene duplications in three individuals with ISS and one with LWD [Benito-Sanz et al., 2011]. This study also included controls, and no *SHOX* whole gene duplication was identified in 340 individuals with normal stature (relative to age and gender) or 104 tall stature individuals with height above the 99th centile. MLPA analysis also identified a duplication in a further four ISS individuals, two from an unspecified number of patients with short stature [Donze et al., 2015], one from a cohort of 245 patients with ISS or LWD [Fukami et al., 2015], and one from a Czech cohort of 352 ISS or LWD patients [Hirschfeldova and Solc 2017]. The significance of the duplication in the latter Czech case was questioned in the manuscript because a different *SHOX* whole gene duplication was identified in one of the 250 population control individuals (whose height was on the 75th-90th centile).
Six other relevant cases in the literature came from cohorts screened specifically for *SHOX* dosage abnormalities (as *SHOX* was considered to be a likely cause of the phenotype under investigation): (1) The first study was a cohort of 81 girls with tall stature [Upners et al., 2017] which identified two relevant *SHOX* whole gene duplications in individuals with a height above the 99th centile and a normal karyotype; (2) Another four cases were identified in a cohort of 816 unrelated individuals with club foot (*talipes equinovarus*) [Sadler et al., 2020], three in probands with bilateral clubfoot and one with unilateral clubfoot. Heights were only available for two of the individuals with club foot; one was on the 8th centile and the other was on the 48th. In this latter manuscript, no similar duplications were detected in any of the 2,645 in-house controls which included 1197 with adolescent idiopathic scoliosis, 334 with Chiari 1 malformation, 433 with male infertility and 637 with amyotrophic lateral sclerosis.
Finally, two published manuscripts reported individuals with *SHOX* whole gene duplications, incidentally ascertained with regard to height, by aCGH genome-wide testing: (1) The first was a cohort of 180 patients with esophageal atresia that identified a *SHOX* duplication in a single patient with ISS [Brosens et al., 2014]. This male patient had a height on the 2nd centile and limb anomalies; (2) The second was a very large and detailed study that tested 26,664 individuals with Autistic Spectrum Disorder (ASD) plus 12,594 controls [Tropeano et al., 2016]. This latter study identified 55 individuals with *SHOX* whole gene duplications (48 in individuals with a neurodevelopmental disorder and 7 in the non-ASD aCGH cohort), but sufficient clinical information for inclusion in this manuscript was only provided for nine of these cases – two where club foot is mentioned and seven where the height is listed (two >97th centile, three <3rd centile, one with a height on the 25th-50th centile and one with a height on the 50th-75th centile). The 12,594 controls were selected on the basis that they did not have ASD but they had been referred for aCGH testing because of other phenotypes such as congenital malformations, physical dysmorphism, growth/skeletal abnormalities, and endocrine/metabolic conditions, so for results purposes we have treated these individuals as a mixed aCGH cohort rather than population-based normal controls.

Results

Details of the nine novel and 29 published duplications are set out in Table 1 and Figure 1. All chromosomal location data are based on the hg19 GRCh37 build. For all probands, the minimum and maximum duplication sizes were sufficiently determined to allow the regulatory element content to be fully defined.
The overall incidence of *SHOX* whole gene duplications detected in patients referred to our laboratory for diagnostic *SHOX* testing was 4/1,959. Of these four cases, two were excluded from Table 1 and Figure 1 because of the presence of a second *SHOX* variant, so the effect of the *SHOX* whole gene duplication in those individuals could not be clearly determined. The overall incidence of *SHOX* whole gene duplications in individuals tested by aCGH in our laboratory was 9/22,018. Two of these nine cases were excluded from Table 1 and Figure 1: the first because we were unable to

obtain any clinical information, and the second because, although their height was <2nd centile, they also had abnormal vertebral segmentation with fused vertebrae and several absent vertebral pedicles which was deemed likely to be the major cause of their height loss, so any compounding effect of the *SHOX* duplication could not be determined. No *SHOX* whole gene duplications were identified by MLPA in 471 anonymised normal controls in our laboratory (previously published in Bunyan et al., 2013). These controls were variant-negative individuals who had undergone carrier-testing for autosomal recessive conditions, or were the parents of patients with a *de novo* structural abnormality. No height data were available on the control group, but all had been seen by a clinical geneticist prior to referral, so are unlikely to have a phenotype that would bring them to clinical attention and are expected to be representative of the general population.

A summary of clinical information is provided for all 38 individuals in Table 1, taken from either the diagnostic referral, the relevant publication or retrospective information from the referring clinician. The clinical information has been used to assign a specific phenotype designation. We categorised 32 of the probands as either LWD (n=2), ISS (n=18), tall stature (n=6) or club foot (n=6). For the remaining six we have listed their heights in centiles and categorised them as "4th-96th centile". While some of the *SHOX* duplications identified in our laboratory or published in the literature were excluded from Table 1, either due to the lack of clinical information or the presence of a second *SHOX* variant, Table 2 includes the frequency of all *SHOX* whole gene duplications detected in every cohort where the total size of the test cohort was provided. Table 2 shows that *SHOX* whole gene duplications are present at a low level in anonymised control cohorts (1/3,721; 0.03%) and in mixed aCGH cohorts (16/34,612; 0.05%). In contrast, in the LWD/ISS group, whole gene duplications were seen in 0.33% of cases (11/3,291).

In aCGH cohorts (our study and Tropeano et al., 2016) where *SHOX* whole gene duplications were detected incidentally (as opposed to targeted *SHOX* testing), a minimum of seven of the 64 probands have ISS (10.9%). As we do not have definitive clinical data for 46 of these 64 probands, the actual incidence of ISS in the mixed aCGH group may be higher.

Discussion/Conclusion

Alterations in *SHOX* expression have significant clinical consequences and are associated with a wide range of phenotypic presentations. *SHOX* haploinsufficiency is a common cause of short stature and can also include additional skeletal features, such as Madelung deformity, in individuals with LWD. However, because the regulation of *SHOX* is so complex and subject to long range position effects, it is difficult to precisely define which *SHOX* variants would result in loss of function. Establishing that specific microdeletions within the *SHOX* regulatory region are pathogenic has been very successful, even if phenotypic variability can confound segregation analysis. Investigating whether duplications within the *SHOX* regulatory region are pathogenic is much more challenging as the mode of pathogenicity is not obvious. The identification of duplications of flanking regulatory elements that do not include the *SHOX* gene in patients with various *SHOX*-related phenotypes [Benito-Sanz et al., 2011; Fukami et al., 2015; Bunyan et al., 2016; Hirschfeldova and Solc 2017; Sadler et al., 2020; Bunyan et al., 2021] suggest that a general disruptive effect on genome architecture may explain the presence of a *SHOX*-related phenotype and that the maintenance of the *SHOX* region genomic architecture is critical to normal gene function.

There is strong evidence, including four *de novo* cases, that duplications of the *SHOX* gene are associated with club foot [Sadler et al., 2020], plus a statistically significant over-representation in individuals with ASD [Tropeano et al., 2016]. However, single reports of *SHOX* CNVs from a specific clinical cohort can lead to ascertainment bias, therefore in this study we have brought together individuals from multiple different cohorts. The duplications were identified in patients with a wide range of phenotypes, including 20 individuals with phenotypes more typically associated with *SHOX* haploinsufficiency (LWD and ISS) and six with tall stature suggestive of *SHOX* over-expression. We present nine individuals with unpublished duplications, three identified through targeted *SHOX* analysis and six detected through diagnostic aCGH testing. We also describe 29 previously reported

individuals with partial duplications of the *SHOX* regulatory region (including the entire *SHOX* gene) where the clinical phenotype was clearly stated.

The prevalence of whole gene duplications in ISS/LWD cohorts provides evidence that they can cause *SHOX* haploinsufficiency (see Table 2): in this study, such duplications have a much higher prevalence in *SHOX*-specific LWD/ISS cohorts (11/3,291; 0.33%) than in population controls (1/3,721; 0.03%), a statistically significant increase (□2 (1, N = 7012) = 9.6, p < .05). In order to limit ascertainment bias, we have also looked at the frequency of *SHOX* whole gene duplications detected by aCGH where the referrals were unrelated to height. Although the overall incidence (16/34,612; 0.05%) was very similar to controls, the number of probands with ISS was higher than expected by chance. ISS has an incidence of 2.3% in the general population [Pedicelli et al., 2009], but in the aCGH cohorts a minimum of seven of the 64 probands (10.9%) have ISS. We do not have definitive clinical data for 46 of these 64 probands, so the actual incidence of ISS in the mixed aCGH group may be even higher. However, many of the mixed aCGH probands may have a secondary genetic diagnosis that includes short stature as part of a wider syndrome, and the same may be true of some members of the *SHOX*/ISS cohort. Although a secondary genetic diagnosis may provide an explanation for the high level of ISS in aCGH probands with *SHOX* whole gene duplications, it would not explain the increased prevalence of such duplications in the *SHOX*/ISS cohort.

The duplications presented in this study are extremely rare, and while some may share a common breakpoint, of the 38 duplications in Table 1 there are at least 36 different breakpoint combinations. However, it is clear from Figure 1 that similar duplications do not always produce a consistent phenotype. The duplications are frequently inherited from a phenotypically-normal parent, so there are high levels of phenotypic variability even in individuals with the same variant, and some of the duplications may be co-incidental findings. This heterogeneity makes it difficult to extrapolate information from one duplication to another, and both laboratory reporting and genetic counselling are challenging. The data provide evidence that a subset of *SHOX* whole gene duplications can result in *SHOX* haploinsufficiency, but assigning causality to any individual duplication is very difficult, and excluding all other causes of ISS is not straightforward. The ACMG guidelines for interpreting copy number variants [Rooney Riggs et al., 2020] are not designed for genes with variable penetrance, and heavy weighting is given to the ClinGen (clinicalgenome.org) triplosensitivity score. For *SHOX*, this is 0, suggesting that *SHOX* whole gene duplications are not currently thought to be clinically important. However, this assessment is likely to be in the context of an intact *SHOX* regulatory region and an assumption that three copies of *SHOX* will result in over-expression.

For whole gene duplications that do not include all enhancers, several factors should to be taken into consideration. Firstly, is the location of the duplicated interval known? In order to disrupt *SHOX* regulation, the duplicated interval would be expected to reside within the *SHOX* regulatory region such that two copies of *SHOX* are competing for the same enhancers. Approximately 95% of large duplications genome-wide are reported to be tandem [Richardson et al., 2018]. The location of the duplicated fragment was only investigated in four of the ISS cases presented in this manuscript. In Patient F1 the duplication was proven to be a direct tandem repeat by Sanger sequencing, and in Patients F2, P3 and B1 fluorescence in situ hybridisation analysis (FISH) gave a single signal at Xp22.3. Where divergent phenotypes are seen in individuals with similar duplications, one explanation is that the duplicated fragments are in different genomic locations. Both individuals (U1 and U2) from the tall stature cohort [Upners et al., 2017] have a maximum duplication size consistent with a terminal rather than an interstitial duplication, and patient P7 (who has a height on the 50th centile) is known to have a terminal duplication, so for these individuals there is a higher likelihood that the extra copy may be translocated elsewhere in the genome [Qian et al., 2018] and therefore not affecting the expression of the other two copies of *SHOX*.

Secondly, have all other causes of the proband's phenotype been excluded? As LWD is specific to *SHOX*, the exclusion of an additional *SHOX* variant in an LWD proband is the sole requirement. However, ISS has multiple aetiologies, so ideally such individuals should have genome-wide testing such as aCGH and whole genome/exome sequencing (WGS/WES) in order to exclude other possible causes of short stature. However, these techniques do not have 100% coverage, they

are unlikely to detect methylation abnormalities or deep intronic variants, and WES/WGS has been reported to provide a molecular genetic diagnosis in only 30-50% of cases [Yang et al., 2013; Gilissen et al., 2014; Soden et al., 2014; Srivastava et al., 2014], so excluding all other causes of ISS is very difficult, even if additional testing is performed. An ideal example is patient P6 who has had aCGH and WGS testing and the only detected variant was the *SHOX* duplication. However, he also has epilepsy and behavioral and sleep difficulties, so if a genetic cause for these symptoms has not been found we cannot be sure that another cause of ISS has also been missed.

Although it is clear from Figure 1 that apparently similar duplications do not always produce a consistent phenotype, the duplication size, regulatory element content and the proximity of one or more breakpoint to the *SHOX* gene may all contribute to the phenotypic consequences of *SHOX* whole gene duplications. With a limited number of positive patients, it is difficult to draw significant conclusions from the breakpoint data. However, five of the six individuals with a duplication of just the *SHOX* gene had LWD (n=2) or ISS (n=3), so breakpoints close to *SHOX* appear more likely to produce a *SHOX* haploinsufficiency phenotype. Patients with normal stature or tall stature have the largest average duplication size (and consequently the highest number of regulatory elements). However, while these two groups have the same average number of regulatory elements, they have a different distribution: tall stature duplications have an excess of upstream regulatory elements while normal stature duplications have an excess of downstream regulatory elements. Similarly, although duplications identified in patients with ISS and club foot also have the same average number of regulatory elements there is a difference in distribution: ISS duplications contain approximately equal numbers of upstream and downstream regulatory elements, while club foot duplications predominantly contain upstream regulatory elements.

Another possible explanation for the phenotypic discrepancy between apparently similar duplications is the influence of modifier genes such as *CYP26C1* [Montalbano et al., 2016]. The pathogenic X:780,550-828,092 deletion that removes the flanking ZED *SHOX* regulatory element was shown to be inherited from a phenotypically normal parent in 43% of cases in one cohort [Bunyan et al., 2013], so high levels of phenotypic variability have previously been observed even in individuals with a known pathogenic *SHOX* variant. In Table 1, only two of the nine parents with the duplication (where the parental heights are known) have the same phenotype designation (ISS) as the proband, with the other seven falling into the 4th-96th centile range.

Alternatively, the two *in cis* copies of *SHOX* in these individuals are effectively competing for the same regulatory elements which may result in inefficient expression of both copies, ultimately leading to *SHOX* under-expression rather than over-expression. Duplications could also change the 3D structure, preventing efficient transcription. In these scenarios, the term triplosensitivity is, therefore, inappropriate and misleading. Distinguishing between phenotypic variability, ascertainment bias and the physical location of the duplicated material makes any conclusive genotype-phenotype correlation difficult.

Without significant intra-familial segregation and/or re-evaluation of the *SHOX* ClinGen "triplosensitivity" score, these duplications will inevitably be classified as variants of uncertain clinical significance. Further studies are required to establish the significance of *SHOX* whole gene duplications that do not include the entire regulatory region. However, there is currently sufficient evidence to suggest that these duplications should not be discounted without considering the extent of the duplication and the patient phenotype, and should be included on diagnostic laboratory reports. However, given the uncertain pathogenicity of these duplications, any reports should encourage the exclusion of all other causes of short stature where possible.

Statements

Acknowledgements

We are grateful to all patients and clinicians who were involved in our *SHOX* and aCGH cohorts, and to Emma-Jane Cassidy for the interpretation of aCGH data and advice on copy number variant interpretation guidelines.

Statement of Ethics

The new probands presented in this study were consented for *SHOX* gene testing or array comparative genome hybridisation (aCGH) analysis as part of their routine clinical care within the UK National Health Service. Written informed consent was obtained from a parent or legal guardian of any participants under 16 years old prior to sample collection.
This retrospective review of patient data did not require ethical approval in accordance with local/national guidelines.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The authors have no funding sources to declare.

Author Contributions

David J Bunyan: Conceptualisation, Validation, Investigation, Writing - original draft, Writing - review & editing. James I. Hobbs: Validation, Investigation. Philippa J. Duncan-Flavell: Validation, Investigation. Rachel J Howarth: Validation, Investigation. Sarah Beal: Validation, Investigation. Diana Baralle; Investigation. Resources. N. Simon Thomas: Conceptualisation, Validation, Writing - original draft, Writing - review & editing, Supervision.

Data Availability Statement

All data generated or analysed during this study are included in this article. Further enquiries can be directed to the corresponding author.

References

Benito-Sanz S, Barroso E, Heine-Suñer D, Hisado-Oliva A, Romanelli V, Rosell J, et al. Clinical and molecular evaluation of SHOX/PAR1 duplications in Leri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS). J Clin Endocrinol Metab. 2011;96(2):E404-12

Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernandez B, Barreda-Bonis AC, Liu P, et al. Identification of the first recurrent PAR1 deletion in Léri-Weill Dyschondrosteosis and idiopathic short stature reveals the presence of a novel *SHOX* enhancer. J Med Genet. 2012; 49:442-50

Benito-Sanz S, Belinchon-Martínez A, Aza-Carmona M, de la Torre C, Huber C, González-Casado I, et al. Identification of 15 novel partial SHOX deletions and 13 partial duplications, and a review of the literature reveals intron 3 to be a hotspot region. J Hum Genet. 2017;62(2):229-34

Binder G, Ranke MB, Martin DD. Auxology is a valuable instrument for the clinical diagnosis of SHOX haploinsufficiency in school-age children with unexplained short stature. J Clin Endocrinol Metab. 2003;88(10):4891-6

Brosens E, de Jong EM, Barakat TS, Eussen BH, D'haene B, De Baere E, et al. Structural and numerical changes of chromosome X in patients with esophageal atresia. Eur J Hum Genet. 2014;22:1077-84

Bunyan DJ, Baker KR, Harvey JF, Thomas NS. Diagnostic screening identifies a wide range of mutations involving the *SHOX* gene, including a common 47.5kb deletion 160kb downstream with a variable effect on gene function. Am J Med Genet Part A. 2013;161A:1329-38

Bunyan DJ, Taylor E-J, Maloney VK, Blyth M. Homozygosity for a novel deletion downstream of the *SHOX* gene provides evidence for an additional long range regulatory region with a mild phenotypic effect. Am J Med Genet Part A. 2014;164(11):2764-8

Bunyan DJ, Gevers E, Hobbs JI, Duncan-Flavell PJ, Howarth RJ, Holder-Espinasse M, et al. Rare dosage abnormalities flanking the *SHOX* gene. Egypt J Med Hum Genet. 2021;22:89

Chen J, Wildhardt G, Zhong Z, Röth R, Weiss B, Steinberger D, et al. Enhancer deletions of the SHOX gene as a frequent cause of short stature: the essential role of a 250 kb downstream regulatory domain. J Med Genet. 2009;46:834–9

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376-80

Donze SH, Meijer CR, Kant SG, Zandwijken GR, van der Hout AH, van Spaendonk RM, et al. The growth response to GH treatment is greater in patients with SHOX enhancer deletions compared to SHOX defects. Eur J Endocrinol. 2015;173:611-21

Durand C, Bangs F, Signolet J, Decker E, Tickle C, Rappold G. Enhancer elements upstream of the SHOX gene are active in the developing limb. Eur J Hum Genet. 2010;18:527-32

Fukami M, Kato F, Tajima T, Yokoya S, Ogata T. Transactivation function of an approximately 800-bp evolutionarily conserved sequence at the SHOX 3'region: implication for the downstream enhancer. Am J Hum Genet. 2006;78:167–70

Fukami M, Naiki Y, Muroya K, Hamajima T, Soneda S, Horikawa R, et al. Rare pseudoautosomal copy-number variations involving SHOX and/or its flanking regions in individuals with and without short stature. J Hum Genet. 2015;60:553-6

Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7

Hirschfeldova K, Solc R. Comparison of SHOX and associated elements duplications distribution between patients (Lèri-Weill dyschondrosteosis/idiopathic short stature) and population sample. Gene. 2017;627:164-8

Huber C, Rosilio M, Munnich A, Cormier-Daire V. High incidence of SHOX anomalies in individuals with short stature. J Med Genet. 2006;43:735-9

Jorge AA, Souza SC, Nishi MY, Billerbeck AE, Libório DC, Kim CA, et al. SHOX mutations in idiopathic short stature and Leri-Weill dyschondrosteosis: frequency and phenotypic variability. Clin Endocrinol (Oxf). 2007;66:130–5

Leri A, Weill J. Une affection congenitale et symetrique du developpement osseux: la dyschondrosteose. Bull Mem Soc Med Hop (Paris). 1929;53:1491

Marchini A, Ogata T, Rappold GA. A track record on SHOX: from basic research to complex models and therapy. Endocr Rev. 2016;37(4):417-48
Montalbano A, Juergensen L, Roeth R, Weiss B, Fukami M, Fricke-Otto S, et al. Retinoic acid catabolizing enzyme CYP26C1 is a genetic modifier in SHOX deficiency. EMBO Mol Med. 2016;8(12):1455-69
Ottesen AM, Aksglaede L, Garn I, Tartaglia N, Tassone F, Gravholt CH, et al. Increased number of sex chromosomes affects height in a nonlinear fashion: a study of 305 patients with sex chromosome aneuploidy. Am J Med Genet A. 2010;152A(5):1206-12
Pedicelli S, Peschiaroli E, Violi E, Cianfarani S. Controversies in the definition and treatment of idiopathic short stature (ISS). J Clin Res Pediatr Endocrinol. 2009;1(3):105-15
Qian YQ, Fu XY, Wang XQ, Luo YQ, Chen M, Yan K, et al. A feasible diagnostic approach for the translocation carrier from the indication of products of conception. Mol Cytogenet. 2018;11:12
Rappold G, Blum WF, Shavrikova EP, Crowe BJ, Roeth R, Quigley CA, et al. Genotypes and phenotypes in children with short stature: clinical indicators of SHOX haploinsufficiency. J Med Genet. 2007;44:306-13
Richardson M, Chong H, Mu W, Conner BR, Hsuan V, Willett S, et al. DNA breakpoint assay reveals a majority of gross duplications occur in tandem reducing VUS classifications in breast cancer predisposition genes. Genet Med. 2018;21:683-93
Rooney Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22(2):245-57
Sabherwal N, Bangs F, Röth R, Weiss B, Jantz K, Tiecke E, et al. Long-range conserved non-coding *SHOX* sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. Hum Mol Genet. 2007;16:210-22
Sadler B, Haller G, Antunes L, Nikolov, M, Amarillo, I, Coe, B, et al. Rare and *de novo* duplications containing *SHOX* in clubfoot. J Med Genet. 2020;57:851-7
Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nuc Acid Res. 2002;30:e57
Skuplik I, Benito-Sanz S, Rosin JM, Bobick BE, Heath KE, Cobb J. Identification of a limb enhancer that is removed by pathogenic deletions downstream of the SHOX gene. Sci Rep. 2018;8:14292
Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Sci Transl Med. 2014;6(265):265ra168
Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. Ann Neurol. 2014;76(4):473-83
Stuppia L, Calabrese G, Gatta V, Pintor S, E Morizio, D Fantasia, et al. SHOX mutations detected by FISH and direct sequencing in patients with short stature. J Med Genet. 2003;40(2):e11
Temple IK, Thomas TG, Baraitser M. Congenital spinal deformity in a three generation family. J Med Genet. 1988;25:831-4
Thomas NS, Harvey JF, Bunyan DJ, Rankin J, Grigelioniene G, Bruno DL, et al. Clinical and molecular characterisation of duplications encompassing the human *SHOX* gene reveal a variable effect on stature. Am J Med Genet A. 2009;149A:1407-14
Tropeano M, Howley D, Gazzellone MJ, Wilson CE, Ahn JW, Stavropoulos DJ, et al. Microduplications at the pseudoautosomal *SHOX* locus in autism spectrum disorders and related neurodevelopmental conditions. J Med Genet 2016;53:536-47
Upners EN, Jensen RB, Rajpert-De Meyts E, Dunø M, Aksglaede L, Juul A. Short stature homeobox-containing gene duplications in 3.7% of girls with tall stature and normal karyotypes. Acta Paediatr. 2017;106(10):1651-7
Wit JM, Clayton PE, Rogol AD, Savage MO, Saenger PH, Cohen P. Idiopathic short stature: definition, epidemiology, and diagnostic evaluation. Growth Horm IGF Res. 2008;18(2):89-110

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502-11

Figure Legends

Fig. 1. The minimum duplication sizes of the nine probands from this study and the 29 probands from the literature where phenotypic data were available.

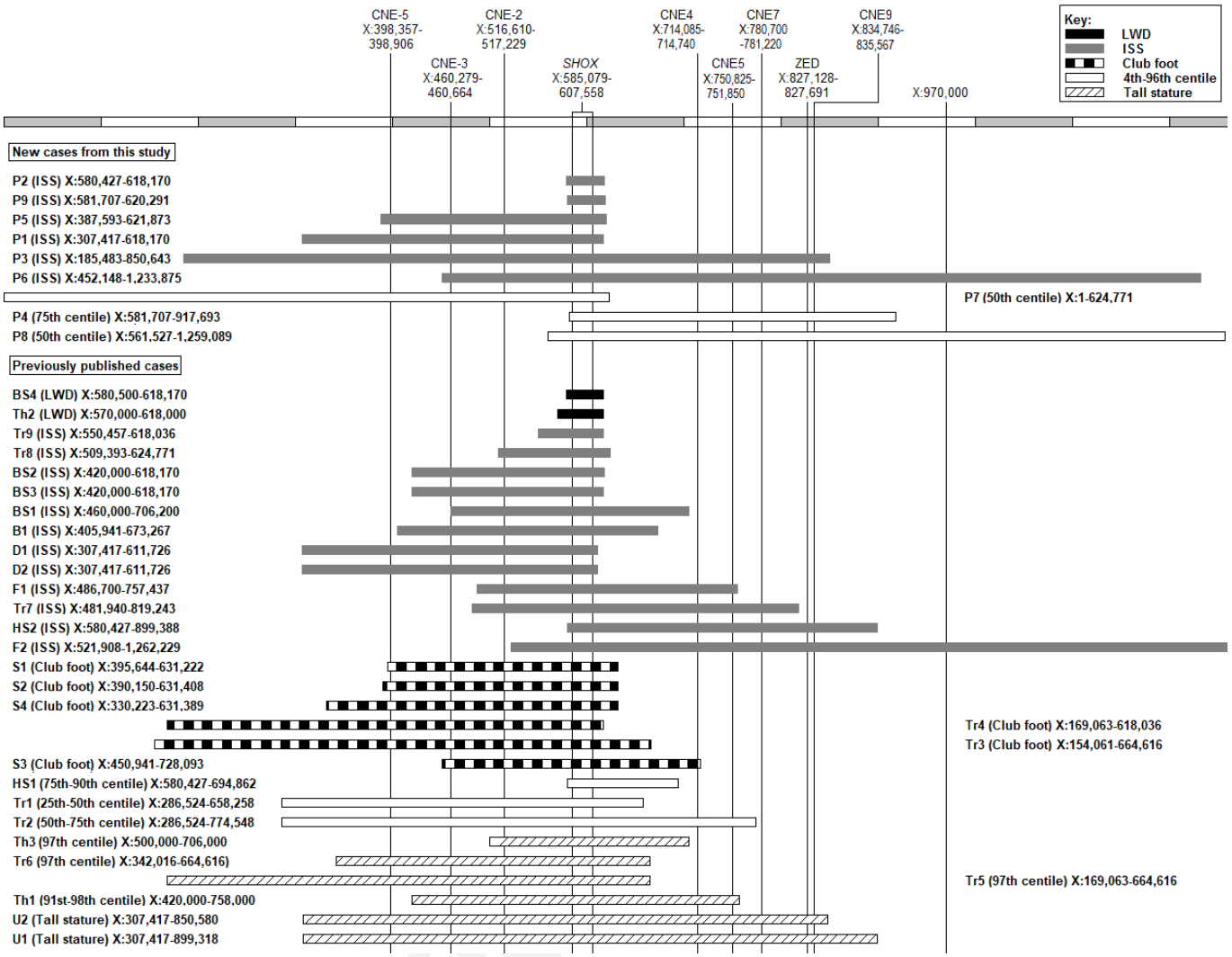
CNE-5
X:398,357-
398,906
CNE-2
X:516,610-
517,229
CNE4
X:714,085-
714,740
CNE7
X:780,700
-781,220
CNE9
X:834,746-
835,567
CNE-3
X:460,279-
460,664
SHOX
X:585,079-
607,558
CNE5
X:750,825-
751,850
ZED
X:827,128-
827,691
X:970,000
Key:
LWD
ISS
Club foot
4th-96th centile
Tall stature
New cases from this study
P2 (ISS) X:580,427-618,170
P9 (ISS) X:581,707-620,291
P5 (ISS) X:387,593-621,873
P1 (ISS) X:307,417-618,170
P3 (ISS) X:185,483-850,643
P6 (ISS) X:452,148-1,233,875
P7 (50th centile) X:1-624,771
P4 (75th centile) X:581,707-917,693
P8 (50th centile) X:561,527-1,259,089
Previously published cases
BS4 (LWD) X:580,500-618,170
Th2 (LWD) X:570,000-618,000
Tr9 (ISS) X:550,457-618,036
Tr8 (ISS) X:509,393-624,771
BS2 (ISS) X:420,000-618,170
BS3 (ISS) X:420,000-618,170
BS1 (ISS) X:460,000-706,200
B1 (ISS) X:405,941-673,267
D1 (ISS) X:307,417-611,726
D2 (ISS) X:307,417-611,726
F1 (ISS) X:486,700-757,437
Tr7 (ISS) X:481,940-819,243
HS2 (ISS) X:580,427-899,388
F2 (ISS) X:521,908-1,262,229
S1 (Club foot) X:395,644-631,222
S2 (Club foot) X:390,150-631,408
S4 (Club foot) X:330,223-631,389
Tr4 (Club foot) X:169,063-618,036
Tr3 (Club foot) X:154,061-664,616
S3 (Club foot) X:450,941-728,093
HS1 (75th-90th centile) X:580,427-694,862
Tr1 (25th-50th centile) X:286,524-658,258
Tr2 (50th-75th centile) X:286,524-774,548
Th3 (97th centile) X:500,000-706,000
Tr6 (97th centile) X:342,016-664,616)
Tr5 (97th centile) X:169,063-664,616
Th1 (91st-98th centile) X:420,000-758,000
U2 (Tall stature) X:307,417-850,580
U1 (Tall stature) X:307,417-899,318

Table 1. The ascertainment and phenotypes of the individuals with *SHOX* whole gene duplications (duplicating only part of the *SHOX* regulatory region) and the minimum duplication sizes. Patients P1-P9 are the novel cases from this study while patients Th1-Th3, BS1-BS4, B1, F1-F2, D1-D2, Tr1-Tr9, HS1-HS2, U1-U2 and S1-S4 are taken from the respective published manuscripts [Thomas et al., 2009; Benito-Sanz et al., 2011; Brosens et al., 2014; Fukami et al., 2015; Donze et al., 2015; Tropeano et al., 2016; Hirschfeldova and Solc 2017; Upners et al., 2017; Sadler et al., 2020]. Parental heights are shown where available.

Patient number	Sex	Age	Test methodology / Clinical details / Other testing	Inheritance	Phenotype designation	Duplication (chrX)
P1	F	5	*SHOX* cohort testing. Height on the 0.4th centile, no skeletal abnormalities. Additional testing unknown.	Unknown	ISS	307,417-618,170
P2	M	16	*SHOX* cohort testing. Height on the 0.4th centile, brachydactyly. No dysmorphism or Madelung deformity. Additional testing unknown.	Unknown	ISS	580,427-618,170
P3	M	10	aCGH testing for height <3rd centile, craniosynostosis, moderate developmental delay, large head, learning difficulties. Also tested for Noonan syndrome and Russell-Silver syndrome but no other abnormality was detected. FISH analysis of the duplication gave a single signal at Xp22.3.	Paternal	ISS	185,483-850,643
P4	M	5	aCGH testing for neonatal oedema, over-riding toes, deep palmar creases. Height on the 75th centile. Additional testing unknown.	Unknown	4th-96th centile	581,707-917,693
P5	F	0.1	aCGH testing for faltering growth, height <0.4th centile with relative preservation of head circumference, almond-shaped eyes, mild bilateral hip immaturity. Not dysmorphic but has a small, blind-ending sacral dimple. Also tested for Russell-Silver syndrome, Spinal Muscular Atrophy, Congenital Generalised Lipodystrophy gene panel, Multi-Locus Imprinting Disorders and Pseudohypoparathyroidism but no other abnormality detected.	Unknown	ISS	387,593-621,873
P6	M	4	aCGH and epilepsy gene panel testing for pharmaco-resistant epilepsy with previous generalized tonic-clonic seizures and frequent absence seizures with myoclonus, learning difficulties, Attention Deficit Hyperactive Disorder, behavioural difficulties, sleep difficulties. Height on the 3rd centile. Also tested by whole genome sequencing but no other abnormality detected.	Maternal	ISS	452,148-1,233,875
P7	M	3	aCGH testing for autism and developmental delay except gross motor skills, height on the 50th centile. Additional testing unknown.	Maternal	4th-96th centile	1-624,771
P8	F	2	aCGH testing for heart defects and an imperforate anus. Height on the 50th centile. Additional testing unknown.	Paternal	4th-96th centile	561,527-1,259,089
P9	F	1	Originally referred for Achondroplasia/Hypochondroplasia testing because of rhizomelia, a short neck and height <0.4th centile. aCGH subsequently requested following a normal *FGFR3* result.	Maternal (height on the 17th centile)	ISS	581,707-620,291
Th1	M	8	aCGH testing for Asperger syndrome. Height on the 91st-98th centile.	Maternal (height on the 75th-91st centile)	Tall stature	420,000-758,000
Th2	F	68	*SHOX* cohort testing. Height on the 0.4th centile, Madelung deformity.	Unknown	LWD	570,000-618,000
Th3	F	1	aCGH testing for cleft palate. Height on the 97th centile.	Paternal (height on the 75th-90th centile)	Tall stature	500,000-706,000
BS1	F	13	*SHOX* cohort testing. Height <3rd centile, slightly shortened neck.	Maternal (height <3rd centile)	ISS	460,000-706,200

BS2	F	>18	*SHOX* cohort testing. Height <3rd centile.	Maternal (height <3rd centile)	ISS	420,000-618,170
BS3	F	13.8	*SHOX* cohort testing. Height <3rd centile.	Paternal (height on the 75th centile)	ISS	420,000-618,170
BS4	F	>18	*SHOX* cohort testing. Height on the 25th-50th centile, Madelung deformity, short ulnars.	Unknown	LWD	580,500-618,170
B1	M	Not given	aCGH testing for esophageal atresia. Height on the 2nd centile, limb anomalies. FISH analysis of the duplication gave a single signal at Xp22.3.	Maternal	ISS	405,941-673,267
F1	F	1.9	*SHOX* cohort testing. Height <3rd centile. Sequencing showed that the duplication is a direct tandem repeat.	Unknown	ISS	486,700-757,437
F2	M	3.2	*SHOX* cohort testing. Height <3rd centile. FISH analysis of the duplication gave a single signal at Xp22.3.	Unknown	ISS	521,908-1,262,229
D1	F	5.6	*SHOX* cohort testing. Height <3rd centile.	Unknown	ISS	307,417-611,726
D2	F	3.3	*SHOX* cohort testing. Height <0.4th centile.	Paternal (height on the 75th centile)	ISS	307,417-611,726
Tr1	F	>40	aCGH testing for Asperger syndrome. Height on the 25th-50th centile.	Unknown	4th-96th centile	286,524-658,258
Tr2	F	>50	aCGH testing for Asperger syndrome. Height on 50th-75th centile.	Unknown	4th-96th centile	286,524-774,548
Tr3	F	<2	aCGH testing as part of a non-ASD control cohort. Also has talipes equinovarus.	Paternal	Club foot	154,061-664,616
Tr4	M	<2	aCGH testing for ASD. Also has bilateral talipes equinovarus.	Unknown	Club foot	169,063-618,036
Tr5	F	<2	aCGH testing as part of a non-ASD control cohort. Height > 97th centile.	Maternal	Tall stature	169,063-664,616
Tr6	M	10	aCGH testing as part of a non-ASD control cohort. Height > 97th centile.	Unknown	Tall stature	342,016-664,616
Tr7	F	<2	aCGH testing as part of a non-ASD control cohort. Height <3rd centile.	Paternal	ISS	481,940-819,243
Tr8	M	54	aCGH testing for ASD. Height <3rd centile.	Maternal	ISS	509,393-624,771
Tr9	M	4	aCGH testing for ASD. Height <3rd centile.	Unknown	ISS	550,457-618,036
HS1	F	Not given	*SHOX* MLPA population control. Height on the 75th-90th centile.	Unknown	4th-96th centile	580,427-694,862
HS2	M	Not given	*SHOX* cohort testing. Height <3rd centile.	Unknown	ISS	580,427-899,388
U1	F	13	*SHOX* MLPA for tall stature. Height >99th centile. Normal karyotype.	Paternal (height on the 75th-90th centile)	Tall stature	307,417-899,318
U2	F	10	*SHOX* MLPA for tall stature. Height >99th centile. Normal karyotype.	Paternal (height on the 75th-90th centile)	Tall stature	307,417-850,580
S1	M	2	*SHOX* MLPA for club foot. Bilateral talipes equinovarus, amniotic band syndrome, bilateral symbrachydactly, sixth nerve palsy, height on the 8th centile.	Inherited	Club foot	395,644-631,222
S2	M	14	*SHOX* MLPA for club foot. Bilateral talipes equinovarus, developmental delay, height unknown. Also has a 16p13.11 duplication.	*De novo*	Club foot	390,150-631,408
S3	M	14	*SHOX* MLPA for club foot. Bilateral talipes equinovarus, height unknown.	Unknown	Club foot	450,941-728,093

S4	F	13	*SHOX* MLPA for club foot. Left talipes equinovarus, adolescent idiopathic scoliosis, height on the 48th centile.	Unknown	Club foot	330,223-631,389

Table 2. *SHOX* whole gene duplication detection rates in this study and in cohorts from the literature.

	Detection rate in the different cohort types						
Publication	LWD/ISS	Club foot	Esophageal atresia	Tall stature	ASD	Unselected aCGH cohorts	Controls
This study plus Thomas et al., 2009 and Bunyan et al., 2013	4/1,959 (0.20%)					9/22,018 (0.04%)	0/471 (0%)
Benito-Sanz et al., 2011	4/735 (0.54%)			0/104 (0%)			0/340 (0%)
Brosens et al., 2014			1/180 (0.56%)				
Fukami et al., 2015	2/245 (0.82%)						0/15 (0%)
Tropeano et al., 2016					48/26,664 (0.18%)	7/12,594 (0.06%)	
Hirschfeldova and Solc 2017	1/352 (0.28%)						1/250 (0.4%)
Upners et al., 2017				2/81 (2.47%)			
Sadler et al., 2020		4/816 (0.49%)					0/2645 (0%)
Totals	11/3,291 (0.33%)	4/816 (0.49%)	1/180 (0.56%)	2/185 (1.08%)	48/26,664 (0.18%)	16/34,612 (0.05%)	1/3721 (0.03%)