



OPEN ACCESS

Short report

# Conclusion of diagnostic odysseys due to inversions disrupting *GLI3* and *FBN1*

Alistair T Pagnamenta <sup>1,2</sup>, Jing Yu,<sup>1,3</sup> Julie Evans,<sup>4</sup> Philip Twiss,<sup>5</sup> Genomics England Research Consortium, Musculoskeletal GeCIP MDT, Amaka C Offiah <sup>6</sup>, Mohamed Wafik,<sup>7</sup> Sarju G Mehta,<sup>8</sup> Mohammed K Javaid,<sup>9</sup> Sarah F Smithson,<sup>10</sup> Jenny C Taylor <sup>1,2</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmg-2022-108753>).

For numbered affiliations see end of article.

## Correspondence to

Dr Alistair T Pagnamenta, Wellcome Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, UK; [alistair@well.ox.ac.uk](mailto:alistair@well.ox.ac.uk)

Received 21 July 2022

Accepted 14 October 2022

## ABSTRACT

Many genetic testing methodologies are biased towards picking up structural variants (SVs) that alter copy number. Copy-neutral rearrangements such as inversions are therefore likely to suffer from underascertainment. In this study, manual review prompted by a virtual multidisciplinary team meeting and subsequent bioinformatic prioritisation of data from the 100K Genomes Project was performed across 43 genes linked to well-characterised skeletal disorders. Ten individuals from three independent families were found to harbour diagnostic inversions. In two families, inverted segments of 1.2/14.8 Mb unequivocally disrupted *GLI3* and segregated with skeletal features consistent with Greig cephalopolysyndactyly syndrome. For one family, phenotypic blending was due to the opposing breakpoint lying ~45 kb from *HOXA13*. In the third family, long suspected to have Marfan syndrome, a 2.0 Mb inversion disrupting *FBN1* was identified. These findings resolved lengthy diagnostic odysseys of 9–20 years and highlight the importance of direct interaction between clinicians and data-analysts. These exemplars of a rare mutational class inform future SV prioritisation strategies within the NHS Genomic Medicine Service and similar genome sequencing initiatives. In over 30 years since these two disease-gene associations were identified, large inversions have yet to be described and so our results extend the mutational spectra linked to these conditions.

## INTRODUCTION

The rare-disease pilot phase of the 100K Genomes Project (100KGP) involved 2183 families spread across 20 different diagnostic categories.<sup>1</sup> Building on previous studies,<sup>2</sup> this has been a major step towards embedding whole-genome sequencing (WGS) into standard healthcare, providing valuable lessons which are being applied in the UK National Health Service (NHS) Genomic Medicine Service. One notable finding was the significant uplift in diagnostic yield made with the help of researchers, which increased the overall yield to 25%. These researcher-enabled findings included 22 non-coding variants, many of which were confirmed experimentally by splicing/luciferase studies, and several repeat expansions.

Many individuals recruited to 100KGP had previously been pre-screened by microarrays, PCR-Sanger, multiplex ligation-dependent probe

amplification, exome sequencing or panel-NGS. These types of genetic analysis are typically inefficient at picking up inversions. Although traditional karyotyping can identify inversions, in most cases this is limited to events of >10 Mb and such methods are nowadays employed infrequently as a first-line test.<sup>3</sup> While the latest optical mapping methods demonstrate a high concordance with traditional approaches<sup>4</sup> and have the potential to be used as first-line test for detecting cryptic SVs,<sup>5</sup> these methods are not yet performed routinely in clinical laboratories. Therefore, one might anticipate 100KGP to be enriched for cryptic structural variants (SVs). Given that the raison d'être of WGS is to pick up all forms of variation, the absence of diagnostic inversions or other complex copy-neutral rearrangements in the 100KGP pilot is notable. Of the 40 variants classed as SVs, all were simple deletions/duplications.<sup>1</sup>

This study was prompted by an unanticipated finding resulting from a virtual multidisciplinary (MDT) meeting involving clinical and academic centres in the UK set up to review genetic/clinical data for unsolved musculoskeletal cases from the 100KGP. These meetings aimed to integrate phenotypic information with dREAMS radiological characterisation<sup>6</sup> and combine with manual review of genomic data. To follow-up our initial findings, which included a family with an inversion disrupting *GLI3*, bioinformatic SV prioritisation tools were developed to search systematically for gene-disrupting inversions across 43 genes that have been linked to well-characterised autosomal dominant forms of skeletal disorders.

## METHODS

The 100KGP was initiated in 2013 to establish diagnoses for patients with rare-disease and cancer and promote the use of WGS in the NHS.<sup>7</sup> The clinical filtering pipeline designed by Genomics England to analyse data from the 100KGP uses a tiering system (online supplemental figure S1A). Variants are assigned as tier 1–3 depending on inheritance, consequence and on whether they lie in a gene assessed as Green in PanelApp (<https://panelapp.genomicsengland.co.uk>), a crowdsourcing knowledgebase containing virtual gene panels relating to a wide range of human disorders. Data from the 100KGP are held in the National Genomic Research



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Pagnamenta AT, Yu J, Evans J, et al. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2022-108753

Library (<https://doi.org/10.6084/m9.figshare.4530893.v6>) and researchers can apply to access data at [www.genomicsengland.co.uk/join-a-gecip-domain](http://www.genomicsengland.co.uk/join-a-gecip-domain). If researchers discover variants that could represent a diagnosis for a participant, they are asked to submit the variants into a review/triage pipeline (online supplemental figure S1B), helping provide assurance to the Genomic Medicine Service that the diagnoses are of high quality and clinical relevance.

In the majority of rare-disease cases, DNA was extracted from blood using the EDTA method and TruSeq PCR-free high throughput library preparation was followed by 150 bp paired-read sequencing on a HiSeqX machine (Illumina). SVs were called using a combination of CANVAS and MANTA algorithms and combined into single 'SV.vcf' files. Mean sequence coverage for the 10 individuals reported here was 35–55x and 341–519 inversions were called, consistent with the numbers seen across the 100KGP as a whole (mean 427; online supplemental table S1). Further quality control statistics are available within the Genomics England research environment.

A monthly virtual MDT meeting process was initiated to scrutinise clinical/WGS data with the aim of helping to solve unsolved musculoskeletal cases from the 100KGP. Further details describing these meetings are available in online supplemental methods. Manual review of read alignments was performed using IGV (v2.11.9), with visibility range threshold setting increased to 100 kb. The SV.vcf file was also loaded into IGV with the feature visibility window size set to 0 kb.

SVs are thought to play a significant role in dominant disease and yet are often missed by WGS analytical pipelines. We therefore sought to extend the preliminary results arising from the MDT meetings by focussing on 43 autosomal genes (online supplemental table S2) listed in the 2019 revision of the skeletal disorder nosology<sup>8</sup> which curators at the Clinical Genome Resource ([www.clinicalgenome.org](http://www.clinicalgenome.org)) assessed as having 'sufficient evidence' supporting haploinsufficiency as a disease mechanism (HI=3). Gene-oriented filtering of SVs in rare disease cases from the main-programme of the 100KGP was performed with SVRare,<sup>9</sup> as described in online supplemental methods. To validate inversions, breakpoint PCR and Sanger sequencing was performed using primers listed in online supplemental table S3.

## RESULTS

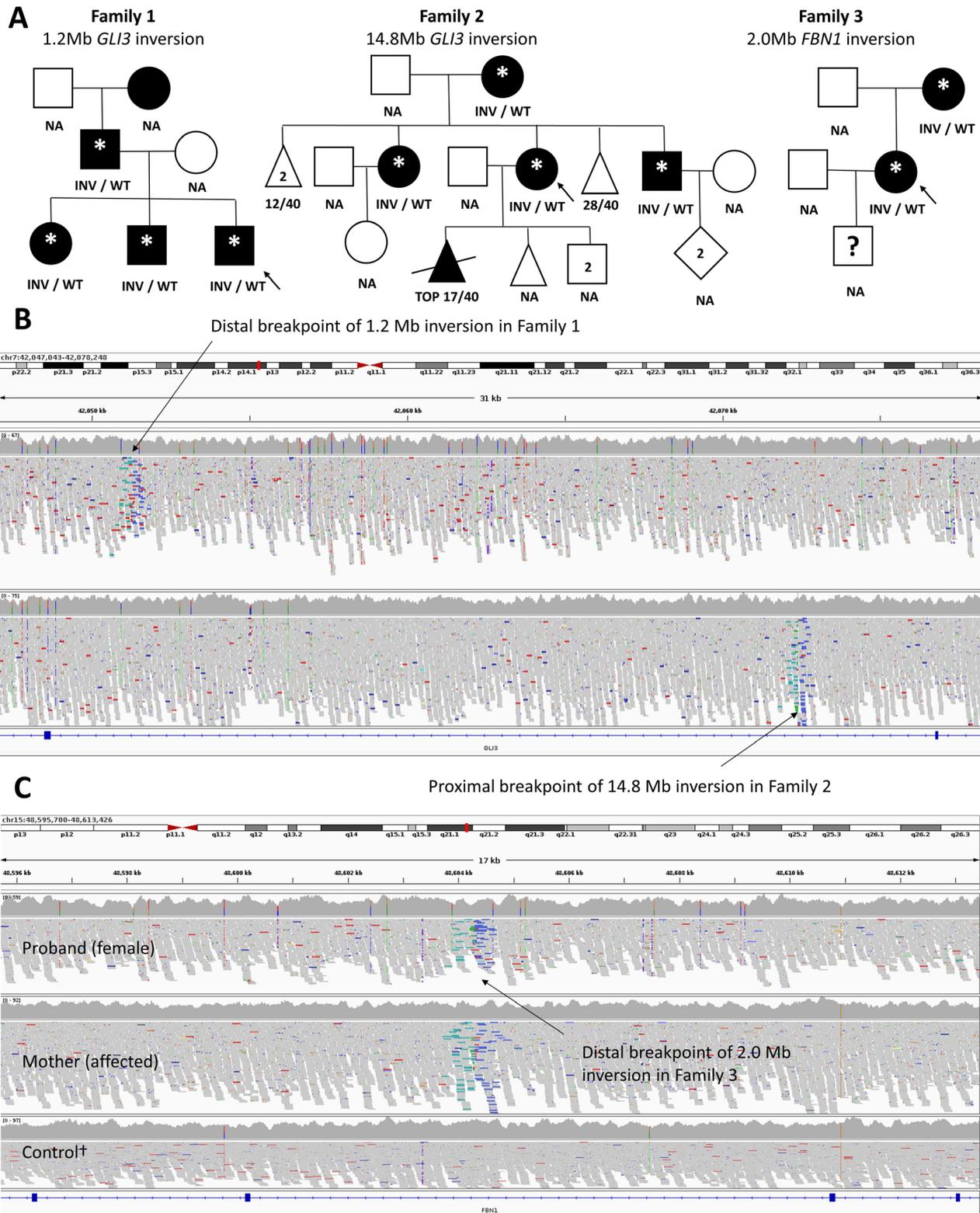
Prior to the first MDT meeting, details were circulated of a boy with clinical features consistent with Greig cephalopolysyndactyly syndrome (GCS) that included relative macrocephaly, hypertelorism, postaxial polysyndactyly of hands and preaxial polysyndactyly of feet (Family 1; online supplemental figure S2). Similarly affected family members included two older siblings, the father and the paternal grandmother (figure 1A). Targeted *GLI3* sequencing in 2004 and again in 2015 had been negative (online supplemental table S4). Due to a confident clinical diagnosis of GCS syndrome, manual inspection of read alignments was performed and in 4/4 affected family members clustering of split read-pairs was identified in intron 4 (figure 1B). Relative strand orientations were consistent with the presence of a 1.2 Mb inversion. This inversion had been called by Manta as chr7:42 051 297–43 254 780 (GRCh38). While the distal breakpoint disrupts *GLI3*, the proximal breakpoint lies within *HECW1*, another gene predicted to be constrained against loss of function variants (pLI=1, gnomAD 2.1.1) but not yet associated with any Mendelian disease. Breakpoints called by Manta were consistent with those seen in the Sanger validation data (online supplemental figure S3), confirming a small ~25 bp

deletion at one end (online supplemental figure S4). The genuine 1.2 Mb *GLI3* inversion lay within a larger 11.6 Mb inversion call. Manual scrutiny of read alignments suggested the latter to be an artefact and increased confidence for genuine inversions may be achieved by the fact that breakpoints are detected separately and represented twice in the SV.vcf file in a reciprocal manner (online supplemental figure S5, table S4).

As inversions are an under-reported class of SV, we sought to replicate this finding using SVRare<sup>9</sup> across 71 408 rare-disease participants from 100KGP. This cohort corresponds to 33 924 families, of which 5222 were recruited under the musculoskeletal domain. Here, we focused on 43 genes linked to skeletal disorders where haploinsufficiency is a known mechanism.<sup>8</sup> Although Manta typically calls ~400 inversions per genome, prioritisation is simpler than for deletions/duplications because only genes overlapping breakpoints are unequivocally disrupted. More detailed information of the filtering/interpretation process is provided in online supplemental figures S6 and S7.

Our systematic prioritisation uncovered Family 2, where a 14.8 Mb inversion (chr7:27 245 456–42 072 394) disrupting *GLI3* was identified in 4/4 affected family members (figure 1A and B). The proband was first reviewed in the genetics clinic in her early 30s, following a termination of pregnancy due to multiple congenital abnormalities. She presented with an unusual combination of distal limb and genitourinary tract malformations. The patient was noted to have a bicornuate uterus with solitary vagina and cervix, a unilateral duplex kidney, bilateral broad and proximally placed thumbs (online supplemental figure S8A), bilateral medial displacement of the great toe ('sandal gap') and bilateral 2/3 toe syndactyly (online supplemental figure S8B). Clinical details for other family members are available in the online supplemental methods. Although hand-foot-genital syndrome (MIM #140000) had been suspected, targeted *HOXA13* analysis and exome sequencing failed to identify any pathogenic variants. While disruption of *GLI3* at the proximal breakpoint likely contributes to the skeletal phenotype, the distal breakpoint in 7p15.2 lies ~45 kb upstream of *HOXA13* and so positional effects may underlie the more variable urogenital anomalies. Breakpoint PCR and Sanger sequencing validated the inversion and confirmed the breakpoints to be consistent with those called by Manta (online supplemental figure S9), although with a small 14 bp insertion at the proximal end and a 6 bp deletion at the distal end (online supplemental figure S10).

Lastly, a mother-daughter duo (Family 3) with Marfan syndrome suspected for ~20 years shared a 2.0 Mb inversion (chr15:46 635 052–48 604 302) disrupting *FBN1* (figure 1A,C). The daughter, first seen in the genetics clinic in her early teens, had skeletal features typical of the condition, with an increased upper segment:lower segment ratio, positive wrist and thumb signs, striae over the knees, upper legs and lower back, mild pectus excavatum and mild scoliosis. An echocardiogram showed marked aortic root dilatation. Despite previous genetic testing of *FBN1* using a variety of methods (online supplemental table S4), the family remained without a diagnosis. Additional clinical details are available in online supplemental methods. Breakpoint PCR and Sanger sequencing validated the inversion in both affected family members and confirmed the breakpoints to be consistent with those called by Manta (online supplemental figures S11 and S12). Finding the molecular cause of disease in this family will have direct clinical utility as there are several relatives for whom we may now be able to provide accurate advice about their risks. Most notably, the proband's son would be difficult to discharge without any molecular testing, as clinical



**Figure 1** Pedigrees and characteristic read-alignment signatures for rare diagnostic inversions in three Families from 100KGP. (A) Pedigrees and genetic segregation. Shading in Family 1 indicates polysyndactyly of hands/feet, relative macrocephaly and suspected Greig syndrome. Shading in Family 2 indicates radial dysplasia, toe syndactyly and variable urogenital features, as detailed in online supplemental figure S13. Shading in Family 3 indicates thoracic aortic aneurysm and suspected Marfan syndrome. Clinical status of the proband's son is unknown. \*WGS data available from 100KGP. NA, genetic testing not performed. (B) Read-alignments viewed with IGV showing inversions of chr7:42 051 297–43 254 780 (Family 1) and chr7:27 245 456–42 072 394 (Family 2). Both *GLI3*-disrupting inversions have breakpoints in intron 4, confirming that truncation of the gene at this point is a *bona fide* disease mechanism. (C) Distal breakpoint of inversion (chr15:46 635 052–48 604 302) disrupting *FBN1* shared by proband (upper track) and mother (middle). †Control (lower) is unrelated individual from 100KGP analysed using similar methods. GRCh38 read-alignments are coloured by pair orientation such that read-pairs where both reads map to the +ve genomic strand are highlighted in green. Read-pairs where both reads map to the –ve strand (blue) are seen on the other side of the breakpoint.

features of Marfan syndrome are often incomplete in childhood and it can be a very variable condition even in adulthood.

## DISCUSSION

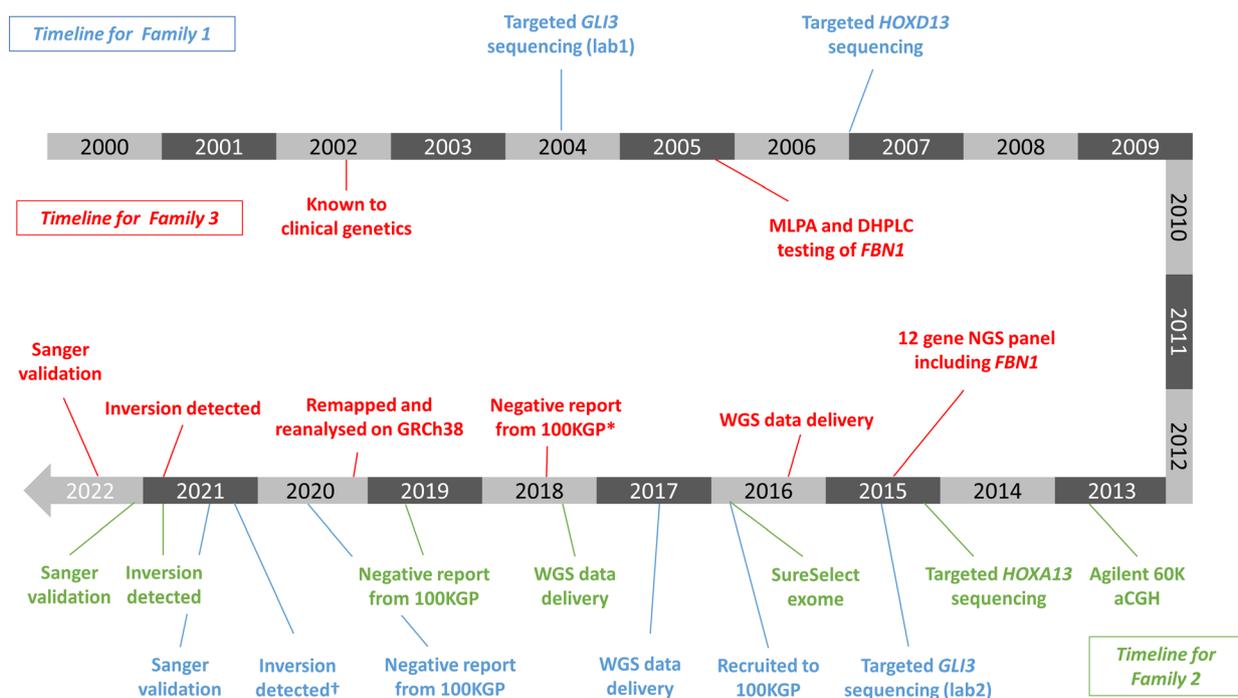
In this study, a combination of MDT discussion, manual review and systematic bioinformatics filtering helped identify rare germline inversions involving *GLI3* and *FBN1*. In family 1, the variant was found by manual assessment of a single candidate gene, prompted by an MDT meeting, highlighting the importance of having detailed phenotypic information to guide analysis. GCS is a highly recognisable condition and we recognise that for a majority of Mendelian disorders, genetic heterogeneity would make manual assessment of read alignments impractical. We therefore performed a systematic analysis of 43 genes involving 33 924 families which identified additional pathogenic inversions disrupting *GLI3* and *FBN1*, highlighting that bioinformatic prioritisation of such variants is possible. Until now, the clinical pipeline used by Genomics England has only used SV calls from Canvas, explaining why copy-neutral changes such as these have been missed. As noted in other studies,<sup>10</sup> optimisation of SV calling/prioritisation is a key area for pipeline development if the full value of clinical WGS is to be realised.

No large germline inversions have been reported for these genes previously, despite both disease-gene associations being described >30 years ago.<sup>11–12</sup> A recent study identified 48 novel cases with causative variants in *GLI3* and performed a review with 314 previously reported *GLI3* variants, looking primarily for genotype-phenotype correlations—none of the variants were inversions.<sup>13</sup> Searching HGMD identified two historical cases of GCS with translocation breakpoints in 7p13,<sup>14–16</sup> which were critical to help pinpoint this disease gene,<sup>12</sup> but no inversions. Literature searches on *FBN1* identified a CAA>TTG variant<sup>17</sup> but this could be classified as a multinucleotide substitution. This variant (NM\_000138.5:c.1881\_1883inv, p.Cys628Asn) was also present in ClinVar, alongside two other small inversions

(c.6617–9\_6617-8inv and c.1875\_1876inv; p.Gly626Arg), but these are all much smaller than the three inversions reported here (1.2–14.8 Mb) and likely result from different mutational processes. Another recent study assessed >373 paediatric patients with Marfan syndrome and did not identify any inversions, although the methods used may have made detection of such variants difficult.<sup>18</sup> Last, the Universal Mutation Database for *FBN1* (www.umd.be/FBN1) contains information about 3077 mutations, but there were no inversions reported.

Although for all three families described, the correct clinical diagnoses had been proposed previously, the precise genetic basis had remained unexplained for 9–20 years and so no specific diagnostic or predictive/prenatal test could be offered. In each case, although multiple genetic techniques were used prior to 100KGP recruitment (figure 2), most of these methods are unable to detect copy-neutral SVs such as inversions. The exception to this is karyotyping which had been performed only for Family 1. However, in that family, the inversion was 1.2 Mb in size and thus below the detection threshold. Another striking observation is that the respective diagnostic odysseys continued, even after the WGS data had been generated and the time between the sequencing data being available and reporting of the variants ranged from 3½ to 5½ years. This lag-time highlights the difficulty in picking up *bona fide* diagnostic inversions in a national clinical WGS project and the importance of understanding the limitations of the methodology employed. In such settings, a high degree of specificity is needed due to limited knowledge regarding the pathogenic importance of copy-neutral SVs.

Both GCS and Pallister-Hall syndrome (PHS) are caused by variants in *GLI3*. Genotype-phenotype correlation studies have indicated that mutations in the N-terminal and C-terminal thirds of the gene lead to GCS whereas mutations in the middle section lead to PHS.<sup>19</sup> A later study confirmed this correlation and suggested the coordinates of the central PHS specific region to be between nucleotides 1998 and 3481.<sup>20</sup> Both inversions



**Figure 2** Diagnostic odyssey timelines for Families 1–3. For Family 1, precise dates were unavailable for karyotyping and array testing. \*Sequence data initially analysed in 2016 using GRCh37 as a reference. The same data were remapped and reanalysed on GRCh38 in February 2020. †Variant identified on Rare Disease Day 2021. WGS, whole-genome sequencing.

reported here for Families 1 and 2 had breakpoints in intron four and so disrupt *GLI3* after cDNA position 474 and therefore these results are largely consistent with the previously reported genotype-phenotype correlation for GCS. However, for Family 1, while disruption to *GLI3* is likely responsible for most of the clinical features seen in this family, we cannot rule out that *HECW1* disruption could be relevant with respect to some of the atypical features. For Family 2, a degree of phenotypic blending seems highly plausible given the prior suspicion for hand-foot-genital syndrome. A study from 2016 used karyotyping/WGS to characterise a homozygous 66 Mb inversion that lies 523 kb upstream of *HOXA13*, found in a patient with hand-foot-genital syndrome.<sup>21</sup> Given that studies using mouse limb cells suggest that expression of HoxA genes can be controlled by enhancer elements located 5' of the gene cluster,<sup>22</sup> the authors suggested that the large pericentric inversion might dysregulate the spatial/temporal expression of *HOXA13*. Due to the large distance involved, any dysregulation would likely be less severe than for dominantly acting mutations that result in disease due to haploinsufficiency, hence leading to the recessive mode of inheritance.<sup>21</sup> Here, the distal breakpoint of the inversion in Family 2 lies just 45 kb from *HOXA13* and so could potentially have a more severe effect of gene regulation.

In summary, our work stresses the need to integrate multiple SV-calling algorithms and the importance of direct interaction between clinicians and data-analysts for cases where clinical suspicion points to a particular gene. Our identification of three unrelated families harbouring inversions disrupting well-known disease genes highlights examples of a rare mutational class that had not been prioritised by Genomics England's pipeline. Manual review prompted by a virtual MDT meeting and subsequent bioinformatic prioritisation of data can help to conclude lengthy diagnostic odysseys for the respective families.

#### Author affiliations

<sup>1</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

<sup>2</sup>Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK

<sup>3</sup>Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

<sup>4</sup>Bristol Genetics Laboratory, North Bristol NHS Trust, Bristol, UK

<sup>5</sup>Cambridge Genomics Laboratory, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

<sup>6</sup>Academic Unit of Child Health, The University of Sheffield, Sheffield, UK

<sup>7</sup>Department of Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, UK

<sup>8</sup>Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

<sup>9</sup>Nuffield Department of Orthopaedics, Rheumatology and Orthopaedic Sciences, University of Oxford, Oxford, UK

<sup>10</sup>Department of Clinical Genetics, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

**Acknowledgements** This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100 000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100 000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

**Collaborators** The Genomics England Research Consortium: J C Ambrose; P Arumugam; R Bevers; M Bleda; F Boardman-Pretty; C R Boustred; H Brittain; M J Caulfield; G C Chan; G Elgar; T Fowler; A Giess; A Hamblin; S Henderson; T J P Hubbard; R Jackson; L J Jones; D Kasperaviciute; M Kayikci; A Kousathanas; L Lahnstein; S E A Leigh; I U S Leong; J F Lopez; F Maleady-Crowe; M McEntagart; F Minneci; L Moutsianas; M Mueller; N Murugaesu; A C Need; P O'Donovan; C A Odhams; C Patch; M B Pereira; D Perez-Gil; J Pullinger; T Rahim; A Rendon; T Rogers; K Savage; K Sawant; R H Scott; A Siddiq; A Sieghart; S C Smith; A Sosinsky; Alexander Stuckey; M Tanguy; A L Taylor-Tavares; E R A Thomas; S R Thompson; A Tucci; M J Welland; E Williams; K Witkowska; S M Wood. Musculoskeletal GeCIP MDT: M Balasubramanian, J Bubbear, C Burren, A Calder, J Fairhurst, E Gevers, D Hunt, M

Irving, MK Javaid, Z Mohsin, A C Offiah, AT Pagnamenta, A Sabir, D Shears, S F Smithson, M Suri, JC Taylor, A Wilkie, L Wilson.

**Contributors** ATP, JY, MKJ and JT conceived study with input from the MGM. GERC, ATP, JE and PT generated experimental data. ATP, JY and JE analysed genetic/genomic data. MW, SGM and SFS recruited families and assessed clinical information. ACO reviewed radiographs. ATP wrote the manuscript with input from all coauthors.

**Funding** Supported by the MRC (MR/W01761X/1), the Wellcome Trust (203141/Z/16/Z) and the Oxford NIHR Biomedical Research Centre. The 100 000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

**Competing interests** None declared.

**Patient consent for publication** Consent obtained directly from patient(s).

**Ethics approval** This study involves human participants and was approved by Cambridge South REC (14/EE/1112). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Alistair T Pagnamenta <http://orcid.org/0000-0001-7334-0602>

Amaka C Offiah <http://orcid.org/0000-0001-8991-5036>

Jenny C Taylor <http://orcid.org/0000-0003-3602-5704>

#### REFERENCES

- Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, Ellingford JM, Arno G, Tucci A, Vandrovцова J, Chan G, Williams HJ, Ratnaik T, Wei W, Stirrups K, Ibanez K, Moutsianas L, Wielscher M, Need A, Barnes MR, Vestito L, Buchanan J, Wordsworth S, Ashford S, Rehmström K, Li E, Fuller G, Twiss P, Spasic-Boskovic O, Halsall S, Floto RA, Poole K, Wagner A, Mehta SG, Gurnell M, Burrows N, James R, Penkett C, Dewhurst E, Graf S, Mapeta R, Kasanik M, Haworth A, Savage H, Babcock M, Reese MG, Bale M, Baple E, Boustred C, Brittain H, de Burca A, Bleda M, Devereau A, Halai D, Haraldsdottir E, Hyder Z, Kasperaviciute D, Patch C, Polychronopoulos D, Matchan A, Sultana R, Ryten M, Tavares ALT, Tregidgo C, Turnbull C, Welland M, Wood S, Snow C, Williams E, Leigh S, Foulger RE, Daugherty LC, Niblock O, Leong IUS, Wright CF, Davies J, Crichton C, Welch J, Woods K, Abulhoul L, Aurora P, Bockenbauer D, Broomfield A, Cleary MA, Lam T, Dattani M, Footitt E, Ganesan V, Grunewald S, Compeyrot-Lacassagne S, Muntoni F, Pilkington C, Quinlivan R, Thapar N, Wallis C, Wedderburn LR, Worth A, Bueser T, Compton C, Deshpande C, Fassih H, Haque E, Izatt L, Josifova D, Mohammed S, Robert L, Rose S, Ruddy D, Sarkany R, Say G, Shaw AC, Wolejko A, Habib B, Burns G, Hunter S, Grocock RJ, Humphray SJ, Robinson PN, Haendel M, Simpson MA, Banka S, Clayton-Smith J, Douzou S, Hall G, Thomas HB, O'Keefe RT, Michaelides M, Moore AT, Malka S, Pontikos N, Browning AC, Straub V, Gorman GS, Horvath R, Quinton R, Schaefer AM, Yu-Wai-Man P, Turnbull DM, McFarland R, Taylor RW, O'Connor E, Yip J, Newland K, Morris HR, Polke J, Wood NW, Campbell C, Camps C, Gibson K, Koelling N, Lester T, Németh AH, Palles C, Patel S, Roy NBA, Sen A, Taylor J, Cacheiro P, Jacobsen JO, Seaby EG, Davison V, Chitty L, Douglas A, Naresh K, McMullan D, Ellard S, Temple IK, Mumford AD, Wilson G, Beales P, Bitner-Grindzicz M, Black G, Bradley JR, Brennan P, Burn J, Chinnery PF, Elliott P, Flinter F, Houlden H, Irving M, Newman W, Rahman S, Sayer JA, Taylor JC, Webster AR, Wilkie AOM, Ouwehand WH, Raymond FL, Chisholm J, Hill S, Bentley D, Scott RH, Fowler T, Rendon A, Caulfield M, 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* 2021;385:1868-80
- Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyt O, Fenwick AL, Flint J, Fugger

- L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lamble S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SR, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AO, Bentley D, Donnelly P, McVean G. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* 2015;47:717–26.
- 3 Ahn JW, Mann K, Walsh S, Shehab M, Hoang S, Docherty Z, Mohammed S, Mackie Ogilvie C. Validation and implementation of array comparative genomic hybridisation as a first line test in place of postnatal karyotyping for genome imbalance. *Mol Cytogenet* 2010;3:9.
- 4 Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, Baatout I, van Beek R, Yammine T, Oorsprong M, Hsoumi F, Olde-Weghuis D, Majdali W, Vermeulen S, Pauper M, Lebbbar A, Stevens-Kroef M, Sanlaville D, Dupont JM, Smeets D, Hoischen A, Schluth-Bolard C, El Khattabi L. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet* 2021;108:1409–22.
- 5 Zhang S, Pei Z, Lei C, Zhu S, Deng K, Zhou J, Yang J, Lu D, Sun X, Xu C, Xu C. Detection of cryptic balanced chromosomal rearrangements using high-resolution optical genome mapping. *J Med Genet* 2022. doi:10.1136/jmedgenet-2022-108553. [Epub ahead of print: 16 Jun 2022] (published Online First: 2022/06/17).
- 6 Offiah AC, Hall CM. The radiologic diagnosis of skeletal dysplasias: past, present and future. *Pediatr Radiol* 2020;50:1650–7.
- 7 Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, Halai D, Baple E, Craig C, Hamblin A, Henderson S, Patch C, O'Neill A, Devereau A, Smith K, Martin AR, Sosinsky A, McDonagh EM, Sultana R, Mueller M, Smedley D, Toms A, Dinh L, Fowler T, Bale M, Hubbard T, Rendon A, Hill S, Caulfield MJ, 100 000 Genomes Project. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* 2018;361.
- 8 Mortier GR, Cohn DH, Cormier-Daire V, Hall C, Krakow D, Mundlos S, Nishimura G, Robertson S, Sangiorgi L, Savarirayan R, Sillence D, Superti-Furga A, Unger S, Warman ML. Nosology and classification of genetic skeletal disorders: 2019 revision. *Am J Med Genet A* 2019;179:2393–419.
- 9 Yu J, Szabo A, Pagnamenta AT, Shalaby A, Giacopuzzi E, Taylor J. SVRare: discovering disease-causing structural variants in the 100K genomes project. *medRxiv* 2022.
- 10 Hyder Z, Calpena E, Pei Y, Tooze RS, Brittain H, Twigg SRF, Cilliers D, Morton JEV, McCann E, Weber A, Wilson LC, Douglas AGL, McGowan R, Need A, Bond A, Tavares ALT, Thomas ERA, Hill SL, Deans ZC, Boardman-Pretty F, Caulfield M, Scott RH, Wilkie AOM, Genomics England Research Consortium. Evaluating the performance of a clinical genome sequencing program for diagnosis of rare genetic disease, seen through the lens of craniosynostosis. *Genet Med* 2021;23:2360–8.
- 11 Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 1991;352:337–9.
- 12 Vortkamp A, Gessler M, Grzeschik KH. Gli3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* 1991;352:539–40.
- 13 Szczakiel HL, Hülsemann W, Holtgrewe M, Abad-Perez AT, Elsner J, Schwartzmann S, Horn D, Spielmann M, Mundlos S, Mensah MA. Gli3 variants causing isolated polysyndactyly are not restricted to the protein's C-terminal third. *Clin Genet* 2021;100:758–65.
- 14 Tommerup N, Nielsen F. A familial reciprocal translocation t(3;7) (p21.1;p13) associated with the Greig polysyndactyly-craniofacial anomalies syndrome. *Am J Med Genet* 1983;16:313–21.
- 15 Krüger G, Götz J, Kvist U, Dunker H, Erfurth F, Pelz L, Zech L. Greig syndrome in a large kindred due to reciprocal chromosome translocation t(6;7)(q27;p13). *Am J Med Genet* 1989;32:411–6.
- 16 Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The human gene mutation database: towards a comprehensive Repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017;136:665–77.
- 17 Mannucci L, Luciano S, Salehi LB, Gigante L, Conte C, Longo G, Ferradini V, Piumelli N, Brancati F, Ruvolo G, Novelli G, Sangiulio F. Mutation analysis of the FBN1 gene in a cohort of patients with Marfan syndrome: a 10-year single center experience. *Clin Chim Acta* 2020;501:154–64.
- 18 Meester JAN, Peeters S, Van Den Heuvel L, Vandeweyer G, Fransen E, Cappella E, Dietz HC, Forbus G, Gelb BD, Goldmuntz E, Hoskoppal A, Landstrom AP, Lee T, Mital S, Morris S, Olson AK, Renard M, Roden DM, Singh MN, Selamet Tierney ES, Tretter JT, Van Driest SL, Willing M, Verstraeten A, Van Laer L, Lacro RV, Loeys BL. Molecular characterization and investigation of the role of genetic variation in phenotypic variability and response to treatment in a large pediatric Marfan syndrome cohort. *Genet Med* 2022;24:1045–53.
- 19 Johnston JJ, Olivos-Glander I, Killoran C, Elson E, Turner JT, Peters KF, Abbott MH, Aughton DJ, Aylsworth AS, Bamshad MJ, Booth C, Curry CJ, David A, Dinulos MB, Flannery DB, Fox MA, Graham JM, Grange DK, Guttmacher AE, Hannibal MC, Henn W, Hennekam RCM, Holmes LB, Hoyme HE, Leppig KA, Lin AE, Macleod P, Manchester DK, Marcellis C, Mazzanti L, McCann E, McDonald MT, Mendelsohn NJ, Moeschler JB, Moghaddam B, Neri G, Newbury-Ecob R, Pagon RA, Phillips JA, Sadler LS, Stoler JM, Tilstra D, Walsh Vockley CM, Zackai EH, Zadeh TM, Brueton L, Black GCM, Biesecker LG. Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of Gli3 mutations. *Am J Hum Genet* 2005;76:609–22.
- 20 Johnston JJ, Sapp JC, Turner JT, Amor D, Aftimos S, Aleck KA, Bocian M, Bodurtha JN, Cox GF, Curry CJ, Day R, Donnai D, Field M, Fujiwara I, Gabbett M, Gal M, Graham JM, Hedera P, Hennekam RCM, Hersh JH, Hopkin RJ, Kayserili H, Kidd AMJ, Kimonis V, Lin AE, Lynch SA, Maisenbacher M, Mansour S, McGaughran J, Mehta L, Murphy H, Raygada M, Robin NH, Rope AF, Rosenbaum KN, Schaefer GB, Shealy A, Smith W, Soller M, Sommer A, Stalker HJ, Steiner B, Stephan MJ, Tilstra D, Tomkins S, Trapane P, Tsai AC-H, Van Allen MI, Vasudevan PC, Zabel B, Zunich J, Black GCM, Biesecker LG. Molecular analysis expands the spectrum of phenotypes associated with Gli3 mutations. *Hum Mutat* 2010;31:1142–54.
- 21 Watson CM, Crinnion LA, Harrison SM, Lascelles C, Antanaviciute A, Carr IM, Bonthron DT, Sheridan E. A chromosome 7 pericentric inversion defined at single-nucleotide resolution using diagnostic whole genome sequencing in a patient with Hand-Foot-Genital syndrome. *PLoS One* 2016;11:e0157075.
- 22 Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. Clustering of tissue-specific sub-TADs accompanies the regulation of HOXA genes in developing limbs. *PLoS Genet* 2013;9:e1004018.

## SHORT REPORT

# Supplemental material: Conclusion of diagnostic odysseys due to inversions disrupting *GLI3* and *FBN1*

## Supplemental methods

### 1. Musculoskeletal MDT meetings

In March 2021 we initiated a virtual multidisciplinary team meeting (MDT) process with the aim of reviewing unsolved musculoskeletal families from the 100KGP. This MDT process is ongoing, with support from an MRC grant. Clinicians are asked to circulate background information prior to these meetings, including any suspected clinical diagnoses and sets of genes/pathways potentially involved. Details of any prior genetic testing are also requested and where appropriate, radiological images are also shared in advance via an online Image Exchange Portal. Clinicians are also asked to comment on most likely modes of inheritance. Multiple disciplines attend these monthly meetings, including clinical geneticists, adult and paediatric endocrinologists/rheumatologists (including trainees), musculoskeletal radiologists and data analysts. Even when no genetic diagnosis could be established, these meetings aimed to help make recommendations for additional clinical assessment and investigations including imaging, genes to prioritise for detailed analysis and optimum sampling of other family members for segregation analysis (e.g. the parents to complete the trio).

Our initial experiences of these meetings were presented in brief at the Bone Research Society 2021 conference<sup>1</sup> and an early outcome from this initiative has been the description of a novel subtype of spondylometaphyseal dysplasia (MIM #619638) due to a homozygous frameshift variant in *PRKG2*.<sup>2</sup>

Regular participants of the MDT meeting have included:

Dr Meena Balasubramanian (Sheffield Children's NHS Foundation Trust, Sheffield, UK), Dr Judith Bubbear (Royal National Orthopaedic Hospital NHS Trust, UK), Dr Christine Burren (University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK), Dr Alistair Calder (Great Ormond Street Hospital for Children, London, UK), Dr Jo Fairhurst (University Hospital Southampton NHS Foundation Trust, Southampton, UK), Dr Evelien Gevers (Queen Mary University of London, London, UK), Dr David Hunt (University Hospital Southampton NHS Foundation Trust, Southampton, UK), Dr Melita Irving (Guy's and St Thomas' NHS Foundation Trust, London, UK), Dr Kassim Javaid (University of Oxford, Oxford, UK), Dr Zaineb Mohsin (University of Oxford, Oxford, UK), Prof Amaka Offiah (University of Sheffield, Sheffield, UK), Dr Alistair Pagnamenta (University of Oxford, Oxford, UK), Dr Ataf Sabir (Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK), Dr Debbie Shears (Oxford University Hospitals NHS Foundation Trust, Oxford, UK), Prof Sarah Smithson (University Hospitals Bristol NHS Foundation Trust, Bristol, UK), Dr Mohnish Suri (Nottingham University Hospitals NHS Trust, Nottingham, UK), Prof Jenny Taylor (University of Oxford, Oxford, UK), Prof Andrew Wilkie (University of Oxford, Oxford, UK) and Dr Louise Wilson (Great Ormond Street Hospital for Children, London, UK).

### 2. Clinical phenotype – Family 1

In advance of the first MDT meeting, details were shared of a boy (then 12 years old) with clinical features of Greig syndrome including relative macrocephaly, hypertelorism, post-axial polysyndactyly of hands and pre-axial polysyndactyly of feet (Fig. S2). Radiographs identified polysyndactyly of hands (post-axial) and feet (pre-axial) and short broad metacarpals, metatarsals and phalanges. Other family members including paternal grandmother, father and two older siblings were similarly affected (Fig. 1A, Fig. S2), suggesting an autosomal dominant mode of inheritance. Additional clinical features were observed in some family members. In the proband, these included spinal cord lipoma with dermal sinus tract and recurrent tethering of spinal cord with associated syrinx, whereas his sister had an umbilical hernia, elder brother had hypospadias and paternal grandmother, a cerebral astrocytoma. These subsidiary findings are not expected in Greig syndrome and their significance is unclear. The suspected diagnosis could not initially be confirmed molecularly as targeted sequencing analysis conducted in 2004, and repeated in 2015 by a different laboratory, had been negative (Table S4). Karyotype and array-CGH testing of the affected children was also negative.

*GLI3/FBN1 inversions - supplemental*

DNA for the paternal grandmother was unavailable and so recruitment to the 100KGP was limited to the other four affected family members. Filtering of variants by Genomic England's clinical pipeline was performed as a quad family using the complete penetrance option. Gene-panels from PanelApp<sup>3</sup> applied were Skeletal dysplasia 1.192 (352 GREEN genes); Limb disorders 1.44 (114 GREEN genes); Rare multisystem ciliopathy disorders 1.119 (89 GREEN genes), where GREEN genes are those for which there are multiple lines of independent evidence confirming the disease-gene association. These all included *GLI3*, as a GREEN-rated candidate gene and comprised a total of 447 genes. This analysis did not yield any TIER1 or TIER2 candidate variants (loss-of-function/*de novo* or missense variants in known candidate genes for the disease, respectively) and the family was signed-off as unsolved in July 2020.

**3. Clinical phenotype – Family 2**

The proband, a female in her early 40s, is one of three siblings of non-consanguineous Caucasian parents (Fig. 1A). She was first reviewed in the genetics clinic following a termination of pregnancy due to multiple congenital abnormalities. Similar to other affected family members (Fig. S13), she presented with an unusual combination of distal limb and genitourinary tract malformations. A diagnosis of hand-foot-uterus syndrome (MIM #140000) was, therefore, suspected upon initial assessment. The patient was noted to have a bicornuate uterus with solitary vagina and cervix, a unilateral duplex kidney, bilateral broad and proximally placed thumbs (Fig. S8A), bilateral medial displacement of the great toe ("sandal gap"), and bilateral 2/3 toe syndactyly (Fig. S8B).

The proband's elder sister had bilateral preaxial polydactyly ("duplicated thumbs") and bilateral 2/3 toe syndactyly. No urogenital abnormalities were reported. The proband's younger brother was noted to have hypospadias, bilateral undescended testes, bilateral thumb brachydactyly, and bilateral 2-3 toe syndactyly. The proband's mother was reported to have uterine didelphys, double cervix, a longitudinal vaginal septum, bilateral broad thumbs, and bilateral 2/3 toe syndactyly. She had an obstetric history that included three miscarriages: two at 12 weeks and one at 28 weeks (Fig. 1A). All affected family members had typical neurodevelopment.

Post-mortem (PM) examination of her aborted male fetus at 17 weeks gestation showed short humeri and lower limbs, marked mandibular recession, broad thumbs showing duplication of the terminal phalanges, absent rectum and anus with narrow colo-vesical fistula, absent prostate gland, single umbilical artery, urethral agenesis with early urethral obstruction sequence including megacystis, bilateral hydroureter, hydronephrosis, and renal dysplasia. Genetic investigations on uncultured placental material (targeted QF-PCR aneuploidy test) revealed trisomy for at least the region of chromosome 15 represented by five informative markers.

Complete trisomy 15, in a non-mosaic form, would be expected to contribute to fetal demise. Nevertheless, the post mortem examination revealed specific familial phenotypic features, for example; broad duplicated thumbs and urogenital anomalies (urethral agenesis). This raised the suspicion of dual genetic diagnoses in the proband's deceased son, who might have harboured the familial 14.8Mb inversion.

The skeletal limb features observed in Family 2 are congruent with a *GLI3* disruption, namely the short broad thumbs, the preaxial polydactyly, and the 1-3 toe syndactyly. The overall phenotype including urogenital abnormalities, however, was more consistent with a *HOXA13*-related disorder. Interestingly, the 7p15.2 breakpoint of the 14.8Mb inversion lies ~45kb upstream of *HOXA13*. We hypothesize that the familial inversion contributed to the phenotype observed in Family 2 via *GLI3* disruption, in addition to a potential *HOXA13* enhancer delocalization effect.

The 100KGP clinical pipeline had yielded a negative result (Sept 2019) with TIERING having been performed prioritising variants in 120 genes made up of the following panels: Limb disorders v1.2 (114 GREEN genes; *GLI3* and *HOXA13* both GREEN), Radial dysplasia 1.6 (47 GREEN genes; *GLI3* absent, *HOXA13* GREEN) and VACTERL-like phenotypes v1.22 (16 green genes; *GLI3* listed as RED, *HOXA13* GREEN).

**4. Clinical phenotype – Family 3**

The proband was first seen in the genetics clinic in her early teens and has had a possible diagnosis of Marfan syndrome for many years. Skeletal features were typical of the condition, with an upper segment:lower segment (US:LS) ratio of 0.79 (normal >0.85), positive wrist and thumb signs, striae over the knees, upper legs, and lower

*GLI3/FBN1 inversions - supplemental*

back, mild pectus excavatum and mild scoliosis, with hypermobility, and a high palate. She had soft, stretchy, skin. An echocardiogram performed at age 11 showed marked aortic root dilatation with a diameter of 3.68cm (>95th centile). She was started on beta-blockers, aged 13. A spinal MRI identified dural ectasia at age 18. The echocardiogram also identified mitral valve prolapse and eventually she had an aortic root replacement aged 23. Genetic testing of *FBN1* was initially performed using DHPLC and MLPA. Aortic gene panel testing via Next Generation Sequencing (12 gene panel) in 2014 also did not identify a cause. Although previous testing identified a variant of unknown significance in *FBN2* (NM\_001999.4:c.976C>T, p.Pro326Ser), this was not shared by the affected mother.

Her mother had a reduced US:LS ratio of 0.72, had easy bruising, striae on her thighs, pes planus, long toes and dental overcrowding. She is myopic. An echocardiogram aged 51 was unremarkable but when reviewed aged 68 she was noted to have pectus carinatum, typical facial features, striae, reduced elbow extension, and a CT aorta identified a mildly dilated sinus of Valsalva of 4.2cm (z score=1.95), and coeliac and splenic artery aneurysms. There was no other relevant family history. Over the years, other conditions were considered but Marfan syndrome was always the primary clinical diagnosis for this family.

The clinical analysis pipeline run by Genomic England focussed on genes in the thoracic aortic aneurysm or dissection panel (v1.112) from PanelApp, which contains *FBN1* as a GREEN gene. The total number of GREEN genes that were on this panel was 31. Although the genomes were originally sequenced in 2016 and reanalysed again in 2020, no TIER1 or TIER2 variants were reported.

#### 5. Filtering structural variants with SVRare

We systematically reviewed results of gene-oriented analyses of SVs using SVRare (17<sup>th</sup> November 2021 version).<sup>4</sup> SVRare was built on a MySQL database that hosted 554 million SVs from 71,408 100KGP participants from 33,924 families. Of these families, 232 (including Family 1) were recruited due to unexplained skeletal dysplasia, 24 (including Family 2) due to radial dysplasia and 664 (including Family 3) due to familial thoracic aortic aneurysm disease. For the current iteration, analysis is limited to deletions, duplications and inversions. In brief, SVRare calculates the similarity between SVs of the same type by using the fraction of the overlap (intersection) over the total length (union), and SVs are considered the same if their similarity score is higher than 80%. This enables estimation of allele frequency for rare SV prioritisation. Once clustered, variants were filtered out if they were seen in >1% of individuals. To further aid discovery of disease-causing SVs, the tool also annotates each SV for familial segregation and predicts protein-coding disruption. SVRare prioritised SVs are freely available to GeCIP members as individual gene reports in the “re\_gecip/shared\_allGeCIPs/JingYu-SV-query” directory within the GEL research environment.

This analysis identified 4 families with predominantly balanced inversions which the breakpoint analysis suggested would unequivocally disrupt gene function and where the phenotype was consistent with well-known genetic conditions. This included re-identification of the 1.2Mb *GLI3* inversion in Family 1 and a 13Mb *TWIST1* inversion, published previously in a mother-son duo with craniosynostosis.<sup>5</sup> Two additional families were identified, as described in more detail below (Families 2 and 3).

#### 6. Detailed summary of inversions involving *GLI3*

The SVRare report for *GLI3* identified 15 rare inversions (3.8kb – 159.3Mb) that overlapped *GLI3* (Fig. S6). Of these, review of breakpoint positions indicated that only 3 would likely lead to complete gene inactivation (i.e. at least 1 breakpoint lying within the gene). One of these inversions was the same 1.2Mb inversion already identified in Family 1. Strikingly, a second quad family (Family 2) was identified where all 4 individuals had been submitted to the 100KGP with a clinical diagnosis of radial dysplasia and all 4 harboured a 14.8Mb inversion disrupting *GLI3*. This inversion had been called by Manta as involving chr7:27,245,456-42,072,394. Like Family 1, the breakpoint for Family 2 lay in intron 4 of *GLI3* but this time it was the proximal not distal end of the rearrangement (Fig. 1B). Clinical information about Family 2 is provided above. The third inversion was a much smaller 6.6kb event that would potentially invert exon 4, however closer scrutiny of read alignments suggested this variant to be an artefact, possibly due to a rare intronic deletion lying nearby.

*GLI3/FBN1 inversions - supplemental***7. Detailed summary of inversions involving *FBN1***

There were 12 rare inversions involving *FBN1* that were prioritised by SVRare that ranged in size by 3 orders of magnitude (78.5kb – 78.2Mb). Of these, only 3 inversions had breakpoints internal to the gene and thus would be predicted to disrupt gene function (Fig. S7). One of these was a 1.97Mb inversion for which the distal breakpoint lay in intron 4 of *FBN1* (Fig. 1C). This variant was called by Manta as chr15:46,635,052-48,604,302 and detected in the proband and her affected mother, both of whom had been recruited to the 100KGP with a diagnosis of “familial thoracic aortic aneurysm disease”. Clinical information about this family is provided above.

The other two inversions involving *FBN1* were present in the same family – further scrutiny of the read alignments suggested they were in fact part of a complex DUP-INV-DUP structural variant (Fig. S14). The two inversions (from Manta) and two duplicated segments (from Canvas) could be explained by at least two different genetic conformations. However, neither conformation would alter gene dosage in terms of the number of complete copies of *FBN1*. Unaltered dosage would be consistent with the fact that individuals in this family did not exhibit a Marfan-like phenotype. This highlights that caution must be taken with DUP-INV-DUP type variants because with short read data there is often ambiguity regarding precisely how the genomic segments are organised and these type of rearrangement can have no effect on gene-dosage.

**8. Supplemental discussion**

This study commenced with MDT-based review of Family 1, a family for whom Genomics England’s clinical pipeline had assessed SNVs and indels in 447 genes. From information shared prior to our MDT meeting it quickly became apparent that the clinician’s initial interest was to double-check just 1 or 2 genes really thoroughly. Prior to 100KGP, targeted sequencing of *GLI3* had been done not once but twice, demonstrating the high level of suspicion for *GLI3* being involved for this family. This long diagnostic odyssey was solved by manually reviewing read alignments in IGV. We provide IGV screenshots showing the characteristic read alignment signature associated with balanced inversions (Fig. 1B,C) and hope this may prompt clinical scientists to scrutinise other unsolved cases with WGS data where there is a strong clinical suspicion pointing to a single gene.

We then developed a robust pipeline for prioritisation of rare high-confidence inversions and used this to focus on 43 genes linked to skeletal disorders, which allowed us to identify 2 more families with rare inversions. By coincidence, for all 3 inversions the breakpoint of interest lay in intron 4 and thus are highly likely to result in loss of function as it is very hard to see how gene-function, in cases with such a rearrangement so early on in a gene, could be retained.

One limitation of our follow up analysis is that it involved a manual review step and thus not every inversion was scrutinised with an equal degree of attention. In addition, inversions can often impact on gene expression via a positional effects and our prioritisation strategy may have missed such cases. Positional effects are exemplified in Family 2 where the distal breakpoint lies just 45kb upstream of *HOXA13*, but the inversion was only picked up on account of the proximal breakpoint which disrupts *GLI3*. In Family 1, we also speculate that disruption to *HECW1* could potentially explain some of the atypical features seen. Together, these cases highlight the importance of considering both breakpoint regions for rare balanced inversions.

Even with WGS data, balanced inversions can only be picked up by Manta or other algorithms that use split-read information. Currently the Genomics England structural variant pipeline only utilises Canvas and even with copy number variants there is typically a 10kb limit to resolution. This study helps emphasise the importance of developing clinical SV prioritisation pathways that can integrate multiple calling algorithms.

**9. The Genomics England Research Consortium**

John C. Ambrose<sup>1</sup>; Prabhu Arumugam<sup>1</sup>; Roel Bevers<sup>1</sup>; Marta Bleda<sup>1</sup>; Freya Boardman-Pretty<sup>1,2</sup>; Christopher R. Boustred<sup>1</sup>; Helen Brittain<sup>1</sup>; Mark J. Caulfield<sup>1,2</sup>; Georgia C. Chan<sup>1</sup>; Greg Elgar<sup>1,2</sup>; Tom Fowler<sup>1</sup>; Adam Giess<sup>1</sup>; Angela Hamblin<sup>1</sup>; Shirley Henderson<sup>1,2</sup>; Tim J. P. Hubbard<sup>1</sup>; Rob Jackson<sup>1</sup>; Louise J. Jones<sup>1,2</sup>; Dalia Kasperaviciute<sup>1,2</sup>; Melis Kayikci<sup>1</sup>; Athanasios Kousathanas<sup>1</sup>; Lea Lahnstein<sup>1</sup>; Sarah E. A.

*GLI3/FBN1 inversions - supplemental*

Leigh<sup>1</sup>; Ivonne U. S. Leong<sup>1</sup>; Javier F. Lopez<sup>1</sup>; Fiona Maleady-Crowe<sup>1</sup>; Meriel McEntagart<sup>1</sup>; Federico Minneci<sup>1</sup>; Loukas Moutsianas<sup>1,2</sup>; Michael Mueller<sup>1,2</sup>; Nirupa Murugaesu<sup>1</sup>; Anna C. Need<sup>1,2</sup>; Peter O'Donovan<sup>1</sup>; Chris A. Odhams<sup>1</sup>; Christine Patch<sup>1,2</sup>; Mariana Buongiorno Pereira<sup>1</sup>; Daniel Perez-Gil<sup>1</sup>; John Pullinger<sup>1</sup>; Tahrima Rahim<sup>1</sup>; Augusto Rendon<sup>1</sup>; Tim Rogers<sup>1</sup>; Kevin Savage<sup>1</sup>; Kushmita Sawant<sup>1</sup>; Richard H. Scott<sup>1</sup>; Afshan Siddiq<sup>1</sup>; Alexander Sieghart<sup>1</sup>; Samuel C. Smith<sup>1</sup>; Alona Sosinsky<sup>1,2</sup>; Alexander Stuckey<sup>1</sup>; Mélanie Tanguy<sup>1</sup>; Ana Lisa Taylor Tavares<sup>1</sup>; Ellen R. A. Thomas<sup>1,2</sup>; Simon R. Thompson<sup>1</sup>; Arianna Tucci<sup>1,2</sup>; Matthew J. Welland<sup>1</sup>; Eleanor Williams<sup>1</sup>; Katarzyna Witkowska<sup>1,2</sup>; Suzanne M. Wood<sup>1,2</sup>.

1. Genomics England, London, UK

2. William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.

## GLI3/FBN1 inversions - supplemental

## Tables

**Table S1:** Selected QC statistics for WGS data from 10 individuals from 3 families with diagnostic inversions in *GLI3* (F1, F2) or *FBN1* (F3). QC based on data mapped to GRCh38 with decoys. \*originally analysed on GRCh37. †mean across all 100KGP samples is 427. The mean across these 10 samples is 418.

ID	F1 proband	F1 brother	F1 sister	F1 father	F2 proband	F2 sister	F2 brother	F2 mother	F3 proband	F3 mother
Delivery date (remapping)	May 2017	Aug 2017	Aug 2017	May 2017	Apr 2018	May 2018	Apr 2018	Apr 2018	Apr2016* (Feb 2020)	May2016* (Feb 2020)
Total aligned reads	871,004,459	1,052,818,353	932,959,733	755,979,204	1,025,504,334	749,957,343	884,423,031	815,883,876	816,243,785	1,281,867,717
Percent duplicate aligned reads	8.77%	9.14%	7.66%	6.78%	5.82%	6.02%	8.03%	8.52%	7.61%	14.55%
Percent aligned reads	93.19%	92.88%	92.73%	92.15%	93.98%	91.81%	92.30%	93.60%	95.15%	94.78%
Percent read pairs aligned to different chromosomes	0.17%	0.17%	0.15%	0.19%	0.31%	0.76%	1.50%	0.53%	0.71%	0.43%
Percent soft-clipped bases	2.01%	1.82%	1.94%	1.91%	1.74%	1.42%	1.64%	1.56%	1.84%	2.02%
Mean coverage	39.64	47.78	43.02	35.18	48.31	35.36	40.67	37.39	37.75	54.83
Coverage at 15X	97.40%	98.07%	97.32%	96.50%	97.56%	97.04%	97.64%	97.22%	97.44%	97.93%
Fragment length median	493	490	473	500	448	495	442	484	437	491
SNVs (All)	3897444	3882458	3914338	3884196	3926566	3945642	3886341	3905448	3910612	3969352
SNV Het/Hom ratio	1.605	1.553	1.562	1.57	1.595	1.606	1.584	1.527	1.556	1.574
SNV Ts/Tv ratio	2.061	2.066	2.062	2.059	2.062	2.06	2.06	2.059	2.057	2.067
SNVs (Percent found in dbSNP)	94.63%	94.69%	94.75%	94.62%	94.59%	94.42%	94.54%	94.59%	94.41%	94.61%
Indels (All)	959057	971729	968234	936189	1003501	967602	977021	974545	971216	1005026
SV Inversions (All)†	406	432	379	341	481	361	400	427	432	519
SV Inversions in genes	235	251	215	176	290	203	226	246	258	310

## GLI3/FBN1 inversions - supplemental

**Table S2:** Details of 43 autosomal genes listed in 2019 revision of the skeletal disorder nosology<sup>6</sup> for which ClinGen assess as having “sufficient evidence” supporting haploinsufficiency (downloaded 10<sup>th</sup> January 2022).

Gene symbol	Approved name	HGNC ID	Location
<i>ALX4</i>	ALX homeobox 4	HGNC:450	11p11.2
<i>ARID1B</i>	AT-rich interaction domain 1B	HGNC:18040	6q25.3
<i>CDKN1C</i>	cyclin dependent kinase inhibitor 1C	HGNC:1786	11p15.4
<i>COL1A1</i>	collagen type I alpha 1 chain	HGNC:2197	17q21.33
<i>COL2A1</i>	collagen type II alpha 1 chain	HGNC:2200	12q13.11
<i>CREBBP</i>	CREB binding protein	HGNC:2348	16p13.3
<i>EFTUD2</i>	elongation factor Tu GTP binding domain containing 2	HGNC:30858	17q21.31
<i>EP300</i>	E1A binding protein p300	HGNC:3373	22q13.2
<i>ERF</i>	ETS2 repressor factor	HGNC:3444	19q13.2
<i>EXT1</i>	exostosin glycosyltransferase 1	HGNC:3512	8q24.11
<i>EXT2</i>	exostosin glycosyltransferase 2	HGNC:3513	11p11.2
<i>FBN1</i>	fibrillin 1	HGNC:3603	15q21.1
<i>FGF10</i>	fibroblast growth factor 10	HGNC:3666	5p12
<i>FGFR1</i>	fibroblast growth factor receptor 1	HGNC:3688	8p11.23
<i>GDF5</i>	growth differentiation factor 5	HGNC:4220	20q11.22
<i>GLI3</i>	GLI family zinc finger 3	HGNC:4319	7p14.1
<i>GNAS</i>	GNAS complex locus	HGNC:4392	20q13.32
<i>HOXD13</i>	homeobox D13	HGNC:5136	2q31.1
<i>KAT6B</i>	lysine acetyltransferase 6B	HGNC:17582	10q22.2
<i>LEMD3</i>	LEM domain containing 3	HGNC:28887	12q14.3
<i>LMX1B</i>	LIM homeobox transcription factor 1 beta	HGNC:6654	9q33.3
<i>MNX1</i>	motor neuron and pancreas homeobox 1	HGNC:4979	7q36.3
<i>MYCN</i>	MYCN proto-oncogene, bHLH transcription factor	HGNC:7559	2p24.3
<i>NF1</i>	neurofibromin 1	HGNC:7765	17q11.2
<i>NIPBL</i>	NIPBL cohesin loading factor	HGNC:28862	5p13.2
<i>NOG</i>	noggin	HGNC:7866	17q22
<i>NSD1</i>	nuclear receptor binding SET domain protein 1	HGNC:14234	5q35.3
<i>PAX3</i>	paired box 3	HGNC:8617	2q36.1
<i>POLR1D</i>	RNA polymerase I and III subunit D	HGNC:20422	13q12.2
<i>PTPN11</i>	protein tyrosine phosphatase non-receptor type 11	HGNC:9644	12q24.13
<i>SALL4</i>	spalt like transcription factor 4	HGNC:15924	20q13.2
<i>SF3B4</i>	splicing factor 3b subunit 4	HGNC:10771	1q21.2
<i>SHH</i>	sonic hedgehog signalling molecule	HGNC:10848	7q36.3
<i>SMAD3</i>	SMAD family member 3	HGNC:6769	15q22.33
<i>SMAD4</i>	SMAD family member 4	HGNC:6770	18q21.2
<i>SMARCB1</i>	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	HGNC:11103	22q11.23
<i>TBX3</i>	T-box transcription factor 3	HGNC:11602	12q24.21
<i>TBX4</i>	T-box transcription factor 4	HGNC:11603	17q23.2
<i>TBX5</i>	T-box transcription factor 5	HGNC:11604	12q24.21
<i>TCF12</i>	transcription factor 12	HGNC:11623	15q21.3
<i>TCOF1</i>	treacle ribosome biogenesis factor 1	HGNC:11654	5q32-q33.1
<i>TRPS1</i>	transcriptional repressor GATA binding 1	HGNC:12340	8q23.3
<i>TWIST1</i>	twist family bHLH transcription factor 1	HGNC:12428	7p21.1

## GLI3/FBN1 inversions - supplemental

**Table S3:** Primers used for Sanger validation in Families 1, 2 and 3.

Family	Primer name	Sequence	End of inversion
Family 1	GLI3_Inv1_F_V1	TACTGCTGAGAAGCAACAGTG	Distal
	GLI3_Inv1_R_V1	CAGCTTTCTTAGATATGATATAC	
	GLI3_Inv2_F_V1	AGTATATACTAGGCTCAGTACATG	Proximal
	GLI3_Inv2_R_V1	GAAGGTTAGGGTGTATAAATGAC	
Family 2	GLI3-1F	CCGGGAGAACTACGTATCCA	Distal
	GLI3-2F	CCCTGCTTTGGAAAATGAAT	
	GLI3-1R	TGTGTGTATGGGAGGAGCAG	Proximal
	GLI3-2R	TGGGAATGTAGGCAATTGGT	
Family 3	FBN1-INV-1F	TCCCAAGACGAAATGAACTT	Proximal
	FBN1-INV-2F	GGCACCTGGATCTCAATACCT	
	FBN1-INV-1R	CCCTCTGTGACAATGCCAAG	Distal
	FBN1-INV-2R	GTGTGTCTTTAGGCATCCCC	

## GLI3/FBN1 inversions - supplemental

Table S4: Details of diagnostic odysseys and prior genetic testing for Families 1-3.

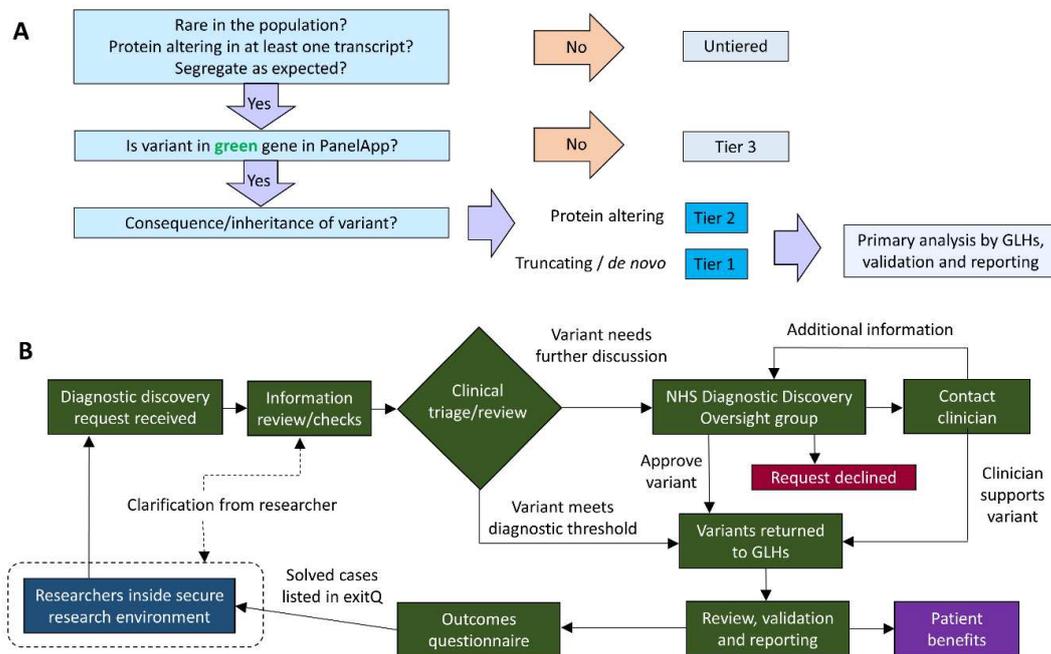
Family number	Family 1	Family 2	Family 3
Date variant detected	February 2021	November 2021	November 2021
Proband WGS date	May 2017	April 2018	April 2016 (Feb 2020 for remapping to GRCh38)
Negative report issued from 100KGP	July 2020	September 2019	May 2018
Time from WGS to variant being identified	3¾ years	3½ years	5½ years
Start of diagnostic odyssey	2004	August 2013	Family known to clinical genetics since ca. 2002
Gender (proband)	Male	Female	Female
Ethnicity	White British	White British	White British
GRCh38 coordinates from Manta (reciprocal call)	chr7:42,051,297-43,254,780 (chr7:42,051,291-43,254,759)	chr7:27,245,456-42,072,394 (chr7:27,245,448-42,072,397)	chr15:46,635,052-48,604,302 (chr15:46,635,053-48,604,300)
Size of inversion	1.20Mb	14.83Mb	1.97Mb
Position of breakpoint in gene	Intron 4 of <i>GLI3</i> (NM_000168.6)	Intron 4 of <i>GLI3</i> (NM_000168.6)	Intron 4 of <i>FBN1</i> (NM_000138.5)
Other breakpoint	Intron 3 of <i>HECW1</i> (NM_015052.5)	Intron 2 of <i>EVX1</i> (NM_001989.5); ~45kb from <i>HOXA13</i>	No genes nearby
Family structure	Affected brother/sister/father all in 100KGP	Affected sister/brother/mother all in 100KGP	Patient and her affected mother are in 100KGP
Recruitment diagnosis	Unexplained skeletal dysplasia	Radial dysplasia	Familial Thoracic Aortic Aneurysm Disease
Cytogenetic testing	Karyotyping done but no details available	No indication that karyotyping was ever done, but has now been requested	N/A
Array testing (date, array type/version)	Array CGH done but no details available	Agilent 60K aCGH (design version 028469). Median resolution: 120kb. Oct 2013	N/A
MLPA testing	N/A	N/A	Genetic testing of <i>FBN1</i> for large deletions and duplications by MLPA in 2005
Targeted sequencing	Targeted sequencing of <i>GLI3</i> first in 2004 (Biesecker lab) and then repeated in 2015 using PCR-Sanger method. <i>HOXD13</i> also sequenced in 2006-7 in Oxford	Analysis of <i>HOXA13</i> and flanking intronic sequences by PCR multiplex AmpliSeq (IAD47762_93, Ion Torrent. Also by PCR and Sanger sequencing. Done in Lille (Hopital Jeanne de Flandre) January 2015	NGS testing of <i>ACTA2</i> , <i>COL3A1</i> , <i>EFEMP2</i> , <i>FBN1</i> , <i>FBN2</i> , <i>FLNA</i> , <i>MYH11</i> , <i>MYLK</i> , <i>NOTCH1</i> , <i>SK1</i> , <i>SLC2A</i> , <i>SMAD3</i> in 2015, using the Illumina Trusight One sequencing panel
Exome testing	N/A	WES at Viapath with Agilent SureSelectXT Clinical Research Exome (SureSelectXT Human All Exon V5 baited with clinically relevant genes). The enriched exome libraries WES using paired-end, 125 cycle chemistry on an Illumina HiSeq2500. November 2016.	N/A

## GLI3/FBN1 inversions - supplemental

<b>Other genetic testing</b>	N/A	The proband's aborted fetus has had QF-PCR aneuploidy test for chromosomes 13, 14, 15, 16, 18, 21, 22, X and Y. The test showed trisomy for at least the region of chromosome 15 represented by five informative markers. The test was done on uncultured placental material and is likely the cause of the fetal loss.	Testing of <i>FBN1</i> by DHPLC in 2005
<b>Other variants from GEL pipeline</b>	No TIER1/2 from GEL pipeline	No TIER1/2 from GEL pipeline	No TIER1/2 from GEL pipeline. Previous testing identified <i>FBN2</i> variant NM_001999.4:c.976C>T (p.Pro326Ser) in proband. Variant not seen in mother and too common in gnomAD so LB/B, see <a href="http://www.ncbi.nlm.nih.gov/clinvar/variation/137316">www.ncbi.nlm.nih.gov/clinvar/variation/137316</a> .
<b>Validation of inversion</b>	PCR of both inversion breakpoints and Sanger sequencing in 4/4 affected individuals - May 2021	PCR of both inversion breakpoints in 4/4 affected individuals and Sanger sequencing (proband only) – January 2022; karyotyping is underway	PCR and Sanger sequencing – May 2022
<b>Reason to check gene and detection method</b>	Clinical suspicion for <i>GLI3</i> mutation shared with data analyst prior to March 2021 MDT meeting. Detected by manual review of read alignments but also called by Manta	AD gene mentioned in Mortier <i>et al</i> 2019 <sup>6</sup> where haploinsufficiency is a known mechanism of pathogenesis. Detection by Manta following genome sequencing in 100KGP and prioritised by SVRare <sup>4</sup>	

## GLI3/FBN1 inversions - supplemental

## Supplemental Figures



**Figure S1:** Summary of clinical tiering and researcher identified variant pipelines linked to the 100K Genomes Project. A) Schematic of the clinical tiering pipeline employed by Genomic England. Although only Tier 1 and 2 variants are assessed routinely by the respective Genomic Laboratory Hub (GLH), Tier 3 variants are also available for review, if required. B) Researchers may discover variants that could represent a diagnosis for a participant. This diagram summarises the official pathway to feed back these findings to the NHS Diagnostic Discovery Oversight Group. This group comprises clinicians and scientists from each GLH, whom meet every 2 months with NHS England and Genomics England representatives. This pipeline provides assurance to the Genomic Medicine Service that the diagnoses being returned are of high quality and clinical relevance. A separate pathway exists for the rapid return of variants considered to be urgent.

*GLI3/FBN1 inversions - supplemental*

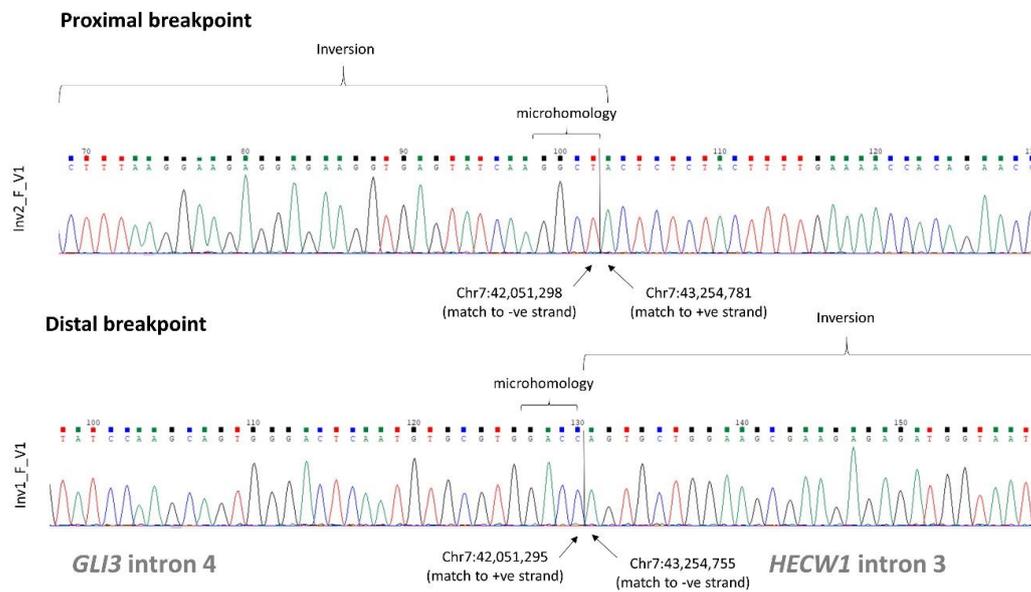
**Figure S2:** Distal limb anomalies observed in siblings of Family 1 are shown, including symmetrical pre-axial polysyndactyly of feet of sister (above) and post-axial polysyndactyly of hands of brother (mid-section). Radiographs of the feet of the proband (below, left and centre) illustrate the interfamilial variability of pre-axial polydactyly with a single proximal phalanx on each side. The first metatarsals are broad. Radiograph of right hand of brother (below right) shows 3-4 soft tissue and bone (terminal phalangeal) syndactyly with relatively short metacarpals and middle/terminal phalanges (post-axial additional digit previously surgically removed).

## GLI3/FBN1 inversions - supplemental



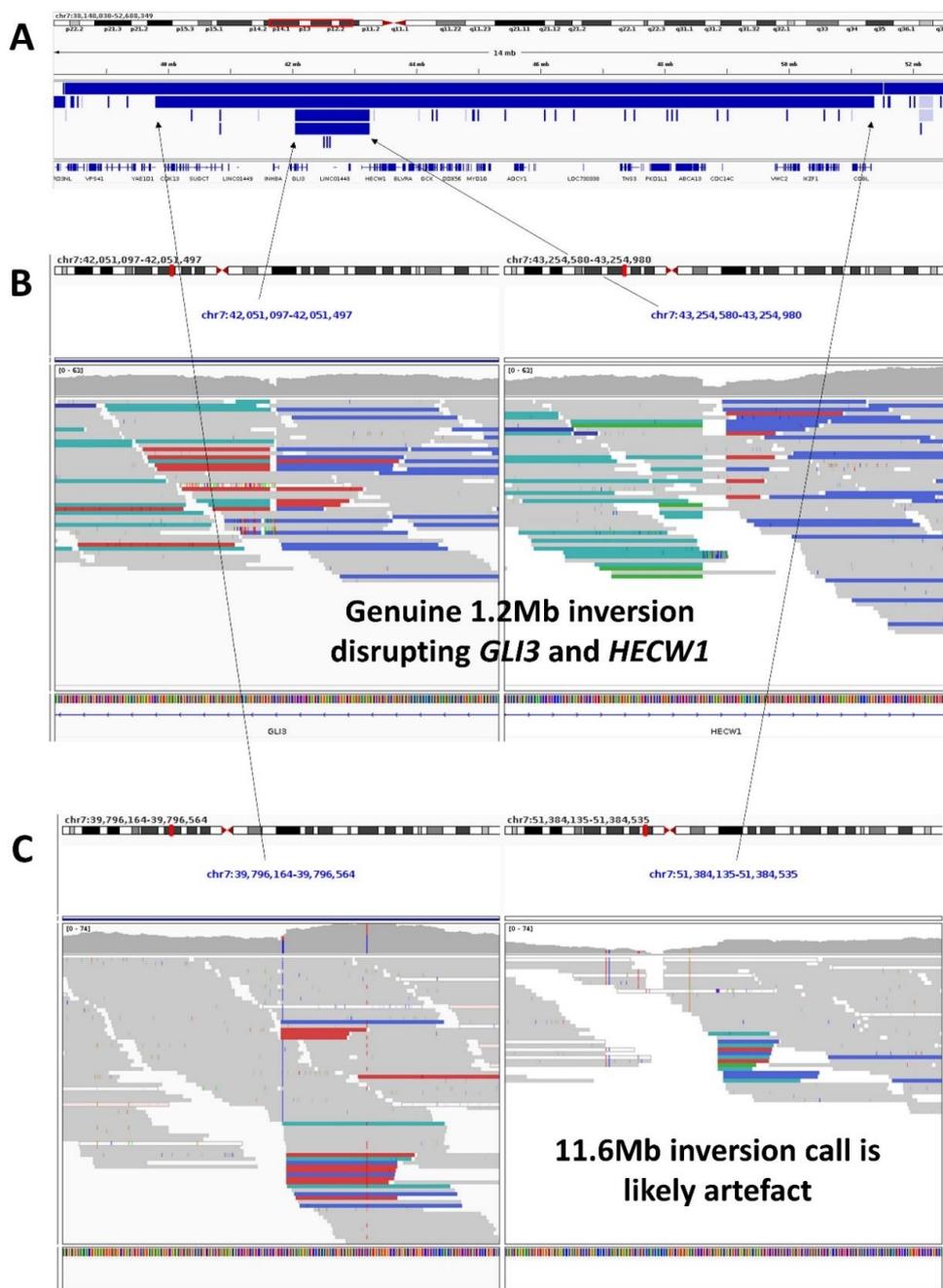
**Figure S3:** Sanger validation and primer positions for distal end of the *GLI3* inversion in Family 1. PCR primer positions are shown for 2 of the 4 primers. Amplicons for 4 affected family members were sequenced bidirectionally and the resulting sequences were uploaded to the UCSC genome browser in FASTA format using the Blat Search tool. An interactive view is shown at [https://genome.ucsc.edu/s/AlistairP/GLI3\\_SANGER](https://genome.ucsc.edu/s/AlistairP/GLI3_SANGER) where one can navigate to the proximal end of the inversion and see the same pattern. The breakpoints defined by Manta are consistent with the Sanger traces.

## GLI3/FBN1 inversions - supplemental



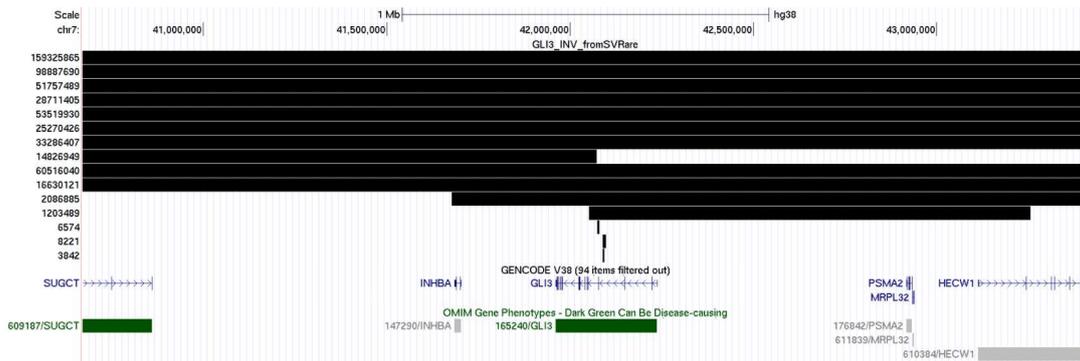
**Figure S4:** Sanger sequencing electropherograms confirming breakpoints for the 1.2Mb inversion in Family 1. Close scrutiny reveals a 25bp deletion at the proximal end and 4bp of microhomology at the respective junctions. Genomic positions are based on GRCh38.

## GLI3/FBN1 inversions - supplemental

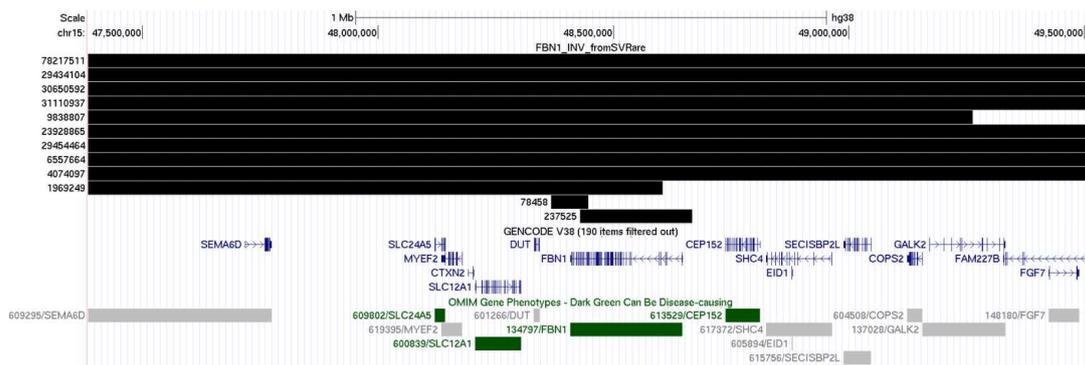


**Figure S5:** IGV screenshot showing Manta/Canvas SV calls and read alignments in the proband from Family 1. A) Manta calls from structural variant vcf file suggests an inversion within a larger inversion on 7p. Genuine inversions such as the 1.2Mb inversion that disrupts *GLI3* are often reported twice in the SV.vcf in a reciprocal manner. B) +ve to +ve read pairs (green) and -ve to -ve read pairs (blue) are seen on each side of the breakpoints. At this level of zoom, the small deleted region is visible at the proximal end of the 1.2Mb inversion. C) Artefactual inversion calls such as the larger 11.6Mb one shown above are only supported by one breakpoint (in this case -ve to -ve strand read pairs) and coverage is more variable.

## GLI3/FBN1 inversions - supplemental



**Figure S6:** Screenshot of UCSC genome browser graphic showing positions of rare inversion calls overlapping *GLI3* in the rare disease arm of the 100KGP. The custom track “*GLI3\_INV\_fromSVRare*” contains 15 entries labelled by size in bp, of which only 2 large inversions directly disrupt *GLI3*. The first of these was reidentification of the 1.2Mb inversion seen in Family 1. The second was a 14.8Mb inversion in Family 2. The 6574bp event would in theory invert a single exon but review of read alignments suggested that this may be an artefact on account of a nearby intronic deletion. An interactive version is available here: [https://genome.ucsc.edu/s/AlistairP/GLI3\\_INVERSION\\_F2](https://genome.ucsc.edu/s/AlistairP/GLI3_INVERSION_F2).



**Figure S7:** Screenshot of UCSC genome browser showing positions of rare inversion calls overlapping *FBN1* in the rare disease arm of the 100KGP. Only 3 of the 12 rare inversions prioritised by SVRare and shown in the track “*FBN1\_INV\_fromSVRare*” have breakpoints which disrupt *FBN1* directly. The 1.97Mb inversion identified in Family 3 likely results in loss of function. The 78.5kb and 237.5kb inversions are from the same family and appear to represent a complex DUP-INV-DUP which is inherited from an unaffected parent. An interactive version is available here: [https://genome.ucsc.edu/s/AlistairP/FBN1\\_INV\\_SVRare](https://genome.ucsc.edu/s/AlistairP/FBN1_INV_SVRare).

## GLI3/FBN1 inversions - supplemental

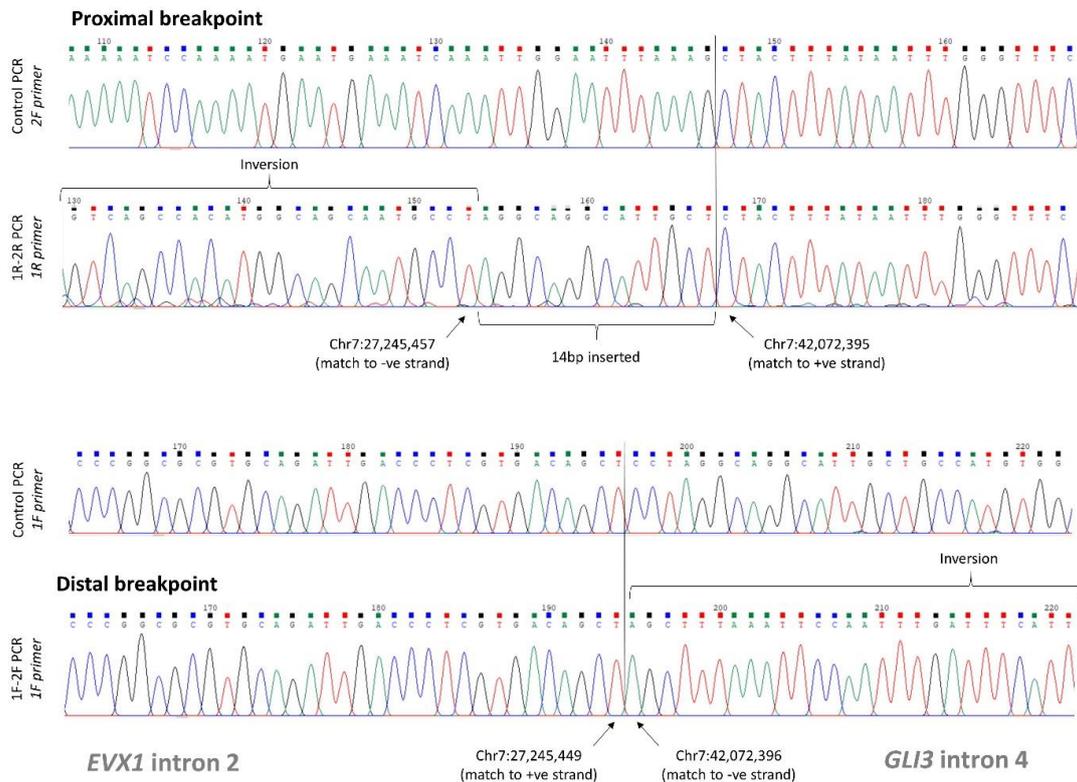


**Figure S8:** Clinical photographs of proband in Family 2. A) Hands showing bilateral short and broad thumbs. B) Feet displaying bilateral 2/3 toe syndactyly and sandal gap.

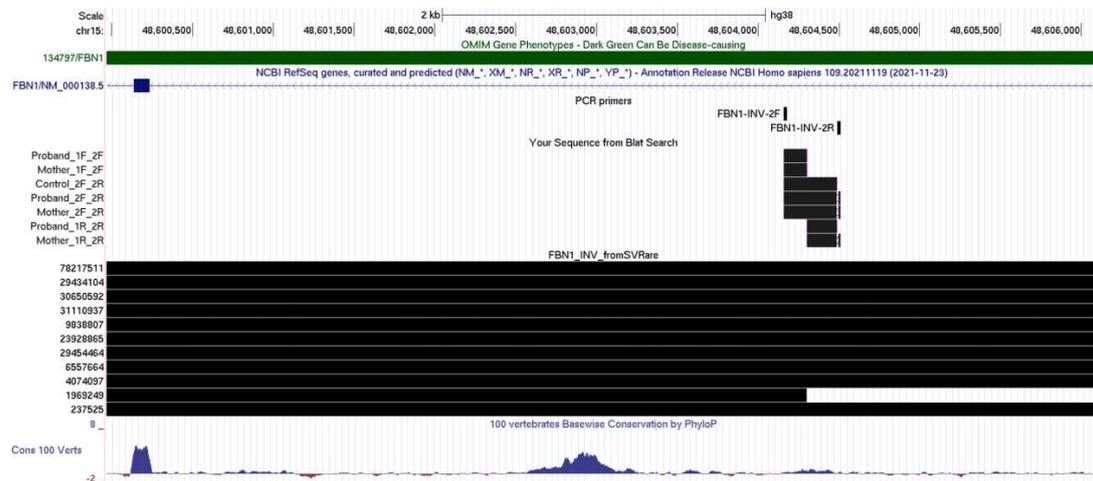


**Figure S9:** Sanger validation and primer positions for proximal end of the *GLI3* inversion in Family 2. PCR primer positions are shown for 2 of the 4 primers. PCR amplicons for the proband were sequenced bidirectionally and the resulting sequences were uploaded to the UCSC genome browser in FASTA format using the Blat Search tool. An interactive view is shown at [https://genome.ucsc.edu/s/AlistairP/GLI3\\_INVERSION\\_F2\\_SANGER](https://genome.ucsc.edu/s/AlistairP/GLI3_INVERSION_F2_SANGER) where one can navigate to the distal end of the inversion and see the same pattern. Breakpoints for the 14.8Mb inversion as defined by Manta are consistent with the Sanger data.

## GLI3/FBN1 inversions - supplemental

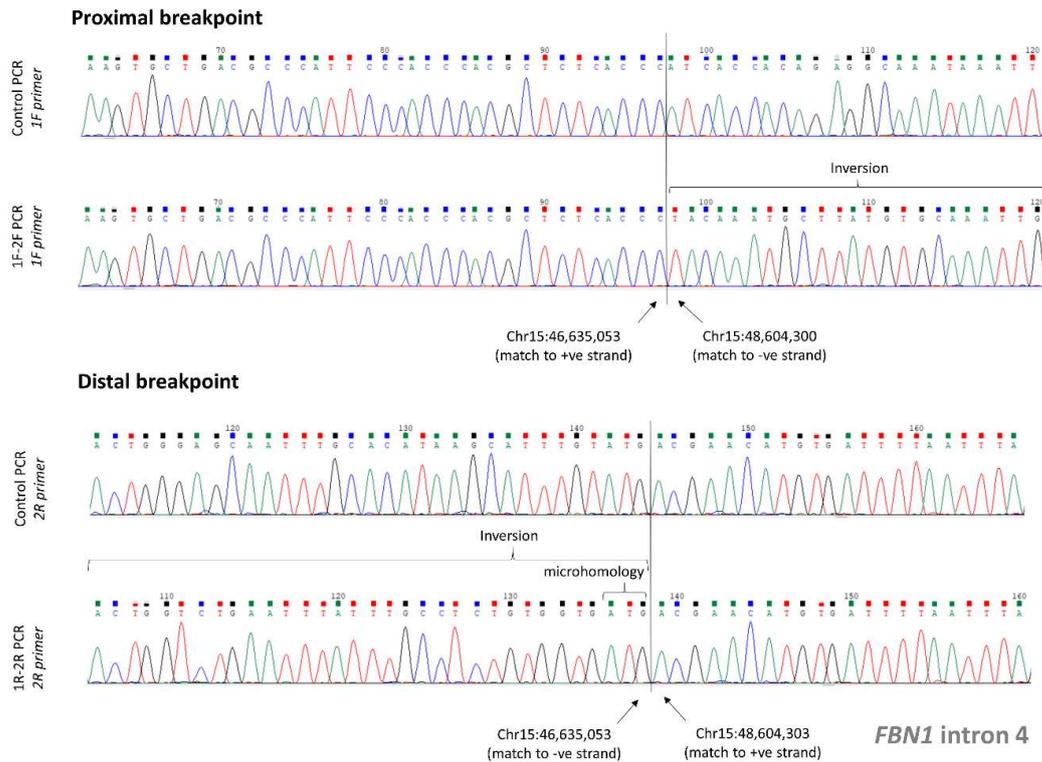


**Figure S10:** Sanger sequencing electropherograms confirming breakpoints for the 14.8Mb inversion in Family 2. Close scrutiny reveals a 14bp insertion at the proximal end and a 6bp deletion at the distal end. Genomic positions are based on GRCh38.



**Figure S11:** Sanger validation and primer positions for distal end of the *FBN1* inversion in Family 3. PCR primer positions are shown for 2/4 primers. Amplicons for the proband, mother and a control were sequenced and the resulting sequences were uploaded to the UCSC genome browser in FASTA format using the Blat Search tool. An interactive view is available at [https://genome.ucsc.edu/s/AlistairP/FBN1\\_INVERSION\\_F3\\_SANGER](https://genome.ucsc.edu/s/AlistairP/FBN1_INVERSION_F3_SANGER) where one can navigate to the proximal end of the inversion and see the same pattern. Breakpoints for the 1.97Mb inversion as defined by Manta are consistent with the Sanger data.

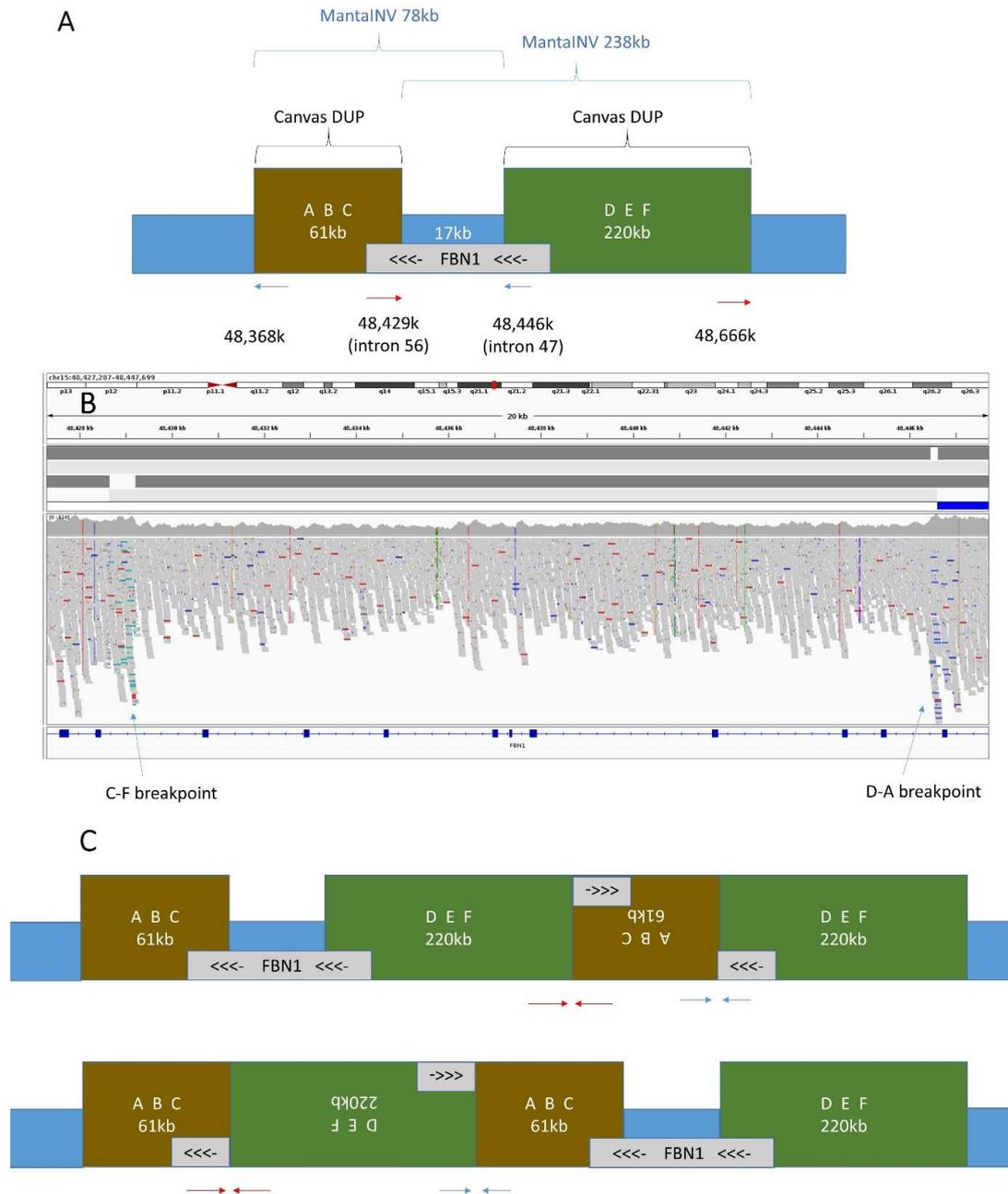
## GLI3/FBN1 inversions - supplemental



**Figure S12:** Sanger sequencing electropherograms confirming breakpoints for the 1.97Mb inversion in Family 3. Close scrutiny reveals 3bp of microhomology at the distal junction - the resulting ambiguity in annotation may explain why the coordinates appear to be 1bp out compared to the MantelINV call. Genomic positions are based on GRCh38.



## GLI3/FBN1 inversions - supplemental



**Figure S14:** DUP-INV-DUP rearrangement in a 100KGP family with non-Marfan phenotype. A) Schematic diagram showing relative positions of 78kb and 238kb inversions and duplication calls with respect to *FBN1*. B) IGV screenshot showing read alignments supporting both junctions internal to *FBN1* in introns 47 and 56. C) Two possible configurations can explain the pattern of split-reads but neither are predicted to impact on dosage of full functional copies of *FBN1*.

## GLI3/FBN1 inversions - supplemental

## References

1. Pagnamenta AT. Bone Research Society 2021 Abstracts - "Experiences of a virtual multidisciplinary team meeting process to review unsolved musculoskeletal families from the 100 K Genomes Project". *JBMR Plus* 2021 doi: <https://doi.org/10.1002/jbm4.10552>
2. Pagnamenta AT, Diaz-Gonzalez F, Banos-Pinero B, et al. Variable skeletal phenotypes associated with biallelic variants in PRKG2. *J Med Genet* 2021 doi: 10.1136/jmedgenet-2021-108027 [published Online First: 2021/11/17]
3. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* 2019;51(11):1560-65. doi: 10.1038/s41588-019-0528-2 [published Online First: 2019/11/05]
4. Yu J, Szabo A, Pagnamenta AT, et al. SVRare: discovering disease-causing structural variants in the 100K Genomes Project. *medRxiv* 2022 doi: <https://doi.org/10.1101/2021.10.15.21265069>
5. Hyder Z, Calpena E, Pei Y, et al. Evaluating the performance of a clinical genome sequencing program for diagnosis of rare genetic disease, seen through the lens of craniosynostosis. *Genet Med* 2021;23(12):2360-68. doi: 10.1038/s41436-021-01297-5 [published Online First: 2021/08/26]
6. Mortier GR, Cohn DH, Cormier-Daire V, et al. Nosology and classification of genetic skeletal disorders: 2019 revision. *Am J Med Genet A* 2019;179(12):2393-419. doi: 10.1002/ajmg.a.61366 [published Online First: 2019/10/22]