Uncovering the burden of hidden ciliopathies in the 100,000 Genomes Project: a reverse phenotyping approach

Sunayna Best[1,2], Jing Yu[3], Jenny Lord[4,5], Matthew Roche[6], Christopher M. Watson[1,7], Roel P. J. Bevers[8], Alex Stuckey[8], Savita Madhusudhan[9], Rosalyn Jewell[2], Sanjay M. Sisodiya[10,11], Siying Lin[12,13], Stephen Turner[12], Hannah Robinson[13], Joseph Leslie[14], Emma Baple[14,15], Genomics England Research Consortium, Carmel Toomes[1], Chris F Inglehearn[1], Gabrielle Wheway[4,5], Colin A. Johnson[1]

1: Division of Molecular Medicine, Leeds Institute of Medical Research at St. James's, University of Leeds, St. James's University Hospital, Leeds, LS9 7TF, UK

2: Yorkshire Regional Genetics Service, Leeds, LS9 7TF, UK

3: Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK

4: University Hospital Southampton NHS Foundation Trust, Southampton, SO16 6YD, UK

5: Faculty of Medicine, Human Development and Health, University of Southampton, Southampton, SO17 1BJ, UK

6: Windsor House Group Practice, Leeds, LS27 9NB, UK

7: North East and Yorkshire Genomic Laboratory Hub, Central Lab, St. James's University Hospital, Leeds, LS9 7TF, UK

8: Genomics England, Queen Mary University of London, Dawson Hall, London, EC1M 6BQ, UK

9: St. Paul's Eye Unit, Royal Liverpool University Hospital, Prescot Street, Liverpool, L7 8XP, UK

10: University College London (UCL) Queen Square Institute of Neurology, Queen Square, London, WC1N 3BG, UK

11: Chalfont Centre for Epilepsy, Chalfont St Peter, SL9 0RJ, UK

12: Department of Ophthalmology, Torbay Hospital, Torbay and South Devon NHS Foundation Trust, Lowes Bridge, Torquay, TQ2 7AA

13: Exeter Genomics Laboratory, Royal Devon and Exeter NHS Foundation Trust, Barrack Road, Exeter, EX2 5DW, UK

14: RILD Wellcome Wolfson Centre, University of Exeter Medical School, Royal Devon & Exeter NHS Foundation Trust, Barrack Road, Exeter, EX2 5AX, UK
15: Peninsula Clinical Genetics Service, Royal Devon and Exeter NHS Foundation Trust, Gladstone Rd, Exeter EX1 2ED, UK

Corresponding authors:
Prof Colin A Johnson (ORCID: 0000-0002-2979-8234)
Email: c.johnson@leeds.ac.uk
telephone: +44 (0)113 343 8443

Dr. Gabrielle Wheway
email: G.Wheway@soton.ac.uk
telephone: +44 (0)23 8120 5029

Conflict of Interest

Disclosure: R.P.J.B and A.S are employed by Genomics England Ltd., UK. GW is employed by Illumina Inc. The other authors declare no conflict of interest.

Abstract

Background

The 100,000 Genomes Project (100K) recruited NHS patients with eligible rare diseases and cancer between 2016 and 2018. PanelApp virtual gene panels were applied to whole genome sequencing data according to Human Phenotyping Ontology (HPO) terms entered by recruiting clinicians to guide focussed analysis.

Methods

We developed a reverse phenotyping strategy to identify 100K participants with pathogenic variants in nine prioritised disease genes (*BBS1, BBS10, ALMS1, OFD1, DYNC2H1, WDR34, NPHP1, TMEM67, CEP290*), representative of the full phenotypic spectrum of multi-systemic primary ciliopathies. We mapped genotype data "backwards" onto available clinical data to assess potential matches against phenotypes. Participants with novel molecular diagnoses and key clinical features compatible with the identified disease gene were reported to recruiting clinicians.

Results

We identified 62 reportable molecular diagnoses with variants in these nine ciliopathy genes. Forty-four have been reported by 100K, five were previously unreported and 13 are new diagnoses. We identified 11 participants with un-reportable, novel molecular diagnoses, who lacked key clinical features to justify reporting to recruiting clinicians. Two participants had likely pathogenic structural variants and one a deep intronic predicted splice variant. These variants would not be prioritised for review by standard 100K diagnostic pipelines.

Conclusion

Reverse phenotyping improves the rate of successful molecular diagnosis for unsolved 100K participants with primary ciliopathies. Previous analyses likely missed these diagnoses because incomplete HPO term entry led to incorrect gene panel choice, meaning that pathogenic variants were not prioritised. Better phenotyping data is therefore essential for accurate variant interpretation and improved patient benefit.

What is already known on this topic

Whole genome sequencing and targeted gene-panel analysis have improved molecular diagnosis rates for patients with multi-systemic ciliopathies.

What this study adds

Reverse phenotyping from 100,00 Genomes Project data has identified 62 reportable molecular diagnoses with variants in nine prioritised ciliopathy genes, of which 18 are new diagnoses not reported by Genomics England (GEL). Furthermore, we identified 11 un-reportable molecular diagnoses in these genes, but these lacked adequate clinical data to justify returning the findings to recruiting clinicians.

How this study might affect research, practice or policy

Reverse phenotyping can improve molecular diagnosis rates from large-scale genomic projects. Comprehensive phenotypic data is essential to facilitate accurate variant interpretation.

Introduction

The 100,000 Genomes Project (100K) is a combined diagnostic and research initiative managed by Genomics England Ltd (GEL). It aimed to sequence 100,000 genomes from 70,000 participants seen within the United Kingdom (UK) National Health Service (NHS) with either selected rare diseases or cancers, the latter allowing comparison of matched germline and somatic tumour genomes [1,2]. To take part in 100K, participants consented to receive a result "relevant to the explanation, main diagnosis or treatment of the disease for which the patient was selected for testing" (the "pertinent finding"), if identified [3]. Furthermore, they consented to allow access to their fully anonymised genome sequence data and phenotype information for approved academic and commercial researchers. Short-read genome sequencing was performed using Illumina "TruSeq" library preparation kits for read lengths 100bp and 125bp (Illumina HiSeq 2500 instruments), or 150bp reads (HiSeq X). These generated a mean read depth of 32× (range, 27 to 54) and a depth greater than 15× for at least 95% of the reference human genome (2). In the Main Programme Data Release 12 (5th June, 2021) used in this study, data was available for 88,844 individuals: 71,597 in the rare diseases arm (33,208 probands and 33,388 relatives) and 17,247 in the cancer arm.

Large-scale genomic studies such as the 100K offer the opportunity to perform reverse phenotyping for genes of interest. In traditional forward genetics, observation of clinical features prompts differential diagnoses and the subsequent evaluation of genes with potentially pathogenic variants (phenotype-to-genotype model). In reverse phenotyping, the search begins with the identification of potentially pathogenic variants, which are then mapped in a reverse strategy against the key clinical features of patients in order to guide phenotyping. Patients with potential causative variants in the selected genes are assessed to see if their clinical features match the associated disease phenotype and inheritance pattern reported in the medical literature (genotype-to-phenotype model).

Reverse phenotyping strategies have been especially successful for diseases characterised by high heterogeneity and complex phenotypes. For example, reverse phenotyping is helping to uncover the genetic architecture of pulmonary arterial hypertension (PAH) [4]. Reverse phenotyping allowed diagnosis of 18/64 previously unsolved patients with steroid-resistant

nephrotic syndrome through analysis of 298 causative genes after whole exome sequencing (WES). This was followed by multi-disciplinary team discussion and recommended additional examinations to detect previously overlooked signs or symptoms of the syndromic genetic disorder that was guided by knowledge of the identified pathogenic variants [5]. Reverse phenotyping also provides an opportunity to extend or refine the phenotype for disease-associated genes, as demonstrated for a family with an *INPP5E*-related ciliopathy [6].

Ciliopathies are a group of rare inherited disorders caused by abnormalities of structure or function of primary cilia (the 'cell's antenna') [7] or motile cilia (organelles responsible for the movement of fluid over the surface of cells) [8 9]. Ciliopathy syndromes present as a clinical spectrum, ranging from relatively common single-system disorders such as retinal or renal ciliopathies, through to rare, complex, multi-system syndromes. There is considerable phenotypic and genetic heterogeneity between the >35 reported ciliopathy syndromes [9 10]. Common, shared clinical features include renal malformations and/or renal dysfunction, retinal dystrophy, developmental delay, intellectual disability, cerebellar abnormalities, obesity and skeletal abnormalities [11]. Collectively, ciliopathies are thought to affect up to 1 in 2000 people based on three common frequent clinical features: renal cysts (1 in 500 adults), retinal degeneration (1 in 3000), and polydactyly (1 in 500) [12]. Multi-systemic ciliopathies can be grouped into metabolic/obesity ciliopathies, neurodevelopmental ciliopathies, and skeletal ciliopathies. The variety in systems involvement reflects the critical role of cilia in development and health [2].

We recently published a study determining a research molecular diagnosis for n=43/83 (51.8%) of probands recruited under primary ciliopathy categories by GEL, comprising the "Congenital Malformations caused by Ciliopathies" cohort (13). We noted that a high proportion of diagnoses were caused by variants in non-ciliopathy disease genes (n=19/43, 44.2%). We hypothesised that this reflects difficulties in the clinical recognition of ciliopathies, as well as practical challenges in recruiting participants to 100K under appropriate rare disease domains. It is therefore reasonable to assume that there are also "hidden" patients with ciliopathies recruited to alternative categories.

Methods

In order to improve the rate of successful molecular diagnosis for unsolved 100K participants with known or suspected ciliopathies, we developed a reverse phenotyping strategy for selected exemplar genes that are most frequently mutated as a cause of primary multi-systemic ciliopathies.

Selection of common multi-systemic ciliopathy genes to assess

A literature review was undertaken to determine the most common genetic causes of multi-systemic primary ciliopathies: Bardet-Biedl syndrome (BBS) and Alström Syndrome (ALMS) (metabolic/obesity ciliopathies); Joubert syndrome (JBTS), Meckel Gruber syndrome (MKS) and oral-facial-digital syndrome (OFD) (neurodevelopmental ciliopathies); the skeletal ciliopathy Jeune Asphyxiating Thoracic Dystrophy (JATD); and nephronophthisis (isolated or syndromic renal ciliopathy) [2]. Disease genes causative of $\geq$ 10% of the total syndrome burden were selected for inclusion in the reverse phenotyping analysis and are summarised alongside referenced literature (**Supplementary Table 1)**. Where disease genes are known to cause multiple ciliopathy syndromes, all associated conditions are included in the table. On this basis, nine disease genes were selected as exemplars that span the extensive phenotypic range of primary multi-systemic ciliopathies: *BBS1, BBS10, ALMS1, OFD1, DYNC2H1, WDR34, NPHP1, TMEM67* and *CEP290*. All have autosomal recessive inheritance except *OFD1* which is associated with X-linked dominant orofaciodigital syndrome type 1 (OFD-1) and X-linked recessive JBTS [13]. Almost all individuals with OFD-1 are female; the few affected males are reported to be malformed fetuses delivered by an affected female.

Identification of solved participants with causative variants in representative ciliopathy disease genes

All analysis on the GEL datasets were performed within a secure workspace called the "Research Environment". Clinical and participant data was integrated and analysed using "LabKey" data management software. Previously reported diagnoses were identified using data in the NHS Genomics Medical Centres (GMC) "Exit Questionnaire". The Exit Questionnaire is completed by the clinicians at the GMC for each closed case, and summarizes the extent to which a participant's diagnosis can be explained by the combined variants

reported to the GMC from GEL and clinical interpretation providers. Data in Exit Questionnaires was filtered for reports containing variants in the nine ciliopathy disease genes, where the "case solved family" was annotated as "yes" (solved) or "partially" (partially solved).

Selection of key clinical terms associated with selected ciliopathy genes

A literature search of review articles prioritized the key clinical terms for each of the nine selected ciliopathy genes. This assessed the potential match against phenotype and justification for reporting new molecular findings. Approved researchers submit a "Researcher Identified Diagnosis" (RID) form using the secure GEL "Airlock" system. This is then sent to the participant's recruiting clinician for consideration of the fit to phenotype and the interpretation of variant pathogenicity, followed by decisions about whether the finding should be reported back to the participant. Usually, such cases are discussed at multi-disciplinary team (MDT) meetings involving clinical scientists, researchers and clinicians. Variants classed as likely pathogenic or pathogenic and felt to be a good clinical match for phenotype, must be molecularly confirmed and formally reported by an NHS-accredited diagnostic laboratory before being fed back to the participant by the clinician responsible for their care [3]. Decisions about feedback of variants of uncertain clinical significance (VUS) to participants are the responsibility of individual clinicians following MDT discussion, but are usually not fed back.

The rationale for selection of key features is presented in **Table 1**, supported by key references from the literature. To allow easier categorisation and to protect participant anonymity, they are grouped into 11 body systems. Without specific participant consent for research studies, we are unable to present clinical features that would potentially identify individuals to within five participants in 100K [3]. Major features (M) are those present in >50% of affected individuals and/or listed as major diagnostic or characteristic features in the cited literature. Minor features (m) are those present in <50% of affected individuals and/or listed as minor diagnostic features. The EMBL-EBI Ontology Lookup Service was used to supplement linked Human Phenotyping Ontology (HPO) terms for each key clinical term, to facilitate capture of a wider selection of appropriate HPO terms that were entered by recruiting

clinicians (available from https://www.ebi.ac.uk/ols/index). The list of acceptable linked HPO terms is available in **Supplementary Table 2**.

Development of a research diagnostic workflow to identify new diagnoses

The full diagnostic workflow developed, from extraction through to reporting of variants, is represented in **Figure 1.**

Steps 1 and 2: Single nucleotide variant (SNV) filtering and prioritisation

The script 'Gene-Variant Workflow' (available from https://research-help.genomicsengland.co.uk/display/GERE/Gene-Variant+Workflow) was used to extract all variants in the nine genes in the 100K dataset from Illumina variant call format (VCF) files, aggregate them together and annotate them using the Ensembl Variant Effect Predictor (VEP) [14]. This includes all intronic and exonic variants within the specified gene region. A custom Python script called filter_gene_variant_workflow.py (available from https://github.com/sunaynabest/filter_100K_gene_variant_workflow) was used to exclude common variants using the following criteria: 100K major allele frequency (MAF) $\geq$ 0.002; gnomAD allele frequency (AF) $\geq$ 0.002) [15]; and variants called in non-canonical transcripts. The allele frequency threshold of 0.002 was calculated using the ImperialCardioGenetics frequency filter calculator (available from https://cardiodb.org/allelefrequencyapp/) [16], as recommended by the Association for Clinical Genomic Science (ACGS) Best Practice Guidelines [17]. Parameters were set as follows: biallelic inheritance, prevalence 1 in 500, allelic heterogeneity 0.1, genetic heterogeneity 0.2, penetrance 1, confidence 0.95, reference population size 121,412 (based on the Exome Aggregation Consortium cohort).

Finally, prioritised sub-lists of SNVs were extracted using filter_gene_variant_workflow.py as follows: i) ClinVar pathogenic (variants annotated by ClinVar as "pathogenic" or "likely_pathogenic") [18]; ii) high impact (variants annotated by VEP as "high impact" (stop_gained, stop_lost, start_lost, splice_acceptor_variant, splice_donor_variant, frameshift_variant, transcript_ablation, transcript_amplification) [14]; iii) SIFT deleterious missenses (missense variants predicted "deleterious" by the *in silico* prediction tool SIFT) [19]. Additional *in silico* missense variant predictions were obtained via the Ensembl VEP web

interface (available from https://www.ensembl.org/Tools/VEP) from Combined Annotation Dependent Depletion (CADD) [20] and PolyPhen-2 [21].

Step 3: SVRare script to prioritise potentially pathogenic structural variants (SVs)

Heterozygous variants in the nine selected genes in either the "ClinVar pathogenic" or "high impact" SNV sub-lists were then analyzed by the SVRare script (29). This utilises a database of 554,060 SVs called by Manta [22] and Canvas [23] aggregated from 71,408 participants in the rare-disease arm of 100K. Common SVs (≥ 10 database calls) were excluded, and the remaining rare SVs that overlapped coding regions of the selected genes were extracted and analysed manually. BAM files for prioritised SVs were inspected in the Integrative Genomics Browser (IGV) [24]. SVs were considered potentially causative if present in >30% of reads. Participants with heterozygous variants identified as "deleterious missense" by SIFT were excluded from further manual analysis by SVRare because of the very high number of such variants and likelihood that they would be classified as VUS. **Supplementary Table 4** summarizes the numbers of SIFT deleterious missense variant calls in each gene; for example, there are 810 calls in *ALMS1* alone.

Step 4: SpliceAI script to prioritise potentially pathogenic splice defects

All rare variants called by the Gene-variant Workflow script in the nine representative ciliopathy disease genes (100K MAF ≤ 0.002; gnomAD AF ≤ 0.002) were run through SpliceAI prediction software with an additional custom Python script ('find_variants_by_gene_and_SpliceAI_score.py'; available at https://github.com/JLord86/Extract_variants). Variants predicted to affect splicing according to the recommended cut-off (SpliceAI delta scores (DS) > 0.5) were extracted and analysed manually [25]. Variants previously annotated by ClinVar as "benign" were excluded.

Step 5: Search for molecular diagnoses amongst prioritised variants

All prioritised variant lists were manually analysed for each gene: these comprised ClinVar pathogenic, high impact and SIFT deleterious missense SNV, SVRare and SpliceAI prioritised variant lists. For recessive genes (all except *OFD1*), homozygous or compound heterozygous

variants were pursued. Heterozygous variants called in female participants and hemizygous variants called in male participants were pursued for X-linked *OFD1*.

Step 6: Link to clinical data and reverse phenotyping

The Gene-Variant Workflow output files contain "plate key" identifiers (IDs; unique identifiers used by GEL for DNA sample tracking and logistics) for all participants in whom each variant was called. These unique IDs for participant samples were used to obtain participant data via LabKey, including GMC exit questionnaires reporting outcomes and participant status. Participants were excluded if recruited as unaffected relatives or "solved" or "partially solved" with variants in alternative genes. For remaining participants (all unsolved probands or affected relatives), parental data was analyzed where available, to determine variant segregation. HPO terms entered at the time of recruitment were also extracted. Further linked clinical data was obtained using the GEL user interface "Participant Explorer". This links to the source data in LabKey to identify participants with particular clinical phenotypes, determine longitudinal phenotypic and clinical data for any participant and allow comparison between multiple participants. From these, the number of key clinical features related to the identified ciliopathy gene was recorded for each participant, as well as the bodily system(s) involved.

Step 7: Decision on reporting of novel molecular diagnoses

We reasoned that the presence of at least one major key clinical feature that was compatible with the implicated gene would be sufficient to report any newly identified potential molecular diagnoses to recruiting clinicians. If no major key clinical features were present, we were unable to justify reporting because they could not be considered a potential match for patients' clinical features, the so-called "pertinent findings".

Step 8: ACMG classification and assignment of diagnostic confidence categories for reportable diagnoses

Variant clinical interpretation was reviewed using ACMG/AMP guidelines [26] and each variant of interest amongst participants with reportable diagnoses was assigned an ACMG

pathogenicity score [17]. Phenotype specificity is a key factor in variant interpretation, so only those deemed potentially pertinent findings, in the presence of at least one major key feature and therefore reportable, underwent variant interpretation and diagnostic confidence scoring. Diagnostic confidence categories were assigned as "confident", "probable" or "possible" based on the assigned ACMG variant classifications (**Figure 1**). A "confident" diagnosis required two pathogenic or likely pathogenic variants in genes with recessive inheritance, or one pathogenic or likely pathogenic variant in *OFD1*. A "probable" diagnosis required one pathogenic/likely pathogenic and one VUS in genes with recessive inheritance; no "probable" classification was possible for *OFD1* variants. A "possible" diagnosis was assigned in the presence of two VUSs in recessive genes or one VUS in *OFD1*.

We exported anonymized data for publication through the Airlock system, after review by the GEL Airlock Review Committee. We present only information about the body systems with key features for each participant rather than specific HPO terms, in order to protect participant anonymity.

Results

100K participants previously solved with causative variants in representative ciliopathy disease genes

Forty-four participants have previously been reported to have "solved" or "partially solved" molecular diagnoses in GMC exit questionnaires with variants in the nine representative ciliopathy disease genes **(Supplementary Table 3)**. Seven of these reported cases overlap with participants described in "Congenital Malformations caused by Ciliopathies" cohort analyses [27]. Interestingly, male participant #32 was reported "solved" with a pathogenic hemizygous *OFD1* frameshift variant in exon 20/23 (NM_003611.3:c.2680_2681del, NP_003602.1:p.(Glu894ArgfsTer6)). Participant #32 was recruited to the "rod-cone dystrophy" category with an apparently milder non-syndromic form of retinal dystrophy that was only identified in late adulthood **(Supplementary Table 3)**. Further clinical information from the recruiting clinicians revealed that the participant had a rod-cone dystrophy that lacked bone spicules typical for retinitis pigmentosa but was similar to Bardet-Biedl syndrome (**Figure 2A-B**). Participant #32 also had intellectual disability, truncal obesity, evidence of renal failure, short fingers, and chronic respiratory disease with mild bronchiectasis ("signet

ring" signs on CT scan of the chest; **Figure 2C**). These are clinical features consistent with a syndromic ciliopathy, and we are not aware of any previous reports of males with hemizygous *OFD1* variants having this combination of features.

Molecular details for two reported variants are incomplete, described as a heterozygous "large delins" in *ALMS1* (participant #6) and a "whole gene deletion" of *NPHP1* (participant #33). Data is also incomplete for participant #43, reported solved with a single heterozygous variant, classified as a VUS, in the recessive disease gene *CEP290.*

New reportable diagnoses identified through the reverse phenotyping research diagnostic workflow

We prioritized a total number of 3666 variants from the SNV, SV and SpliceAI outputs **(Supplementary Table 4)** through our research diagnostic workflow; 30 variants led to potential reportable diagnoses in 18 previously unsolved participants through reverse phenotyping **(Table 2)**. However, upon further investigation, n=5/18 participants (#45, #47, #48, #50 and #51) had causative variants that were already included in their GMC Exit Questionnaires, but had reporting outcomes annotated as "unknown" or without listing the ciliopathy disease genes of interest. Although these outcomes may be due to inadvertent coding errors, we did not include the data from these participants for further analysis. Our workflow therefore identified a total of n=13/18 participants with new reportable diagnoses.

Identification of reportable SVs

Two participants have been identified with new potentially causative SVs through the SVRare script **(Figure 3).** Participant #45 had a maternally inherited, 116969bp chr2 inversion and a 63550bp gain (identified using Manta and Canvas, respectively), both including coding regions of *ALMS1*. After a careful inspection of the IGV plot, we also observed a monoallelic, complex SV in the *ALSM1* gene spanning from chr2: g.73424245 to chr2: g.73544334 (GRCh38). We interpreted this as a paired-duplication inversion **(Figure 3A-B)**. Ideally, this would be confirmed experimentally; we have contacted the recruiting clinician about performing these studies but no response has been received. Participant #45 also has a paternally inherited, known pathogenic *ALMS1* frameshift variant (NM_015120.4:c.10775del,

NP_055935.4:p.Thr3592LysfsTer6). Therefore, segregation analysis is consistent with autosomal recessive inheritance as expected. Participant #45 was recruited to the cone dysfunction category and has one *ALMS1* key feature involving the ophthalmic system that allowed this research finding to be reported to the recruiting clinician.

Participant #70 had a maternally inherited, 56371bp chromosome 11 deletion (identified by Canvas), including the terminal four exons of *DYNC2H1* **(Figure 3C)**. This individual also has a ClinVar "likely pathogenic" paternally-inherited *DYNC2H1* synonymous variant (NM_001377.3: c.11049G>A, NP_001368.2: p.Pro3683=). This variant is predicted to cause a splice acceptor loss by SpliceAI (DS_AL 0.51). No clinical detail is provided with the ClinVar entry (from the Rare Disease Group, Karolinska Institutet), but the "likely pathogenic" listing in association with Jeune syndrome provides some confidence in this assessment of pathogenicity. Participant #70, recruited to the proteinuric renal disease category, has two Jeune syndrome key features from the renal and skeletal systems, allowing this research finding to be reported to the recruiting clinician. Furthermore, the participant's affected sibling, also recruited to 100K with three Jeune syndrome clinical key features from the renal and skeletal systems, was found to have the same two variants, strengthening the confidence in the diagnosis.

Identification of reportable non-canonical splice defects

One new homozygous *CEP290* intronic variant has been identified by using our SpliceAI script, predicted to cause a splice acceptor gain (SpliceAI DS_AG 0.64) (NM_025114.4:c.6011+874G>T) and gain of a potential splice acceptor site (Alamut screenshot; **Figure 3D**). This variant was identified in participant #49, recruited to the cystic kidney disease category. The proband's father is heterozygous for the variant, but there is no maternal sample available in 100K. The recruiting clinician has been contacted and relevant tissues (blood, urinary renal epithelial cells) requested to perform functional splicing assays, but no response has been received. Therefore, the variant has been called a VUS, allowing classification of only a "possible" diagnosis to be made.

Novel un-reportable diagnoses identified through research workflow

Eleven participants have un-reportable, novel diagnoses in the nine ciliopathy disease genes **(Table 3).** These participants have no major key clinical features amongst their entered HPO terms, or identifiable amongst the additional clinical data available on Participant Explorer, that can justify reporting to recruiting clinicians as potentially pertinent clinical findings. Four of these 11 have novel missense variants, which can only be classified as VUSs. The other seven (#60, #61, #64, #65, #71, #72, #73) have at least one more definitively damaging variant, including high impact frameshifts, stop gains, splice acceptors and ClinVar pathogenic missenses.

Discussion

Reportable diagnoses

We have used a reverse phenotyping strategy to identify 62 reportable molecular diagnoses with variants in nine prioritised, multi-systemic ciliopathy genes (*BBS1, BBS10, ALMS1, OFD1, DYNC2H1, WDR34, NPHP1, TMEM67, CEP290).* The nine genes chosen were representative exemplars that, from the literature review, span the extensive phenotypic range of ciliopathies. The addition of other ciliopathy genes (such as *CPLANE1* for JBTS) would, of course, further increase diagnostic yield. Forty-four have been previously reported by 100K in GMC Exit Questionnaires, five were previously unreported and 13 represent new diagnoses that are compatible with the entered clinical features for unsolved participants (**Table 2**). Based on ACMG classifications of underlying variants, six are classified as confident diagnoses, two as probable diagnoses and ten as only possible diagnoses. In summary, 14 molecular diagnoses are in *ALMS1*, 13 in *BBS1*, two in *BBS10*, 16 in *CEP290*, three in *DYNC2H1*, seven in *OFD1*, four in *NPHP1* and three in *TMEM67*. No molecular diagnoses have been made in *WDR34.* These ciliopathy findings fit with what has previously been reported for reverse phenotyping studies; namely, that this approach proves particularly useful in conditions with high genetic heterogeneity and/or complex phenotypes [4-6].

We have reported VUS results to recruiting clinicians in this project by using RID forms submitted through the secure GEL Airlock. The ACMG advises that VUS results cannot be used in clinical decision-making [26]. Not only does this apply to the index patient, but also to cascade

testing of other family members and to prenatal testing. If reported to patients, VUS can cause significant anxiety and make decision-making challenging [28 29]. We do not anticipate that VUS results identified through this study will be immediately reported back to patients by recruiting clinicians, but there is a high probability that at least some are the correct molecular diagnosis. Therefore, we believe it is important to report them from the research setting for current and future consideration, especially with the emergence of improved functional variant interpretation tools. The problem lies is the lack of available lines of evidence to perform definitive variant classification, especially for missense and splice variants. The ACMG advises that "efforts to resolve the classification of the variant as pathogenic or benign should be undertaken" when VUS are identified [26]. Currently, functional work to provide additional "strong" evidence is largely limited to the research setting, done on a case-by-case basis where resources are available and interested researchers are involved. Improved variant sharing will also facilitate better variant classification because the recurrent identification of potential disease alleles amongst individuals with convincing shared phenotypes adds weight to the assessment of variant pathogenicity.

Alström syndrome is one of the rarer ciliopathies, with an estimated prevalence of 1:100,000 to 1:1,000,000 and around 950 affected individuals reported worldwide [30]. Biallelic *ALMS1* variants have recently also been published as rare causes of non-syndromic retinal dystrophy and cardiomyopathy **(Supplementary Table 1)**. The identification of 14 patients with biallelic, predicted pathogenic *ALMS1* variants is therefore higher than anticipated and may reflect under-recognition of this disease gene in the clinical setting. We expected to find a higher rate of *TMEM67* diagnoses than the three identified, given that *TMEM67* is the leading cause of JBTS and MKS, and is also associated with NPHP and COACH syndrome **(Supplementary Table 1)**. Potentially, given the greater disease burden and therefore familiarity with *TMEM67,* more straightforward *TMEM67* diagnoses may have been identified by mainstream testing, preventing the need for those participants to be enrolled into 100K. However, this explanation may not be true for other genes because all 12 of the GEL reported *BBS1* cases had at least one copy of the founder missense variant NM_024649.5:c.1169T>G, NP_078925.3:p.(Met390Arg). This is known to be the most frequent cause of BBS [31], and it would be expected to be identified by routine testing.

To illustrate the challenge of diagnosing biallelic *BBS1* variants, even when one copy is the common founder missense variant NM_024649.5:c.1169T>G, NP_078925.3:p.(Met390Arg), we further reviewed the "Congenital Malformations caused by Ciliopathies" (CMC) cohort, as described previously (13), for potential compound heterozygous *BBS1* variants. Through manual inspection of aligned sequence reads in IGV, we identified a soft-clipped read signature in exon 13 in CMC cohort participant #59 (**Figure 3E**) that was consistent with a recently described mobile SVA F family element insertion of size 2.4 kb (37). Analysis of parental alignments supported the variant being in *trans* with the maternally inherited c.1169T>C missense mutation (**Figure 3E**). A duplex PCR screening assay (**Supplementary Data 1**) and sequencing confirmed the presence of the mobile element insertion in the proband and their father (**Figure 3E**). This case further demonstrates that reanalysis of 100K data improves diagnostic yield, and allows refinement (**Supplementary Data 1**) of an existing diagnostic screening strategy (37) that allows interpretation of unusual alignment profiles in short-read sequencing datasets.

Sources of additional diagnoses from the reverse phenotyping research diagnostic workflow

The Genomics England Rare Disease Tiering Process includes an automated variant triaging algorithm to classify variants on virtual PanelApp panel(s) selected according to entered HPO terms into a series of "Tiered" categories, which have been described previously [2 32]. Tiered variants are primarily limited to those variants affecting coding sequences, and splice donor or acceptor sites. The standard 100K pipeline requires diagnostic labs to analyse variants that are triaged into Tier one or two. Tier three variants (rare coding SNVs in genes not included in the selected panel or panels) and un-tiered variants are not routinely analysed in the diagnostic setting. The selection of incorrect panels that prevents appropriate tiering of causative variants, and the fact that certain types of variant are not routinely tiered, will therefore both contribute to missed diagnoses. Furthermore, inaccurate or incomplete HPO term entries at the time of recruitment will lead to inappropriate virtual gene panel selections that will not allow the analysis of the correct causative disease gene. These problems of missed diagnoses for both the present reverse phenotyping study and our previous analysis of the "Congenital Malformations caused by Ciliopathies" cohort (13), suggests that a change

in protocol should be considered. This would permit further gene panel selection in the absence of good phenotyping data, or when the answer is not found from the first panel(s) applied.

SVs and single heterozygous SNVs in recessive disease genes are not routinely tiered, even when the genes are on the panel(s) applied. Filtering of all variants in our selected genes independent of the GEL tiering system, followed by independent annotation and analysis, has allowed us to identify SNVs most likely to be pathogenic, even when they are a single hit in a recessive disease gene. If the second variant in the same gene is difficult to find, for example if it is an SV or intronic variant, then their identification in our pipeline could improve diagnostic yield. In particular, the introduction of the SVRare script (29), permitting exclusion of SV calls from analysis if they appear in more than ten 100K participants, has facilitated diagnosis of two previously unsolved participants (#45 and #70) with un-tiered, likely pathogenic SVs. SVRare provides a fast and systematic approach to SV analysis, which will be invaluable for future genomic studies. All 100K participants have SV.vcf files available in the Research Environment, called using the Manta and Canvas pipelines [22 23]. To date, strategies to filter the huge number of SVs from these outputs, most of which are common and benign, have been limited. Alongside manual IGV inspection, the SVRare pipeline also allowed more accurate definition of the complex *ALMS1* SV found in participant #45, since it was called as both a rare inversion (Manta) and duplication (Canvas).

A further source of un-tiered, potentially pathogenic variants is our custom SpliceAI script. Currently, novel intronic variants are not routinely tiered. SpliceAI has provided one possible new diagnosis in participant #49, with the identification of a rare, homozygous intronic variant predicted to cause a *CEP290* splice acceptor gain (NM_025114.4:c.6011+874G>T, SpliceAI DS_AG 0.64; **Figure 3D**).

These sources of potentially missed causative variants shows the value of research collaborations to make the most of available genomic data. In particular, comprehensive SV and intronic variant analysis facilitates diagnoses not easily achievable through whole exome sequencing and gene-panel testing, but the standard 100K diagnostic pipelines do not yet take full advantage of these analyses.

The challenge of poor phenotyping data that prevents accurate variant interpretation

The quality of phenotyping has proven highly significant in determining the accuracy of variant interpretation in this study. At the time of recruitment to 100K, the HPO term entry for participants was frequently sparse, comprising one or two terms only, often from just one organ system. The Participant Explorer user interface can provide additional clinical data from longitudinal patient records, which summarize medical history, and timelines for in-patient and out-patient observations, treatments and procedures. However, this data is of variable quality, and clinical features are not collated in a form amenable for genotype-phenotype correlation analyses. Given the frequently sparse clinical data available, we decided to report identified molecular diagnoses amongst participants with at least one major key clinical feature. This was to maximise the number of potential new diagnoses. With the limited data and systems available, we must pass responsibility on to the recruiting clinicians to refine any phenotypic fit in light of any additional clinical data to which they have access.

Effective communication with recruiting clinicians, providing additional clinical information not entered at the time of recruitment to 100K, has proven invaluable for accurate variant interpretation. However, of the 20 researcher-identified diagnosis forms and clinical collaboration request forms submitted via the GEL Airlock in the last three months, we have only received responses from four recruiting clinicians. Participant #62, recruited under the "epilepsy plus other features" category with an "unsolved" status on their GMC exit questionnaire, illustrates the value of effective researcher-clinician collaboration. We identified a ClinVar pathogenic *CEP290* frameshift variant (NM_025114.4:c.5434_5435del, NP_079390.3:p.Glu1812LysfsTer5) and a deep intronic *CEP290* variant known to cause a strong splice-donor site and insertion of a cryptic exon (NM_025114.4:c.2991+1655A>G) [33]. Participant #62 had one *CEP290*-related key clinical feature from the ophthalmic system category (keratoconus), permitting us to report the finding. The recruiting clinician confirmed the presence of key ophthalmological features not entered during recruitment to 100K, comprising a formal diagnosis of Leber Congenital Amaurosis (bilateral keratoconus and cataracts, no detectable ERG responses to light) that was not previously specified. This

strengthened confidence that the molecular diagnosis is correct and that this participant is highly likely to have a *CEP290*-related syndromic ciliopathy. It is unclear if the neurological features reported for participant #62 (diffuse cerebellar atrophy confirmed by MRI, but no evidence of structural brain abnormalities or intellectual disability), in addition to epilepsy, are associated with syndromic ciliopathy or comprise a separate phenotype. Nevertheless, reporting the molecular diagnosis is especially important in this instance, because the *CEP290* c.2991+1655A>G variant is a target for the development of antisense oligonucleotides that may offer a personalised therapy for patients [34 35].

Reverse phenotyping facilitates expansion of ciliopathy disease-gene associations

As was previously demonstrated for a family with an *INPP5E*-related ciliopathy [6], this study widens the phenotypic spectrum of known ciliopathy disease-gene associations through reverse phenotyping. For example, male participant #32 was reported "solved" with a pathogenic hemizygous *OFD1* frameshift variant in exon 20/23 (NM_003611.3:c.2680_2681del, NP_003602.1:p.Glu894ArgfsTer6). Although participant #32 was recruited to the "rod-cone dystrophy" category with apparently non-syndromic retinal dystrophy, reverse phenotyping revealed that he had clinical features that were consistent with a syndromic ciliopathy. Truncating variants in the C-terminal end of *OFD1* (exons 20-21) have recently been associated with the motile ciliopathy primary ciliary dyskinesia (PCD) without the characteristic skeletal, neurological or renal features of other *OFD1*-related disorders [36 37]. The OFD1 protein is a component of ciliary basal bodies and centrioles, and has been shown to be essential for both primary and motile ciliogenesis [38]. Therefore, it is entirely plausible that pathogenic *OFD1* variants could cause features compatible with both motile and primary ciliopathies, therefore accounting for participant #32's full constellation of features (retinal dystrophy, renal failure and intellectual disability in keeping with primary ciliopathies, and PCD-like respiratory disease with motile ciliopathies). Further reports of patients with both motile and primary ciliopathy features that carry pathogenic *OFD1* variants would strengthen this potential broadening of associated phenotypes. It is possible that the exon 20 frameshift variant identified in participant #32 could just explain part of his phenotype, for example his PCD-like respiratory disease, in keeping with the published literature [36 37]. Conversely, retinal dystrophy may be an additional feature, as has been

reported in association with X-linked recessive JBTS caused by pathogenic *OFD1* variants in affected males [39]. We therefore suggest that individuals with a suspected *OFD1*-associated ciliopathy undergo a formal ophthalmological assessment to strengthen the diagnosis.

Un-reportable diagnoses

As well as the 18 reportable molecular diagnoses, we also identified 11 un-reportable molecular diagnoses for the nine ciliopathy disease genes **(Table 3)**. Parental sequence is not available for any of the participants with un-reportable diagnoses apart from one (#52). Lack of segregation analyses hamper accurate variant interpretation. Nevertheless, it is highly likely that some of these molecular diagnoses are correct and clinically actionable, with implications not only for the proband but also for their relatives. The inability to report these findings is likely to be driven by inaccurate HPO term entry, which is a great loss to the participants. A review of reporting guidelines, given this important observation, may prove beneficial. For example, a system could be devised that marks potential pathogenic variants of interest that then requests further clinical information, but these remain un-reportable until further, actionable data is available.

Conclusion

This study reveals the power of reverse phenotyping approaches to improve diagnosis rates for rare disease participants entered into large-scale genomic studies such as the 100K. Through the application of additional novel screening methodologies such as the SVRare suite, and with domain-specific knowledge, we have confirmed existing ciliopathy diagnoses and identified additional ones in a series of 100K participants who were not originally recruited as having a primary ciliopathy. Our findings suggest that diagnoses may be missed when screening of limited gene panels is directed by incorrect or incomplete HPO term entry, and that inaccurate phenotyping may prevent participants from accessing clinically valuable findings. We have discussed the challenges of 100K analyses more extensively in our recent commentary article and suggest potential improvements for future use of 100K data [32]. Clearly, open dialogue between researchers, clinicians and clinical scientists is essential to fully exploit the available data for patient benefit in the post-genomic era.

Data availability

Full data is available in the Genomic England Secure Research Environment. All datasets are available in the re_gecip shared folder of the GEL research environment for approved researchers. Access to our folder containing variant data (re_gecip/shared_allGeCIPs/GW_SB) can be requested from the GEL Helpdesk.

Acknowledgements

S.B acknowledges support from the Wellcome Trust 4Ward North Clinical PhD Academy (ref. 203914/Z/16/Z). J.Y acknowledges support from Retina UK (grant HMR03950). G.W acknowledges support from Wellcome Trust Seed Award (ref. 204378/Z/16/Z). C.A.J acknowledges support from Medical Research Council project grants MR/M000532/1 and MR/T017503/1. J.L is supported by a National Institute for Health Research (NIHR) Research Professorship awarded to Prof Diana Baralle (DB NIHR RP-2016-07-011). S.M.S is supported by the Epilepsy Society and the NIHR University College London Hospitals Biomedical Research Centre.

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 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. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

Author Information

Conceptualisation: S.B, C.T, C.F.I, C.A.J, G.W; Data curation: S.B, J.Y, Genomics England Research Consortium; Formal analysis: S.B, J.Y, J.L, M.R, C.M.W, S.M, R.J, C.T, C.F.I, C.A.J, G.W; Funding acquisition: S.B, J.L, C.T, C.F.I, C.A.J, G.W; Investigation: S.B, J.Y, J.L, M.R, C.M.W, S.M.S, R.J, C.T, C.F.I, C.A.J, G.W; Methodology: S.B, J.Y, J.L, M.R, C.T, C.F.I, C.A.J, G.W; Software: J.Y, J.L, M.R, R.P.J.B, A.S, Genomics England Research Consortium; Project administration: S.B, J.Y, Genomics England Research Consortium, G.W; Resources: S.B, S.M.S, S.S.L, S.T, E.B,

Genomics England Research Consortium; Supervision: C.T, C.F.I, C.A.J, G.W; Validation: S.M.S, S.S.L, S.T, E.B; Writing – original draft: S.B, G.W; Writing – review and editing: all authors.

Collaborator Information

The Genomics England Research Consortium:

Ambrose, J. C. [1]; Arumugam, P.[1]; Bevers, R.[1]; Bleda, M. [1]; Boardman-Pretty, F. [1,2]; Boustred, C. R. [1]; Brittain, H.[1]; Brown, M.A. [1]; Caulfield, M. J. [1,2]; Chan, G. C. [1]; Fowler, T. [1]; Giess A. [1]; Hamblin, A.[1]; Henderson, S.[1,2]; Hubbard, T. J. P. [1]; Jackson, R. [1]; Jones, L. J. [1,2]; Kasperaviciute, D. [1,2]; Kayikci, M. [1]; Kousathanas, A. [1]; Lahnstein, L. [1]; Leigh, S. E. A. [1]; Leong, I. U. S. [1]; Lopez, F. J. [1]; Maleady-Crowe, F. [1]; McEntagart, M.[1]; Minneci F. [1]; Moutsianas, L. [1,2]; Mueller, M. [1,2]; Murugaesu, N. [1]; Need, A. C. [1,2]; O'Donovan P. [1]; Odhams, C. A. [1]; Patch, C. [1,2]; Perez-Gil, D. [1]; Pereira, M. B.[1]; Pullinger, J. [1]; Rahim, T. [1]; Rendon, A. [1]; Rogers, T. [1]; Savage, K. [1]; Sawant, K. [1]; Scott, R. H. [1]; Siddiq, A. [1]; Sieghart, A. [1]; Smith, S. C. [1]; Sosinsky, A. [1,2]; Stuckey, A. [1]; Tanguy M. [1] ; Taylor Tavares, A. L.[1]; Thomas, E. R. A. [1,2]; Thompson, S. R. [1]; Tucci, A. [1,2]; Welland, M. J. [1]; Williams, E. [1]; Witkowska, K. [1,2]; Wood, S. M. [1,2].

1: Genomics England, London, UK

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

All collaborators contributed to data curation, formal analysis, software, project administration and resources.

Ethics Declaration

Written informed consent was obtained from all participants (or from their parent/legal guardian) in the 100,000 Genomes Project (IRAS ID 166046; REC reference 14/EE/1112). Access to the secure online Research Environment within the Genomics England Ltd (GEL) Data Embassy was provided by the GEL Access Review Committee, and research project RR185 'Study of cilia and ciliopathy genes across the 100,000 GP cohort' was registered and approved by GEL. This research study received ethical approval from University of Southampton Faculty of Medicine Ethics Committee (ERGO#54400).

References

1. 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. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ (Clinical research ed)* 2018;361:k1687. doi: 10.1136/bmj.k1687 [published Online First: 2018/04/25]
2. Wheway G, Mitchison HM, Genomics England Research C. Opportunities and Challenges for Molecular Understanding of Ciliopathies-The 100,000 Genomes Project. *Front Genet* 2019;10:127-27. doi: 10.3389/fgene.2019.00127
3. The National Genomic Research Library v5.1. 2020 01/04/2020. (accessed 14/09/2021).
4. Swietlik EM, Prapa M, Martin JM, Pandya D, Auckland K, Morrell NW, Gräf S. 'There and Back Again'-Forward Genetics and Reverse Phenotyping in Pulmonary Arterial Hypertension. *Genes (Basel)* 2020;11(12) doi: 10.3390/genes11121408 [published Online First: 2020/12/02]
5. Landini S, Mazzinghi B, Becherucci F, Allinovi M, Provenzano A, Palazzo V, Ravaglia F, Artuso R, Bosi E, Stagi S, Sansavini G, Guzzi F, Cirillo L, Vaglio A, Murer L, Peruzzi L, Pasini A, Materassi M, Roperto RM, Anders HJ, Rotondi M, Giglio SR, Romagnani P. Reverse Phenotyping after Whole-Exome Sequencing in Steroid-Resistant Nephrotic Syndrome. *Clin J Am Soc Nephrol* 2020;15(1):89-100. doi: 10.2215/cjn.06060519 [published Online First: 2019/12/14]
6. de Goede C, Yue WW, Yan G, Ariyaratnam S, Chandler KE, Downes L, Khan N, Mohan M, Lowe M, Banka S. Role of reverse phenotyping in interpretation of next generation sequencing data and a review of INPP5E related disorders. *Eur J Paediatr Neurol* 2016;20(2):286-95. doi: 10.1016/j.ejpn.2015.11.012 [published Online First: 2016/01/11]
7. Singla V, Reiter JF. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* 2006;313(5787):629-33. doi: 10.1126/science.1124534 [published Online First: 2006/08/05]
8. Oud MM, Lamers IJ, Arts HH. Ciliopathies: Genetics in Pediatric Medicine. *J Pediatr Genet* 2017;6(1):18-29. doi: 10.1055/s-0036-1593841 [published Online First: 2017/02/10]
9. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nature reviews Molecular cell biology* 2017;18(9):533-47. doi: 10.1038/nrm.2017.60 [published Online First: 2017/07/13]
10. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. *The Journal of pathology* 2017;241(2):294-309. doi: 10.1002/path.4843 [published Online First: 2016/11/20]
11. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. *Pediatric nephrology (Berlin, Germany)* 2011;26(7):1039-56. doi: 10.1007/s00467-010-1731-7 [published Online First: 2011/01/07]
12. Quinlan RJ, Tobin JL, Beales PL. Modeling ciliopathies: Primary cilia in development and disease. *Current topics in developmental biology* 2008;84:249-310. doi: 10.1016/s0070-2153(08)00605-4 [published Online First: 2009/02/03]

13. Toriello HV, Franco B, Bruel AL, Thauvin-Robinet C. Oral-Facial-Digital Syndrome Type I. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
14. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biology* 2016;17(1):122. doi: 10.1186/s13059-016-0974-4
[dataset] 15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv* 2020:531210. doi: 10.1101/531210
16. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* 2017;19(10):1151-58. doi: 10.1038/gim.2017.26 [published Online First: 2017/05/18]
17. Ellard S BE, Berry I, Forrester N, Turnbull C, Owens M, Eccles DM, Abbs S, Scott R, Deans Z, Lester T, Campbell J, Newman W, McMullan D. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020. https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf.
[dataset] 18. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* 2016;44(D1):D862-8. doi: 10.1093/nar/gkv1222 [published Online First: 2015/11/20]
19. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research* 2012;40(W1):W452-W57. doi: 10.1093/nar/gks539
20. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* 2019;47(D1):D886-d94. doi: 10.1093/nar/gky1016 [published Online First: 2018/10/30]
21. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* 2013;Chapter 7:Unit7.20. doi: 10.1002/0471142905.hg0720s76 [published Online First: 2013/01/15]
22. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline

and cancer sequencing applications. *Bioinformatics* 2016;32(8):1220-2. doi: 10.1093/bioinformatics/btv710 [published Online First: 2015/12/10]
23. Ivakhno S, Roller E, Colombo C, Tedder P, Cox AJ. Canvas SPW: calling de novo copy number variants in pedigrees. *Bioinformatics* 2018;34(3):516-18. doi: 10.1093/bioinformatics/btx618 [published Online First: 2017/10/14]
24. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 2012;14(2):178-92. doi: 10.1093/bib/bbs017
25. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 2019;176(3):535-48.e24. doi: 10.1016/j.cell.2018.12.015 [published Online First: 2019/01/22]
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17(5):405-24. doi: 10.1038/gim.2015.30 [published Online First: 2015/03/06]
27. Best S, Lord J, Roche M, Watson CM, Poulter JA, Bevers RPJ, Stuckey A, Szymanska K, Ellingford JM, Carmichael J, Brittain H, Toomes C, Inglehearn C, Johnson CA, Wheway G. Molecular diagnoses in the congenital malformations caused by ciliopathies cohort of the 100,000 Genomes Project. *J Med Genet* 2021 doi: 10.1136/jmedgenet-2021-108065 [published Online First: 20211029]
28. Han PKJ, Umstead KL, Bernhardt BA, Green RC, Joffe S, Koenig B, Krantz I, Waterston LB, Biesecker LG, Biesecker BB. A taxonomy of medical uncertainties in clinical genome sequencing. *Genet Med* 2017;19(8):918-25. doi: 10.1038/gim.2016.212 [published Online First: 2017/01/19]
29. Makhnoon S, Shirts BH, Bowen DJ. Patients' perspectives of variants of uncertain significance and strategies for uncertainty management. *Journal of genetic counseling* 2019;28(2):313-25. doi: 10.1002/jgc4.1075 [published Online First: 2019/01/13]
30. Paisey RB, Steeds R, Barrett T, Williams D, Geberhiwot T, Gunay-Aygun M. Alström Syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
31. Mykytyn K, Nishimura DY, Searby CC, Shastri M, Yen HJ, Beck JS, Braun T, Streb LM, Cornier AS, Cox GF, Fulton AB, Carmi R, Lüleci G, Chandrasekharappa SC, Collins FS, Jacobson SG, Heckenlively JR, Weleber RG, Stone EM, Sheffield VC. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat Genet* 2002;31(4):435-8. doi: 10.1038/ng935 [published Online First: 2002/07/16]
32. Best S, Inglehearn CF, Watson CM, Toomes C, Wheway G, Johnson CA. Unlocking the potential of the UK 100,000 Genomes Project-lessons learned from analysis of the "Congenital Malformations caused by Ciliopathies" cohort. *Am J Med Genet C Semin Med Genet* 2022 doi: 10.1002/ajmg.c.31965 [published Online First: 20220315]

33. den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, Zonneveld MN, Strom TM, Meitinger T, Brunner HG, Hoyng CB, van den Born LI, Rohrschneider K, Cremers FP. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet* 2006;79(3):556-61. doi: 10.1086/507318 [published Online First: 2006/08/16]
34. Dulla K, Aguila M, Lane A, Jovanovic K, Parfitt DA, Schulkens I, Chan HL, Schmidt I, Beumer W, Vorthoren L, Collin RWJ, Garanto A, Duijkers L, Brugulat-Panes A, Semo M, Vugler AA, Biasutto P, Adamson P, Cheetham ME. Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. *Mol Ther Nucleic Acids* 2018;12:730-40. doi: 10.1016/j.omtn.2018.07.010 [published Online First: 2018/08/17]
35. Duijkers L, van den Born LI, Neidhardt J, Bax NM, Pierrache LHM, Klevering BJ, Collin RWJ, Garanto A. Antisense Oligonucleotide-Based Splicing Correction in Individuals with Leber Congenital Amaurosis due to Compound Heterozygosity for the c.2991+1655A>G Mutation in CEP290. *Int J Mol Sci* 2018;19(3) doi: 10.3390/ijms19030753 [published Online First: 2018/03/10]
36. Guo Z, Chen W, Wang L, Qian L. Clinical and Genetic Spectrum of Children with Primary Ciliary Dyskinesia in China. *J Pediatr* 2020;225:157-65.e5. doi: 10.1016/j.jpeds.2020.05.052 [published Online First: 2020/06/06]
37. Bukowy-Bieryllo Z, Rabiasz A, Dabrowski M, Pogorzelski A, Wojda A, Dmenska H, Grzela K, Sroczynski J, Witt M, Zietkiewicz E. Truncating mutations in exons 20 and 21 of OFD1 can cause primary ciliary dyskinesia without associated syndromic symptoms. *J Med Genet* 2019;56(11):769-77. doi: 10.1136/jmedgenet-2018-105918 [published Online First: 2019/08/02]
38. Ferrante MI, Zullo A, Barra A, Bimonte S, Messaddeq N, Studer M, Dollé P, Franco B. Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. *Nat Genet* 2006;38(1):112-7. doi: 10.1038/ng1684 [published Online First: 2005/11/29]
39. Bruel AL, Franco B, Duffourd Y, Thevenon J, Jego L, Lopez E, Deleuze JF, Doummar D, Giles RH, Johnson CA, Huynen MA, Chevrier V, Burglen L, Morleo M, Desguerres I, Pierquin G, Doray B, Gilbert-Dussardier B, Reversade B, Steichen-Gersdorf E, Baumann C, Panigrahi I, Fargeot-Espaliat A, Dieux A, David A, Goldenberg A, Bongers E, Gaillard D, Argente J, Aral B, Gigot N, St-Onge J, Birnbaum D, Phadke SR, Cormier-Daire V, Eguether T, Pazour GJ, Herranz-Pérez V, Goldstein JS, Pasquier L, Loget P, Saunier S, Mégarbané A, Rosnet O, Leroux MR, Wallingford JB, Blacque OE, Nachury MV, Attie-Bitach T, Rivière JB, Faivre L, Thauvin-Robinet C. Fifteen years of research on oral-facial-digital syndromes: from 1 to 16 causal genes. *J Med Genet* 2017;54(6):371-80. doi: 10.1136/jmedgenet-2016-104436 [published Online First: 2017/03/16]
40. Forsyth R, Gunay-Aygun M. Bardet-Biedl Syndrome Overview. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
41. Keppler-Noreuil KM, Adam MP, Welch J, Muilenburg A, Willing MC. Clinical insights gained from eight new cases and review of reported cases with Jeune syndrome (asphyxiating thoracic dystrophy). *Am J Med Genet A* 2011;155a(5):1021-32. doi: 10.1002/ajmg.a.33892 [published Online First: 2011/04/06]

42. Stokman M, Lilien M, Knoers N. Nephronophthisis. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
43. Parisi M, Glass I. Joubert Syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
44. Hartill V, Szymanska K, Sharif SM, Wheway G, Johnson CA. Meckel-Gruber Syndrome: An Update on Diagnosis, Clinical Management, and Research Advances. *Front Pediatr* 2017;5(244)
45. Kumaran N, Pennesi ME, Yang P, Trzupek KM, Schlechter C, Moore AT, Weleber RG, Michaelides M. Leber Congenital Amaurosis / Early-Onset Severe Retinal Dystrophy Overview. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved. 1993.
46. Brown DE, Pittman JE, Leigh MW, Fordham L, Davis SD. Early lung disease in young children with primary ciliary dyskinesia. *Pediatr Pulmonol* 2008;43(5):514-6. doi: 10.1002/ppul.20792

Tables

Table 1: Key clinical features for ciliopathy syndromes associated with the nine selected ciliopathy genes of interest. Key features are grouped into 11 body systems. Clinical features marked "M" are major features (present in >50% and/or listed as major diagnostic or characteristic feature in the literature cited). Features marked with "m" are minor features (present in <50% and/or listed as a minor diagnostic feature in the literature cited).

System	Ciliopathy syndrome	BBS	ALMS	JATD	OFD-1	Nephronophthisis	JBTS	MKS	LCA / EOSRD
	Reference(s)	40	30	41	13 39	42	43	44	45
	Chosen ciliopathy gene(s) associated with syndrome	*BBS1, BBS10, TMEM67, CEP290*	*ALMS1*	*DYNC2H1, WDR34*	*OFD1*	*NPHP1* (isolated + syndromic), *TMEM67 + CEP290* (syndromic)	*TMEM67, CEP290, NPHP1, OFD1*	*TMEM67, CEP290*	*CEP290*
Ophthalmic	Retinal dystrophy	**M**	**M**	**M**		**m**[1, 2, 3]	**m**[1, 2, 3]		**M**
	Abnormality of eye movement					**m**[1, 2, 3]	**M**		**M**
	Lens opacities								**M**
	Keratoconus								**M**
Gastro-intestinal	Abnormality of the liver		**m**	**M**	**m**	**m**[1,2, 3]	**m**[1, 2, 3]	**M**	
	Abnormality of the gut	**m**		**m**					
Renal	Abnormal renal morphology / dysfunction	**M**	**M**	**M**	**M**	**M**	**m**[1, 2]	**M**	
Genito-urinary	Abnormality of the genitourinary system	**M**	**m**					**m**	
Cardiovascular	Cardiomyopathy		**M**						
	Laterality defect	**m**				**m**[1, 2]	**m**[1, 2]	**m**	
	Congenital heart disease	**m**		**m**				**m**	
	Hypertension		**m**						
Sensory	SNHL	**m**	**M**						
	Glue ear		**m**						
	Chronic otitis media		**m**		**m**				
	Abnormality of the sense of smell	**M**							
Endocrine / metabolic	Hypogonadotrophic hypogonadism	**M**	**M**						
	Glucose intolerance		**M**						
	Obesity	**M**	**M**						
	Hypertriglyceridemia		**M**						
	Thyroid abnormality	**m**	**m**					**m**	
	Polycystic ovarian syndrome	**m**			**m**				
Neurological	Intellectual disability	**M**	**m**		**M**	**m**[1, 2]	**M**		
	Neurodevelopmental delay	**M**	**m**				**M**		
	Hypotonia		**m**				**M**		
	Ataxia		**m**				**M**		

	Abnormality of brain morphology			**m**	**M**	**m**[1, 2]	**M**	**M**	
	Seizures		**m**						
	Unusual sleep patterns		**m**						
Skeletal	Polydactyly	**M**		**m**	**M**		**m**	**M**	
	Short stature			**M**					
	Narrow chest			**M**					
	Brachydactyly			**M**	**M**				
	Micromelia			**M**	**M**			**m**	
	Leg cramps		**M**						
Facial / oral	Dental abnormalities	**M**							
	Abnormal oral morphology	**M**			**M**		**m**	**m**	
	Dysmorphic facial features				**M**				
Respiratory	Abnormal pattern of respiration						**M**		
	Chronic airway infection		**m**						
	Asthma		**m**						
	Pulmonary hypoplasia							**m**	
	Cystic lung							**m**	

[1] Feature of *NPHP1* associated JBTS-plus syndrome (Senior-Loken syndrome)
[2] Feature of *CEP290* associated JBTS-plus syndrome (Senior-Loken syndrome, Joubert syndrome w/retinal disease (JS-Ret), Joubert syndrome w/renal disease (JS-Ren), COACH syndrome (cerebellar vermis hypoplasia, oligophrenia, ataxia, coloboma, and hepatic fibrosis))
[3] Feature of *TMEM67* associated JBTS-plus syndrome (COACH syndrome) (cerebellar vermis hypoplasia, oligophrenia, ataxia, coloboma, and hepatic fibrosis)

Abbreviations: BBS = Bardet Biedl syndrome, ALMS = Alström syndrome, JATD = Jeune Asphyxiating Thoracic Dystrophy, OFD-1 = Oral-facial-digital-syndrome 1, JBTS = Joubert syndrome, MKS = Meckel Gruber syndrome, LCA = Leber Congenital Amaurosis, SNHL = Sensorineural Hearing Loss, M = major clinical feature, m = minor clinical feature

Table 2: Reportable new diagnoses identified via reverse phenotyping research diagnostic workflow

Research ID	Dx confidence	Reported Sex	Recruitment category	Gene(s)	Variant zygosity	Consequence	HGVSc	HGVSp	gnomAD AF	100K MAF	SIFT	PolyPhen	CADD	PubMed	ClinVar listing	Segregation	ACMG classification	# of key features	System(s) involved
45	Conf	Ma	Cone dysfunction syndrome	*ALMS1*	Het	FS	NM_015120.4:c.10775del	NP_055935.4:p.Thr3592LysfsTer6	5.23E-05	4.77E-04				11941369, 11941370, 17594715	Path	Pat	Path	1M	M: Oph
					Het	SV	NC_000002.12:g.[73424245_73544334inv;73424245_73427355dup;73484777_73544334dup]							abs	abs	Mat	Lik_path		
47	Poss	Fe	Bardet Biedl Syndrome	*BBS10*	Hom	Mis	NM_024685.4:c.1790G>A	NP_078961.3:p.Gly597Asp	0	2.54E-05	Delet	Prob_dam		abs	abs	Bi-par	VUS	4M	M: Ren, oph, skel, endo / met
48	Conf	Ma	Rod dysfunction syndrome	*NPHP1*	Hom	Mis	NM_001128178.3:c.1027G>A	NP_000263.2:p.Gly343Arg	0.0001155	0.00034312	Delet	Prob_dam	35	10839884	Path	1 par (other unk)	Path	1M, 1m	M: Ren m: Oph
49	Poss	Fe	Cystic kidney disease	*CEP290*	Hom	Intr	NM_025114.4:c.6011+874G>T	-	0	3.81E-05				abs	abs	1 par (other unk)	VUS	1M, 1m	M: Ren m: CVS
50	Poss	Fe	Syndromic cleft lip and / or cleft palate	*OFD1*	Het	Mis	NM_003611.3:c.635G>C	NP_003602.1:p.Arg212Pro	0	1.27E-05	Delet	Poss_dam	21.2	abs	abs	De Novo	VUS	3M	M: Fac / ora (n=2), skel
51	Prob	Ma	Joubert Syndrome	*CEP290*	Het	Mis	NM_025114.4:c.104T>G	NP_079390.3:p.Val35Gly	0	1.00E-04	Delet	Prob_dam	33	abs	abs	Mat	VUS	4M	M: Oph, neu (n=3)
					Het	SG	NM_025114.4:c.5668G>T	NP_079390.3:p.Gly1890Ter	9.49E-05	2.50E-04				18414213, 26092869, 16682970, 16682973, 17564967	Path	Pat	VUS		
53	Poss	Ma	Cystic kidney disease	*ALMS1*	Het	Mis	NM_015120.4:c.8735A>G	NP_055935.4:p.Gln2912Arg	0	5.00E-05	Delet	Poss_dam	19.03	abs	abs	Unk	VUS	2M, 2m	M: Ren, endo / met m: GI, CVS
					Het	Mis	NM_015120.4:c.7412A>G	NP_055935.4:p.Asp2471Gly	0	5.00E-05	Delet	Prob_dam	25.9	abs	abs	Unk	VUS		
54	Poss	Fe	Rod-cone dystrophy	*ALMS1*	Het	Mis	NM_015120.4:c.10831A>G	NP_055935.4:p.Arg3611Gly	3.22E-05	5.00E-05	Delet	Poss_dam	23.5	abs	abs	Unk	VUS	2M, 1m	M: Oph, ren m: Resp
					Het	Mis	NM_015120.4:c.10377C>G	NP_055935.4:p.Ile3459Met	3.63E-05	5.00E-05	Delet	Poss_dam	23.1	abs	VUS	Unk	VUS		
55	Conf	Fe	Single autosomal recessive mutation in rare disease	*ALMS1*	Het	FS	NM_015120.4:c.11794del	NP_055935.4:p.Glu3932LysfsTer18	3.99E-06	1.27E-05				abs	abs	Unk	Path	4M	M: Oph, endo / met (n=2), CVS
					Het	FS	NM_015120.4:c.1735del	NP_055935.4:p.Arg579GlyfsTer17	1.61E-05	3.18E-05				26104972, 32581362, 17594715, 24462884	Path	Unk	Path		
56	Prob	Ma	Intellectual disability	*ALMS1*	Het	FS	NM_015120.4:c.10775del	NP_055935.4:p.Thr3592LysfsTer6	5.23E-05	0.00047656				11941369	Path	Unk	Path	1M, 2m	M: Sens m: CVS, neu
					Het	Mis	NM_015120.4:c.7510G>T	NP_055935.4:p.Ala2504Ser	8.93E-05	0.00019062	Delet	Prob_dam	25	abs	VUS	Unk	VUS		
57	Poss	Ma		*ALMS1*	Het	Mis	NM_015120.4:c.11429A>G	NP_055935.4:p.Tyr3810Cys	-	1.27E-05	Delet	Prob_dam	27.5	abs	abs	Mat	VUS	1M	M: Sens

			Congenital hearing impairment		Het	Mis	NM_015120.4:c.9148A>G	NP_055935.4:p.Ile3050Val	0.0002007	0.00012708	Delet	Poss_dam	24.2	abs	VUS	Pat	VUS		
58	Poss	Fe	Syndromic congenital heart disease	*BBS1*	Het	Mis	NM_024649.5:c.734C>T	NP_078925.3:p.Pro245Leu	7.16E-05	6.35E-06	Delet	Ben	23.4	abs	VUS	Unk	VUS	1M, 1m	M: Neu m: CVS
					Het	Mis	NM_024649.5:c.1313C>G	NP_078925.3:p.Thr438Arg	7.96E-05	4.45E-05	Delet	Prob_dam	25.3	abs	VUS	Mat	VUS		
62	Conf	Fe	Epilepsy plus other features	*CEP290*	Het	FS	NM_025114.4:c.5434_5435del	NP_079390.3:p.Glu1812LysfsTer5	1.14E-05	4.45E-05				abs	Path	Unk	Path	2M	M: Oph, Neu
					Het	Intr	NM_025114.4:c.2991+1655A>G	-	0	0.00040031				20301475, 17964524, 20301500, 16909394, 17564967, 17345604	Path	Unk	Path		
63	Conf	Fe	Cystic kidney disease	*CEP290*	Het	SL	NM_025114.4:c.2T>A	NP_079390.3:p.Met1?	4.07E-06	2.54E-05				abs	path/lik_path	Pat	Path	2M	M: Ren, oph
					Het	SG	NM_025114.4:c.4966G>T	NP_079390.3:p.Glu1656Ter	3.60E-05	1.59E-04				23559409, 25525159, 16909394, 20079931	path/lik_path	Mat	Path		
66	Poss	Ma	Leber Congenital Amaurosis or Early-Onset Severe Retinal Dystrophy	*CEP290*	Hom	Mis	NM_025114.4:c.182T>C	NP_079390.3:p.Met61Thr	-	2.54E-05	Delet	Poss_dam	25.6	abs	abs	Bi-par	VUS	3M	M: Oph, neu (n=2)
67	Poss	Fe	Ultra-rare undescribed monogenic disorders	*CEP290*	Hom	Mis	NM_025114.4:c.5284C>T	NP_079390.3:p.Arg1762Cys	3.78E-05	9.53E-05	Delet	Poss_dam	29.6	25741868	VUS	Unk	VUS	2M, 1m	M: Oph, neu m: GI
70	Conf	Fe	Proteinuric renal disease	*DYNC2H1*	Het	Syn	NM_001377.3:c.11049G>A	NP_001368.2:p.Pro3683=	1.21E-05	0.00014996				abs	Lik_path	Pat	Lik_path	2M	M: Ren, skel
					Het	SV	NC_000011.9:g.103445518_1035018del							abs	abs	Mat	Lik_path		
75	Poss	Ma	Distal myopathies	*TMEM67*	Het	Mis	NM_153704.6:c.2035G>C	NP_714915.3:p.Glu679Gln	0	5.00E-05	Delet	Prob_dam	26.5	abs	abs	Unk	VUS	1M	M: Ren
					Het	Mis	NM_153704.6:c.755T>C	NP_714915.3:p.Met252Thr	8.36E-05	2.00E-04	Delet	Ben	23.7	26092869, 19508969, 21866095	Path	Unk	Path		

Abbreviations: Dx = diagnostic, HGVSc = Human Genome Variation Society coding, HGVSp = Human Genome Variation Society protein, AF = allele frequency, MAF = maximum allele frequency, 100K = 100,000 Genomes Project, ACMG = American College of Medical Genetics and Genomics, Conf = confident, Prob = probable, Poss = possible, Ma = male, Fe = female, Het = heterozygous, Hom = homozygous, Hemi = hemizygous, FS = frameshift, SV = structural variant, SG = stop gain, SL = start loss, Intr = intronic, Syn = synonymous, Mis = missense, Spl A = splice acceptor, Spl Reg = splice region, Abs = absent Mat = maternal, Pat = Paternal, Bi-par = bi-parental, 1 par (other unk) = 1 parent, other unknown, Unk = unknown, Path = pathogenic, Lik_path = likely pathogenic, VUS = variant of uncertain significance, Oph = ophthalmic, Neu = neurological, Ren = renal, Skel = skeletal, Endo / met = endocrine / metabolic, Fac / ora = facial / oral, Sens = sensory, Resp = respiratory, M = major clinical feature, m = minor clinical feature

Table 3: Novel, unreportable diagnoses identified via reverse phenotyping research diagnostic workflow

Research ID	Recruitment category	Gene	Variant zygosity	Consequence	HGVSc	HGVSp	gnomAD AF	100K MAF	SIFT	PolyPhen	CADD	PubMed	ClinVar listing	Segregation	# of key features
52	Intellectual Disability	*ALMS1*	Het	Mis	NM_015120.4:c.7738A>T	NP_055935.4:p.Ile2580Phe	4.01E-06	0.00009997	Delet	Ben	22.5	abs	abs	Mat	0M, 2m
			Het	Mis	NM_015120.4:c.346C>T	NP_055935.4:p.His116Tyr	2.41E-05	0.00019994	Delet	Ben	16.05	abs	abs	Pat	
59	Hereditary Spastic Paraplegia	*BBS1*	Het	Mis	NM_024649.5:c.235G>A	NP_078925.3:p.Glu79Lys	0.000756	0.00120728	Delet	Poss_dam	23.8		VUS	Unk	0M, 0m
			Het	Mis	NM_024649.5:c.1714G>T	NP_078925.3:p.Gly572Cys	-	1.27E-05	Delet	Prob_dam	32		abs	Unk	
60	Primary Immunodeficiency	*CEP290*	Het	FS	NM_025114.4:c.6154_6161del	NP_079390.3:p.Asp2052LeufsTer17	0	1.27E-05				abs	abs	Unk	0M, 0m
			Het	FS	NM_025114.4:c.7412_7415del	NP_079390.3:p.Glu2471ValfsTer13	0	1.91E-05				abs	abs	Unk	
61	Primary Lymphoedema	*CEP290*	Het	SG	NM_025114.4:c.7048C>T	NP_079390.3:p.Gln2350Ter	1.90E-05	1.91E-05				abs	lik_path	Unk	0M, 0m
			Het	Mis	NM_025114.4:c.4063C>T	NP_079390.3:p.Arg1355Cys	4.97E-05	9.53E-05	Delet	Prob_dam	32	abs	VUS	Unk	
64	Limb Girdle Muscular Dystrophy	*CEP290*	Hom	Mis	NM_025114.4:c.4805C>T	NP_079390.3:p.Thr1602Met	0.000226	0.00027958	Delet	Poss_dam	27.7	abs	lik_path, VUS	Unk	0M, 0m
65	Undiagnosed Monogenic Disorders	*CEP290*	Het	Mis	NM_025114.4:c.5909C>A	NP_079390.3:p.Thr1970Asn	0	1.27E-05	Delet	Prob_dam	25.6	abs	abs	Unk	0M, 0m
			Het	SG, FS	NM_025114.4:c.7283_7286dup	NP_079390.3:p.Tyr2429Ter	2.11E-05	2.54E-05				abs	abs	Unk	
68	Early Onset Dementia	*CEP290*	Het	Mis	NM_025114.4:c.31A>G	NP_079390.3:p.Met11Val	7.72E-05	6.35E-06	Delet	Ben	23.2	-	VUS	Unk	0M, 0m
			Het	Mis	NM_025114.4:c.2447G>A	NP_079390.3:p.Arg816His	3.13E-05	6.35E-05	Delet	Prob_dam	26.4	-	VUS	Unk	
69	Epilepsy plus other features	*CEP290*	Het	Mis	NM_025114.4:c.2446C>T	NP_079390.3:p.Arg816Cys	5.00E-05	2.54E-05	Delet	Prob_dam	32	25741868	VUS	Unk	0M, 0m
			Het	Mis	NM_025114.4:c.4741C>T	NP_079390.3:p.Leu1581Phe	1.32E-05	1.27E-05	Delet	Poss_dam	25.1	25741868	VUS	Unk	
71	Hereditary Ataxia	*DYNC2H1*	Het	Mis	NM_001377.3:c.10142C>T	NP_001368.2:p.Pro3381Leu	3.62E-05	1.00E-04	Delet	Prob_dam	31	abs	lik_path, path	Unk	0M, 0m
			Het	Mis	NM_001377.3:c.3419G>T	NP_001368.2:p.Gly1140Val	0.0003938	0.00044987	Delet	Ben	23.2	abs	VUS	Unk	
72	Early Onset Dementia	*OFD1* (♀)	Het	Spl_A	NM_003611.3:c.936-1G>A	-	0	6.35E-06				abs	abs	Unk	0M, 0m
73	Early Onset Dystonia	*OFD1* (♀)	Het	FS	NM_003611.3:c.1911del	NP_003602.1:p.Glu637AspfsTer29	0	6.35E-06				abs	abs	Unk	0M, 0m

Abbreviations: HGVSc = Human Genome Variation Society coding, HGVSp = Human Genome Variation Society protein, AF = allele frequency, MAF = maximum allele frequency, 100K = 100,000 Genomes Project, ACMG = American College of Medical Genetics and Genomics, M = male, F = female, Het = heterozygous, Hom = homozygous, Hemi = hemizygous, FS = frameshift, SV = structural variant, SG = stop gain, SL = start loss, Intr = intronic, Syn = synonymous, Mis = missense, Spl A = splice acceptor, Spl Reg = splice region, Abs = absent, Mat = maternal, Pat = Paternal, Unk = unknown, Path = pathogenic, Lik_path = likely pathogenic, VUS = variant of uncertain significance, M = major clinical feature, m = minor clinical feature

Figure legends

Figure 1: Reverse phenotyping diagnostic research workflow.

Figure 2: Clinical features of participant #32 consistent with a syndromic ciliopathy. **A** (left eye) & **B** (right eye): upper panels, colour fundoscopy of retina; lower panels, fundus autofluorescence images showing perimacular pigment changes (arrowheads) and relatively hypofluorescent central macula. **C:** Optical coherence tomography (OCT) for left eye (L; left panel) and right eye (R; right panel), with the plane of OCT shown by green arrows in left-hand regions of each panel, showing loss of ellipsoid zone outside of the central macula with disruption of the outer nuclear later (*) indicative of rod-cone photoreceptor dystrophy. Arrowhead indicates cystoid macular oedema for the left retina. Scale bars = 200μm. **D:** Computed tomography (CT) axial section of chest showing "signet ring" signs (arrowheads; detail shown in inset) typical of bronchiectasis [46]. Abbreviations: A, anterior; FP, foveal pit; INL, inner nuclear layer; IS/OS, inner segment/outer segment; L, left; ONL, outer nuclear layer; P, posterior; R, right; RPE-CC, retinal pigment epithelium-choriocapillaris complex.

Figure 3. Likely pathogenic structural variants and other variants in selected ciliopathy genes identified through the reverse phenotyping research diagnostic workflow. **A:** IGV plot of *ALMS1* (NM_015120.4) in participant #45. We observed a monoallelic complex SV in the *ALSM1* gene spanning from chr2:g.73424245 to chr2:g. 73544334 (GRCh38). **B:** Diagrammatic representation of complex *ALMS1* SV in participant #45. After inspection of the IGV plots, we surmised that the alternative allele is a paired-duplication inversion, with block A at chr2:g.73424245_73427355, covering exons 4 and 5 (NM_015120.4), block B at chr2:g.73427355_73484777, covering exons 6-9, and block C at chr2:g. 73484777_ 73544334. Note that the boundary between block B and C is an estimate as it is within a region with relatively low alignment quality. **C:** IGV plot of heterozygous 56kb deletion identified in *DYNC2H1* (NM_001377.3) in participant #70. The terminal four exons (86 – 89) have been deleted. **D:** Alamut screenshot for *CEP290* c.6011+874G>T variant in participant #49. Top tracks are donor/acceptor splice site predictions for the reference sequence and the bottom tracks are donor/acceptor predictions for the mutated sequence.

Green highlighting identifies increasing scores for a potential splice acceptor site in the non-reference mutated sequence track. **E:** Analysis of the *BBS1* locus for CMC cohort participant #59 following trio whole genome sequencing. **i:** The maternally-inherited pathogenic variant, NM_024649.5:c.1169T>G, NP_078925.3:p.(Met390Arg) (highlighted by the black frames) is in *trans* with a paternally-inherited mobile element insertion for which the target site duplication sequence is highlighted (red frames). Soft-clipped junction spanning reads, showing inserted nucleotides and the terminal poly(A) tract, are visible. **ii:** Sanger sequencing confirmation of the maternally-inherited c.1169T>C mutation. Exon 12 coding sequence is highlighted in peach. **iii:** Duplex screening assay (37) confirming that the mobile element insertion was present in the proband and his father (270 bp band). Upstream (**iv**) and downstream (**v**) junction fragments confirm that the target site duplication sequence is as previously reported (37). Exon 13 coding sequence is highlighted in grey. Genomic coordinates are according to Human Genome build Hg38. Variant nomenclature is according to transcript NM_024649.5. Abbreviations: Ex, exon; ctrl, control.

Step 1: genes of interest submitted to Gene-Variant Workflow script

BBS1	BBS10	ALMS1	DYNC2H1	WDR34	OFD1	NPHP1	TMEM67	CEP290

List of all variants across 100K dataset with Ensembl VEP annotations and linked Plate Key IDs

Step 2: filtering and prioritisation of SNVs using custom Python script

A. Exclude common variants: 100K MAF ≥ 0.002; gnomAD AF ≥ 0.002
B. Exclude variants called in non canonical transcripts
C. Create prioritised SNV sub-lists: ClinVar Pathogenic → VEP High Impact → SIFT deleterious missense

Step 3: search for potentially pathogenic SVs using SVRare script

A. All unsolved, affected individuals with heterozygous variants on ClinVar pathogenic or Ensembl VEP High Impact prioritized sub-lists submitted to SVRare script
B. SVs overlapping coding regions of genes of interest extracted

Step 4: search for novel splicing variants using custom SpliceAI script

A. All rare variants submitted to SpliceAI using custom Python script
B. Variants potentially affecting splicing extracted (SpliceAI delta scores (DS) > 0.5)

Step 5: search for molecular diagnoses amongst prioritised SNV sub-lists, SVRare prioritised variants and SpliceAI prioritised variants

Recessive gene(s): *BBS1, BBS10, ALMS1, DYNC2H1, WDR34, NPHP1, TMEM67, CEP290*
- Homozygous variants, compound heterozygous variants

X-linked gene(s): *OFD1*
- Heterozygous variants in females, Hemizygous variants in males

Step 6: reverse phenotyping – link to clinical data

A. Extract participant data from LabKey
   - Exclude unaffected relatives
B. Extract GMC exit questionnaires from LabKey
   - Exclude participants already marked ”solved” with variants in alternative genes
C. Extract entered HPO terms from LabKey and look for linked clinical data via Participant Explorer
   - Check for presence of key clinical features associated with variants in gene of interest

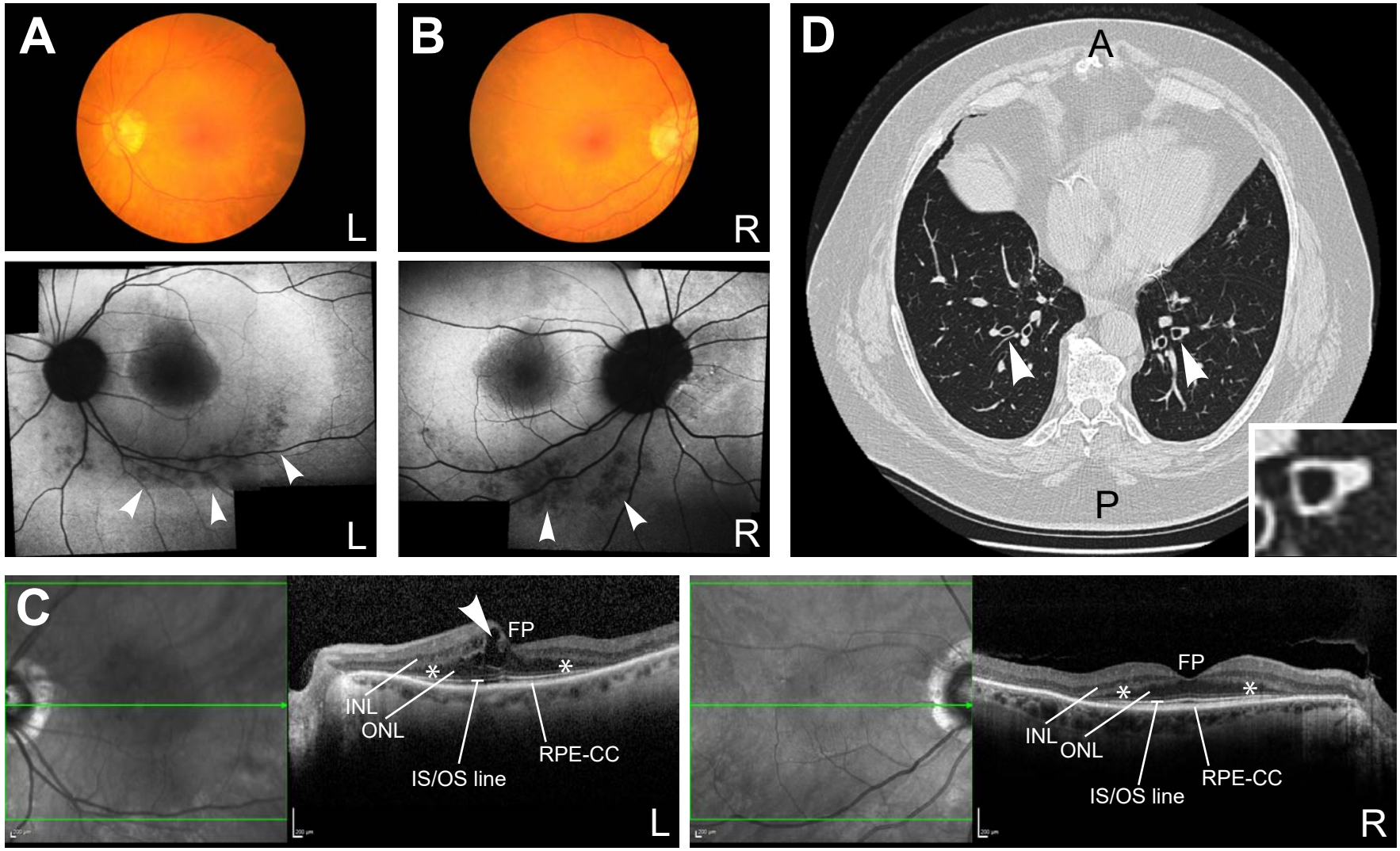
Step 7: determine whether novel molecular diagnoses can be reported

≥ 1 key clinical feature present related to identified molecular diagnosis → REPORT to recruiting clinician

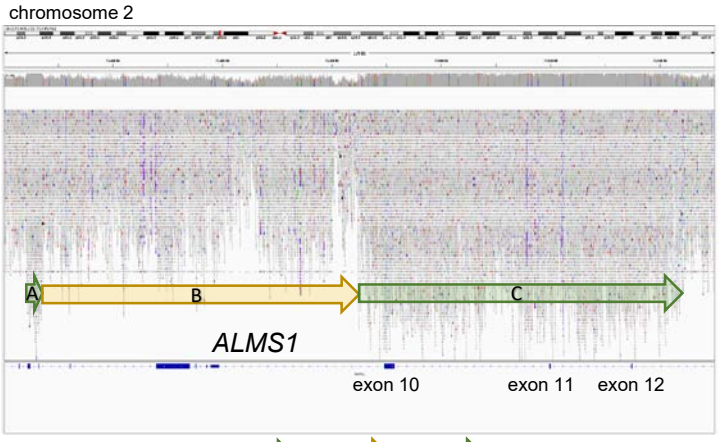
No key clinical features present related to identified molecular diagnosis → DO NOT REPORT

Step 8: ACMG assessment and assignment of diagnostic confidence

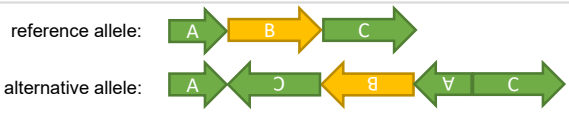
Mode of inheritance	Confident diagnosis	Probable diagnosis	Possible diagnosis
Recessive	2 pathogenic / likely-pathogenic variants	1 pathogenic / likely pathogenic variant + 1 VUS	2 VUSs
X-linked	1 pathogenic / likely pathogenic variant	N/A	1 VUS


A
L
B
R
L
R
D
A
P
C
FP
INL
ONL
IS/OS line
RPE-CC
L
FP
INL
ONL
IS/OS line
RPE-CC
R

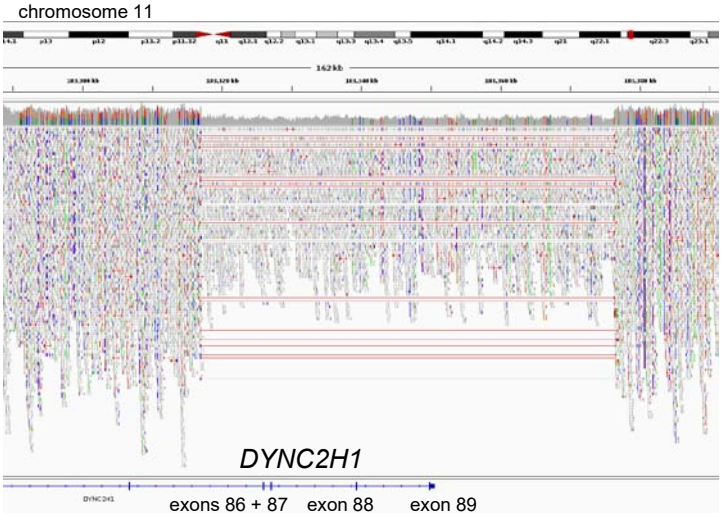
A complex *ALMS1* SV, participant #45



B

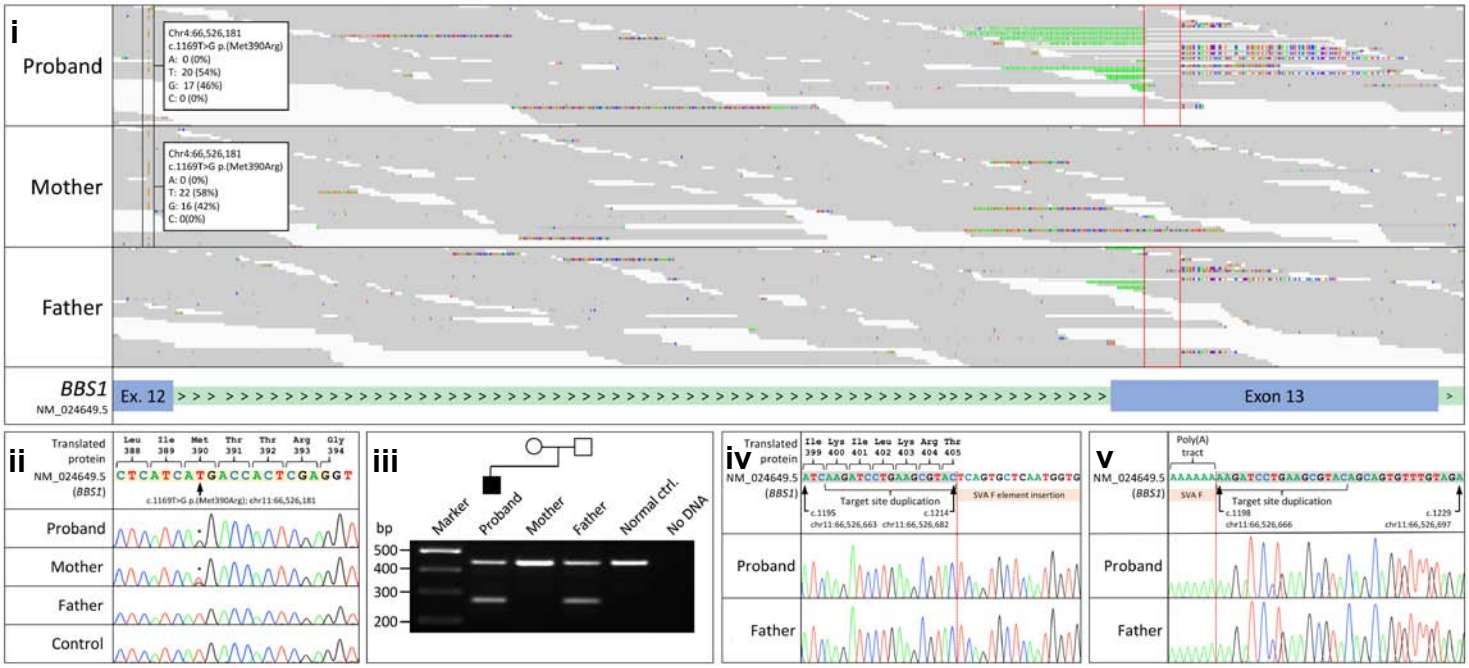


C heterozygous *DYNC2H1* deletion, participant #70



D *CEP290* c.6011+874G>T variant, participant #49

E *BBS1* c.1169T>C mutation and mobile element insertion, participant #59



Supplementary Table 1: Selection of leading multi-systemic ciliopathy disease genes from the medical literature

Ciliopathy syndrome	Leading genetic cause(s)	Mode of inheritance	Further ciliopathies associated with gene	Reference(s)
Bardet-Biedl syndrome (BBS)	*BBS1* (23.4% of all BBS)	Recessive	N/A	(1-3)
	BBS10 (14.5% of all BBS)	Recessive	N/A	
Alström Syndrome (ALMS)	*ALMS1* (only causative gene)	Recessive	-Non-syndromic retinal dystrophy -Non-syndromic cardiomyopathy	(4-8)
Joubert syndrome (JBTS) and Meckel Gruber syndrome (MKS)	*TMEM67* (6-26% of all JBTS; 16% of all MKS)	Recessive	-NPHP with hepatic fibrosis -COACH syndrome (cerebellar vermis hypo/aplasia, oligophrenia, ataxia, ocular coloboma, and hepatic fibrosis)	(9-17)
	CEP290 (6-22% of all JBTS, 2nd most common cause of MKS)	Recessive	-Leber Congenital Amaurosis (LCA) / Early-Onset Severe Retinal Dystrophy (EOSRD) (15-20% of LCA / EOSRD cases) -NPHP -BBS -Senior-Løken syndrome -COACH syndrome	(14, 18-24)
Jeune Asphyxiating Thoracic Dystrophy (JATD)	*DYNC2H1* (~50% of all JATD)	Recessive	N/A	(25-28)
	WDR34 (~10% of all JATD)	Recessive		
Nephronophthisis (NPHP)	*NPHP1* (20-25% of all NPHP)	Recessive	JBTS	(29-31)
Oral-facial-digital syndrome (OFD) Type 1	*OFD1* (only genetic cause)	X-linked dominant	JBTS (X-linked recessive)	(9, 32)

Supplementary Table 2: HPO terms linked to clinical key terms for ciliopathy syndromes

Key term	HPO ID	HPO descriptor	Linked HPO terms included in analysis
Retinal dystrophy	HP:0000556	Breakdown of light-sensitive cells in back of eye	• Cone/cone-rod dystrophy + sub-terms • Rod-cone dystrophy + sub-terms • Pattern dystrophy of the retina + sub-terms
Abnormality of eye movement	HP:0000496	An abnormality in voluntary or involuntary eye movements or their control	• Oculomotor apraxia (JBTS) • Nystagmus (LCA) • Roving eye movements (LCA)
Abnormal renal morphology / renal insufficiency	HP:0012210	Any structural anomaly of the kidney	• Abnormal localisation of kidney + sub-terms • Abnormal renal cortex morphology + sub-terms • Abnormal renal echogenicity + sub-terms • Abnormal renal medulla morphology + sub-terms • Abnormal renal pelvis morphology + sub-terms • Renal cyst + sub-terms • Renal dysplasia + sub-terms • Renal fibrosis + sub-terms • Renal hypoplasia/aplasia + sub-terms
	HP:0000083	A reduction in the level of performance of the kidneys in areas of function comprising the concentration of urine, removal of wastes, the maintenance of electrolyte balance, homeostasis of blood pressure, and calcium metabolism	• Chronic kidney disease + sub-terms
Abnormality of the liver	HP:0001392	An abnormality of the liver	• Abnormal liver morphology + sub-terms • Abnormal liver physiology + sub-terms • Abnormality of the biliary system + sub-terms
Abnormality of the genitourinary system	HP:0000119	The presence of any abnormality of the genitourinary system	• Abnormality of the genital system + sub-terms • Abnormality of the urinary system + sub-terms
Cardiomyopathy	HP:0001638	A myocardial disorder in which the heart muscle is structurally and functionally abnormal, in the absence of coronary artery disease, hypertension, valvular disease and congenital heart disease sufficient to cause the observed myocardial abnormality.	• All sub-terms
Sensorineural hearing impairment	HP:0000407	A type of hearing impairment in one or both ears related to an abnormal functionality of the cochlear nerve.	• All sub-terms

Abnormality of the sense of smell	HP:0004408	An anomaly in the ability to perceive and distinguish scents (odors).	• All sub-terms
Abnormal pattern of respiration	HP:0002793	An anomaly of the rhythm or depth of breathing	• Apnoea + sub-terms • Tachypnoea + sub-terms
Hypogonadotrophic hypogonadism	HP:000044	Hypogonadotropic hypogonadism is characterized by reduced function of the gonads (testes in males or ovaries in females) and results from the absence of the gonadal stimulating pituitary hormones: follicle stimulating hormone (FSH) and luteinizing hormone (LH).	• All sub-terms
Glucose intolerance	HP:0001952	Glucose intolerance (GI) can be defined as dysglycemia that comprises both prediabetes and diabetes. It includes the conditions of impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) and diabetes mellitus (DM).	• Type II diabetes mellitus + sub-terms • Impaired glucose tolerance + sub-terms
Obesity	HP:0001513	Accumulation of substantial excess body fat.	• All sub-terms
Hypertriglyceridemia	HP:0002155	An abnormal increase in the level of triglycerides in the blood	• All sub-terms
Intellectual disability	HP:0001249	Subnormal intellectual functioning which originates during the developmental period. Intellectual disability, previously referred to as mental retardation, has been defined as an IQ score below 70.	• All sub-terms
Neurodevelopmental delay	HP:0012758	None listed	• All sub-terms
Hypotonia	HP:0001252	Hypotonia is an abnormally low muscle tone (the amount of tension or resistance to movement in a muscle). Even when relaxed, muscles have a continuous and passive partial contraction which provides some resistance to passive stretching. Hypotonia thus manifests as diminished resistance to passive stretching. Hypotonia is not the same as muscle weakness, although the two conditions can co-exist.	• All sub-terms
Ataxia	HP:0001251	Cerebellar ataxia refers to ataxia due to dysfunction of the cerebellum. This causes a variety of elementary neurological deficits including asynergy (lack of coordination between muscles, limbs and joints), dysmetria (lack of ability to judge distances that can lead to under- or overshoot in grasping movements), and dysdiadochokinesia (inability to perform	• All sub-terms

		rapid movements requiring antagonizing muscle groups to be switched on and off repeatedly).	
Abnormality of brain morphology	HP:0012443	A structural abnormality of the brain, which has as its parts the forebrain, midbrain, and hindbrain.	• Abnormal brainstem morphology + sub-terms • Abnormal cerebral ventricle morphology + sub-terms • Abnormal midbrain morphology + sub-terms • Abnormality of forebrain morphology + sub-terms • Abnormality of hindbrain morphology + sub-terms
Polydactyly	HP:0010442	A congenital anomaly characterized by the presence of supernumerary fingers or toes.	• All sub-terms
Short stature	HP:0004322	A height below that which is expected according to age and gender norms. Although there is no universally accepted definition of short stature, many refer to "short stature" as height more than 2 standard deviations below the mean for age and gender (or below the 3rd percentile for age and gender dependent norms).	• All sub-terms
Thoracic hypoplasia	HP:0005257	None listed	• All sub-terms
Brachydactyly / micromelia	HP:0001156	Digits that appear disproportionately short compared to the hand/foot.	• All sub-terms
Micromelia	HP:0002983	The presence of abnormally small extremities.	• All sub-terms
Abnormality of dentition	HP:0000164	Any abnormality of the teeth	• All sub-terms
Abnormal oral morphology	HP:0031816	Any structural anomaly of the mouth, which is also known as the oral cavity.	• All sub-terms
OFD1-specific facial dysmorphic features	HP:0000316	Hypertelorism: Interpupillary distance more than 2 SD above the mean (alternatively, the appearance of an increased interpupillary distance or widely spaced eyes)	• This term only
	HP:0000430	Underdeveloped nasal alae: Thinned, deficient, or excessively arched ala nasi.	• This term only
	HP:0000347	Micrognathia: Developmental hypoplasia of the mandible.	• This term only

Supplementary Table 3: Participants reported solved or partially solved in GMC exit questionnaires with variants in ciliopathy genes of interest

RESEARCH ID	GMC exit report outcome	Reported Sex	100K Recruitment Category	Gene	Variant Zygosity	Consequence	HGVSc	HGVSp	GMC exit questionnaire ACMG Class
1	Solved	MALE	BBS	ALMS1	Het	FS	NM_015120.4:c.10775del	NP_055935.4:p.Thr3592LysfsTer6	Path
					Het	SG	NM_015120.4:c.11107C>T	NP_055935.4:p.Arg3703Ter	Path
2	Solved	FEMALE	CDS	ALMS1	Het	SG	NM_015120.4:c.10975C>T	NP_055935.4:p.Arg3659Ter	Path
					Het	SG; FS	NM_015120.4:c.4571dup	NP_055935.4:p.Tyr1524Ter	Path
3	Solved	MALE	RCD	ALMS1	Het	FS	NM_015120.4:c.284del	NP_055935.4:p.Pro95ArgfsTer19	Path
					Het	FS	NM_015120.4:c.1793del	NP_055935.4:p.Glu598GlyfsTer3	Path
4	Solved	FEMALE	LCA or EOSRD	ALMS1	Het	SG	NM_015120.4:c.10483C>T	NP_055935.4:p.Gln3495Ter	Path
					Het	FS	NM_015120.4:c.6590del	NP_055935.4:p.Lys2197SerfsTer10	Path
5	Solved	FEMALE	ID; RCD	ALMS1	Het	FS	NM_015120.4:c.6570del	NP_055935.4:p.Ser2191HisfsTer16	Path
					Het	FS	NM_015120.4:c.10831_10832del	NP_055935.4:p.Arg3611AlafsTer6	Path
6	Solved	MALE	BBS	ALMS1	Het	FS	NM_015120.4:c.11881dup	NP_055935.4:p.Ser3961PhefsTer12	Path
					Het	“Large delins”	Data missing	Data missing	Likely path
7	Solved	MALE	URUMD	ALMS1	Hom	FS	NM_015120.4:c.2515dup	NP_055935.4:p.Ser839PhefsTer8	Path
8	Solved	MALE	BBS	ALMS1	Hom	FS	NM_015120.4:c.4684_4690dup	NP_055935.4:p.Ile1564AsnfsTer20	Path
9	Solved	FEMALE	RCD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
10	Solved	FEMALE	RCD	BBS1	Hom	Mis	19)	NP_078925.3:p.Met390Arg	Path
11	Solved	FEMALE	RCD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
12	Solved	MALE	RCD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
13	Solved	FEMALE	RCD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
14	Solved	FEMALE	SEOO +/- OEF + SS	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
15	Solved	FEMALE	ID	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
16	Solved	MALE	BBS	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
17	Solved	MALE	RCD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
18	Solved	MALE	CKD	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
19	Partially	MALE	ID	BBS1	Het	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
					Het	SG	NM_024649.5:c.871C>T	NP_078925.3:p.Gln291Ter	Path
20	Solved	FEMALE	Mito D	BBS1	Hom	Mis	NM_024649.5:c.1169T>G	NP_078925.3:p.Met390Arg	Path
21	Partially	MALE	RDS	BBS10	Het	Mis	NM_024685.4:c.1230T>G	NP_078961.3:p.His410Gln	Likely path
					Het	FS	NM_024685.4:c.271dup	NP_078961.3:p.Cys91LeufsTer5	Path
22	Solved	MALE	CAKUT	CEP290	Het	FS	NM_025114.4:c.2848dup	NP_079390.3:p.Gln950ProfsTer6	Path
					Het	Mis	NM_025114.4:c.2817G>T	NP_079390.3:p.Lys939Asn	Likely path

23	Solved	FEMALE	JBTS	CEP290	Hom	SG	NM_025114.4:c.5932C>T	NP_079390.3:p.Arg1978Ter	Path
24	Solved	MALE	LCA or EOSRD	CEP290	Hom	In-frame deletion	NM_025114.4:c.4661_4663del	NP_079390.3:p.Glu1554del	Likely path
25	Solved	FEMALE	LCA or EOSRD	CEP290	Het	FS	NM_025114.4:c.5434_5435del	NP_079390.3:p.Glu1812LysfsTer5	Path
					Het	SG	NM_025114.4:c.5668G>T	NP_079390.3:p.Gly1890Ter	Path
26	Solved	FEMALE	CAKUT	CEP290	Hom	SG	NM_025114.4:c.4174G>T	NP_079390.3:p.Glu1392Ter	Likely path
27	Partially	MALE	ID	CEP290	Het	SG	NM_025114.4:c.322C>T	NP_079390.3:p.Arg108Ter	Path
					Het	FS	NM_025114.4:c.3422dup	NP_079390.3:p.Leu1141PhefsTer5	Path
28	Solved	MALE	RCD	CEP290	Het	SG	NM_025114.4:c.1984C>T	NP_079390.3:p.Gln662Ter	Path
					Het	SG	NM_025114.4:c.7048C>T	NP_079390.3:p.Gln2350Ter	Path
29	Solved	FEMALE	BBS	CEP290	Het	SG	NM_025114.4:c.5668G>T	NP_079390.3:p.Gly1890Ter	Path
					Het	SG	NM_025114.4:c.322C>T	NP_079390.3:p.Arg108Ter	Path
30	Solved	MALE	RCD	DYNC1H1	Hom	SG	NM_001080463.2:c.9836C>A	NP_001073932.1:p.Ser3279Ter	Path
31	Solved	MALE	USD	DYNC1H1	Het	Spl A	NM_001080463.2:c.10834-1G>A	-	Path
					Het	Spl Reg	NM_001080463.2:c.6140-5A>G	-	Likely path
32	Solved	MALE	RCD	OFD1	Hemi	FS	NM_003611.3:c.2680_2681del	NP_003602.1:p.Glu894ArgfsTer6	Path
33	Solved	FEMALE	RCD	NPHP1	Het	Mis	NM_001128178.3:c.1882C>T	NP_001121650.1:p.Arg628Trp	Likely path
					Het	"Whole gene deletion"	Data missing	Data missing	Not specified
34	Solved	MALE	UKFIYP	NPHP1	Hom	Mis	NM_001128178.3:c.859G>A	NP_001121650.1:p.Gly287Arg	Path
35	Solved	MALE	UKFIYP	NPHP1	Hom	SG	NM_001128178.3:c.1142G>A	NP_001121650.1:p.Trp381Ter	Path
36	Solved	FEMALE	UKFIYP	OFD1	Het	FS	NM_003611.3:c.1651_1654del	NP_003602.1:p.Thr551ProfsTer2	Path
37	Solved	FEMALE	SARMIRD	OFD1	Het	Mis	NM_003611.3:c.1363A>C	NP_003602.1:p.Lys455Gln	VUS
38	Solved	FEMALE	Craniosyn S	OFD1	Het	Spl Reg	NM_003611.3:c.382-4A>G	-	VUS
39	Solved	FEMALE	CKD	OFD1	Het	Spl A	NM_003611.3:c.112-1G>A	-	Path
40	Partially	FEMALE	RMCD	OFD1	Het	FS	NM_003611.3:c.306del	NP_003602.1:p.Glu103LysfsTer42	Likely path
41	Solved	MALE	CKD	TMEM67	Het	FS	NM_153704.6:c.103del	NP_714915.3:p.Gln35ArgfsTer52	Path
					Het	FS	NM_153704.6:c.415_416del	NP_714915.3:p.Asp139HisfsTer2	Path
42	Partially	MALE	ID	TMEM67	Het	Mis	NM_153704.6:c.1319G>A	NP_714915.3:p.Arg440Gln	Path
					Het	Mis	NM_153704.6:c.2498T>C	NP_714915.3:p.Ile833Thr	Likely path
43	Solved	MALE	RCD	CEP290	Het	FS	NM_025114.4:c.254dup	NP_079390.3:p.Asn85LysfsTer6	Likely path
44	Solved	MALE	LCA or EOSRD	CEP290	Hom	Mis	NM_025114.4:c.21G>T	NP_079390.3:p.Trp7Cys	Likely path

Abbreviations: 100K = 100,000 Genomes Project, GMC = Genomic Medicine Centre, ACMG = American College of Medical Genetics and Genomics, BBS = Bardet-Biedl syndrome, CDS = cone dysfunction syndrome, RCD = rod-cone dystrophy, LCA or EOSRD = Leber Congenital Amaurosis or Early-Onset Severe Retinal Dystrophy, ID = intellectual disability, URUMD = Ultra-rare undescribed monogenic disorders, SEOO +/- OEF + SS = Significant early-onset obesity with or without other endocrine features and short stature, CKD = cystic kidney disease, Mito D = mitochondrial disorders, RDS = rod-dysfunction syndrome, CAKUT = Congenital Anomaly of the Kidneys and Urinary Tract, JBTS = Joubert

syndrome, USD = Unexplained skeletal dysplasia, UKFIYP = Unexplained kidney failure in young people, SARMIRD = Single autosomal recessive mutation in rare disease, Craniosyn S = craniosynostosis syndromes, RMCD = Rare multisystem ciliopathy disorders, Het = heterozygous, Hom = homozygous, Hemi = hemizygous, FS = frameshift, SG = stop gain, Mis = missense, Spl A = splice acceptor, Spl Reg = splice region, Path = pathogenic, Likely path = likely pathogenic, VUS = variant of uncertain significance

Supplementary Table 4: Prioritised variants extracted through reverse phenotyping diagnostic research workflow

Step 2 workflow inputs and outputs: filtering and prioritisation of SNVs using custom Python script																		
INPUTS																		
INPUT SNV DATA: All SNVs from the 100K dataset for each selected ciliopathy gene generated by Gene-Variant Workflow. Separate lists for participants called on GrCh37 and GrCh38																		
Gene	**ALMS1**		**BBS1**		**BBS10**		**DYNC2H1**		**WDR34**		**OFD1**		**NPHP1**		**TMEM67**		**CEP290**	
Build	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38	GrCh37	GrCh38
# un-filtered Gene-Variant Workflow variants	52420	287121	24050	71969	166	601	80615	284569	7636	234958	2122	27257	30997	104051	28384	95596	19436	96000
PROCESS: filter using custom python script filter_gene_variant_workflow.py A: Exclude common variants: 100K MAF ≥ 0.002; gnomAD AF ≥ 0.002 B: Exclude variants called in non-canonical transcripts ⇓																		
# filtered variants: rare, canonical transcripts only	11862	43098	1217	3802	153	588	16127	59165	1465	4939	279	4365	3399	12254	2810	10226	3740	14200
PROCESS: extract prioritised SNV sub-lists using custom python script filter_gene_variant_workflow.py: • ClinVar pathogenic/likely pathogenic • VEP High Impact (stop_gained, stop_lost, start_lost, splice_acceptor_variant, splice_donor_variant, frameshift_variant, transcript_ablation, transcript_amplification) • SIFT deleterious missense																		
OUTPUTS																		
Gene	**ALMS1**		**BBS1**		**BBS10**		**DYNC2H1**		**WDR34**		**OFD1**		**NPHP1**		**TMEM67**		**CEP290**	
Total ClinVar Pathogenic	13	43	1	14	5	22	16	58	2	9	0	64	3	8	10	36	22	78
Total VEP High Impact	30	130	2	22	5	28	19	141	4	38	0	70	7	35	11	57	36	167
Total SIFT deleterious missense	167	643	33	86	18	86	125	556	32	107	5	75	26	79	33	167	84	344

DISTRIBUTION OF PRIORITISED VARIANTS BETWEEN DIFFERENT PRIORITISED SNV SUB-LISTS																		
Gene	**ALMS1**		**BBS1**		**BBS10**		**DYNC2H1**		**WDR34**		**OFD1**		**NPHP1**		**TMEM67**		**CEP290**	
# ClinVar Pathogenic + VEP High Impact	13	43	0	11	5	17	5	26	1	6	0	58	2	7	4	20	19	73
# ClinVar pathogenic + SIFT deleterious missense	0	0	1	3	0	5	10	30	1	3	0	5	1	1	6	14	2	4
# VEP High Impact (only)	17	87	2	11	0	11	13	115	3	32	0	12	5	28	7	37	17	94
# SIFT deleterious missense (only)	167	643	32	83	18	81	115	526	31	104	5	70	25	78	27	153	82	340
# ClinVar Pathogenic (only)	0	0	0	0	0	0	1	2	0	0	0	1	0	0	0	2	1	1
Total	197	773	35	108	23	114	144	699	36	145	5	146	33	114	44	226	121	512

Step 3 workflow inputs and outputs: search for potentially pathogenic SVs using SVRare script

INPUTS

INPUT DATA: PlateKey identifiers for all unsolved 100K participants (probands and affected relatives) with heterozygous ClinVar pathogenic or VEP high impact prioritised SNVs in one of the nine ciliopathy genes
N = 801 participants

PROCESS: Submitted to SVRare script (Yu et al, 2021)
Extracts participants with SVs called by Manta and/or Canvas with ≤ 10 calls across the 100K database, overlapping coding regions of the 9 ciliopathy genes
⇓

OUTPUTS

Gene	**ALMS1**		**BBS1**		**BBS10**		**DYNC2H1**		**WDR34**		**OFD1**		**NPHP1**		**TMEM67**		**CEP290**	
# Prioritised SNVs	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
Impression	N/a	LP	N/a	N/a	N/a	N/a	LP	N/a	N/a	N/a	N/a	Excl: 2nd hit in different gene	N/a	N/a	N/a	N/a	N/a	Excl: alternative diagnosis

Step 4 workflow inputs and outputs: search for novel splicing variants using custom SpliceAI script

INPUTS

INPUT DATA: all rare variants (100K MAF ≤ 0.002; gnomAD AF ≤ 0.002) called in canonical transcripts in the nine ciliopathy genes identified in unsolved 100K participants
AS PER Step 2: Gene-Variant Workflow rare SNVs called in canonical transcripts filtered through custom python script (filter_gene_variant_workflow.py)

PROCESS: Run through custom SpliceAI Python script (find_variants_by_gene_and_SpliceAI_score.py)

⇓

FILTERING:

- Variants called in unaffected relatives excluded
- Variants with SpliceAI delta score (DS) > 0.5 retained
- Variants already assessed on other SNV prioritised sub-lists excluded

⇓

OUTPUTS

Gene	ALMS1		BBS1		BBS10		DYNC2H1		WDR34		OFD1		NPHP1		TMEM67		CEP290	
# rare variants with SpliceAI DS >0.5	1	22	3	10	0	1	7	53	1	9	0	10	3	12	2	15	4	34

The number of variants input, filtered and prioritised in steps 2, 3 and 4 of the reverse phenotyping diagnostic research workflow. Note that 100K participants had genomes called on GrCh37 or GrCh38 depending on when they were recruited to the project.
Abbreviations: SNV = single nucleotide variant, 100K = 100,000 Genomes Project, AF = allele frequency, MAF = maximum allele frequency, VEP = Variant Effect Predictor, SV = structural variant, Excl = excluded

Supplementary Data 1: Duplex PCR assay of a *BBS1* exon 13 mobile element insertion

The patient presented with congenital right ptosis, childhood onset high myopia, rod/cone dysfunction, autism, dyspraxia and postaxial polydactyly on the left hand and foot that were removed in childhood. The patient was recruited to the 100,000 Genomes Project (100K) for whole genome sequencing, following identification of a heterozygous pathogenic variant in an autosomal recessive disease gene through mainstream testing. The *BBS1* missense mutation, NM_024649.5:c.1169T>G, NP_078925.3:p.(Met390Arg), was insufficient to confirm the diagnosis in the absence of a second pathogenic variant. 100K tiering failed to identify a second deleterious allele in *BBS1*. Manual inspection of the aligned sequence reads using the Integrative Genome Browser (IGV) v.2.4.10 (http://software.broadinstitute.org/software/igv/) (33) and interrogation of soft-clipped reads using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) (34), revealed a soft-clipped read signature that was consistent with a 2.4 kb insertion of an SVA F family element mobile element (35).

To confirm the *BBS1* heterozygous missense variant, c.1169T>C, a PCR amplicon was first optimised; each reaction comprised 0.5 µL of genomic DNA (~50 ng/µL) 19.3 µL MegaMix PCR reagent (Microzone Ltd., Haywards Heath, UK) and 0.1 µL each of 10 µM forward (dTGTAAAACGACGGCCAGTAAAGGCAGCATTGTGAAGGG) and reverse (dCAGGAAACAGCTATGACCCCCTTCACTCCCGACTTCAA) primers. Thermocycling conditions comprised 94°C for 5 minutes then 30 cycles of 94°C for 30 seconds, 55°C for 1 minute and 72°C for 2 minutes before a final extension step at 72°C for 5 minutes. Amplification products were resolved on a 1% Tris-borate-EDTA agarose gel, before being extracted and purified using a QIAquick column (Qiagen GmbH, Hilden, Germany), then Sanger sequenced using an ABI3730 following manufacturer's protocols throughout (Life Technologies Ltd., Paisley, UK). Sequence chromatograms were analysed using 4Peaks v.1.8 (http://nucleobytes.com/4peaks/index.html). Universal sequence tags (underlined) were incorporated into primer tails for use with our routine diagnostic workflow.

To verify the apparent *BBS1* exon 13 mobile element insertion, we implemented the duplex PCR assay as described previously (35). Each reaction comprised 0.5 µL of genomic DNA (~50 ng/µL) 19.2 µL of MegaMix PCR reagent and 0.1 µL each of 10 µM primer. These included a common intron 12 forward (dCACAGTACTCCACAAATAACTGCT), an intron 13 reverse

(dATTCCCCCAGCTTTGCTGT) and insertion-specific reverse (dCAGCCTGGGCACCATTGA) primer. Thermocycling conditions required 35 cycles, but were otherwise as described above. Amplification products specific for the normal (440 bp) and insertion-containing (270 bp) allele were resolved on a 2% TRIS-borate-EDTA agarose gel prior to gel extraction and Sanger sequencing. To determine the precise sequence of the downstream target site duplication a further PCR was optimised for Sanger sequencing, using previously reported forward (F9: dAGTACCCAGGGACAAACACT) and reverse (R5: dGTCTTTCGGGGCACATTGAG) primers (35). Analysis of parental alignments supported the mobile element insertion being in *trans* with the maternally-inherited c.1169T>C mutation, with Sanger sequencing confirming the presence of the insertion in the proband and his father.

Supplementary references

1. Niederlova V, Modrak M, Tsyklauri O, Huranova M, Stepanek O. Meta-analysis of genotype-phenotype associations in Bardet-Biedl syndrome uncovers differences among causative genes. Hum Mutat. 2019;40(11):2068-87.
2. Shamseldin HE, Shaheen R, Ewida N, Bubshait DK, Alkuraya H, Almardawi E, Howaidi A, Sabr Y, Abdalla EM, Alfaifi AY, Alghamdi JM, Alsagheir A, Alfares A, Morsy H, Hussein MH, Al-Muhaizea MA, Shagrani M, Al Sabban E, Salih MA, Meriki N, Khan R, Almugbel M, Qari A, Tulba M, Mahnashi M, Alhazmi K, Alsalamah AK, Nowilaty SR, Alhashem A, Hashem M, Abdulwahab F, Ibrahim N, Alshidi T, AlObeid E, Alenazi MM, Alzaidan H, Rahbeeni Z, Al-Owain M, Sogaty S, Seidahmed MZ, Alkuraya FS. The morbid genome of ciliopathies: an update. Genet Med. 2020;22(6):1051-60.
3. Forsyth R, Gunay-Aygun M. Bardet-Biedl Syndrome Overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
4. Paisey RB, Steeds R, Barrett T, Williams D, Geberhiwot T, Gunay-Aygun M. Alström Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
5. Aldrees A, Abdelkader E, Al-Habboubi H, Alrwebah H, Rahbeeni Z, Schatz P. Non-syndromic retinal dystrophy associated with homozygous mutations in the ALMS1 gene. Ophthalmic Genet. 2019;40(1):77-9.
6. Hull S, Kiray G, Chiang JP, Vincent AL. Molecular and phenotypic investigation of a New Zealand cohort of childhood-onset retinal dystrophy. Am J Med Genet C Semin Med Genet. 2020;184(3):708-17.
7. Lazar CH, Kimchi A, Namburi P, Mutsuddi M, Zelinger L, Beryozkin A, Ben-Simhon S, Obolensky A, Ben-Neriah Z, Argov Z, Pikarsky E, Fellig Y, Marks-Ohana D, Ratnapriya R, Banin E, Sharon D, Swaroop A. Nonsyndromic Early-Onset Cone-Rod Dystrophy and Limb-Girdle Muscular Dystrophy in a Consanguineous Israeli Family are Caused by Two Independent yet Linked Mutations in ALMS1 and DYSF. Hum Mutat. 2015;36(9):836-41.
8. Louw JJ, Corveleyn A, Jia Y, Iqbal S, Boshoff D, Gewillig M, Peeters H, Moerman P, Devriendt K. Homozygous loss-of-function mutation in ALMS1 causes the lethal disorder mitogenic cardiomyopathy in two siblings. Eur J Med Genet. 2014;57(9):532-5.
9. Bachmann-Gagescu R, Dempsey JC, Phelps IG, O'Roak BJ, Knutzen DM, Rue TC, Ishak GE, Isabella CR, Gorden N, Adkins J, Boyle EA, de Lacy N, O'Day D, Alswaid A, Ramadevi AR, Lingappa L, Lourenco C, Martorell L, Garcia-Cazorla A, Ozyurek H, Haliloglu G, Tuysuz B, Topcu M, Chance P, Parisi MA, Glass IA, Shendure J, Doherty D. Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. J Med Genet. 2015;52(8):514-22.
10. Vilboux T, Doherty DA, Glass IA, Parisi MA, Phelps IG, Cullinane AR, Zein W, Brooks BP, Heller T, Soldatos A, Oden NL, Yildirimli D, Vemulapalli M, Mullikin JC, Nisc Comparative Sequencing P, Malicdan MCV, Gahl WA, Gunay-Aygun M. Molecular genetic findings and

clinical correlations in 100 patients with Joubert syndrome and related disorders prospectively evaluated at a single center. Genet Med. 2017;19(8):875-82.
11. Parisi M, Glass I. Joubert Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
12. Suzuki T, Miyake N, Tsurusaki Y, Okamoto N, Alkindy A, Inaba A, Sato M, Ito S, Muramatsu K, Kimura S, Ieda D, Saitoh S, Hiyane M, Suzumura H, Yagyu K, Shiraishi H, Nakajima M, Fueki N, Habata Y, Ueda Y, Komatsu Y, Yan K, Shimoda K, Shitara Y, Mizuno S, Ichinomiya K, Sameshima K, Tsuyusaki Y, Kurosawa K, Sakai Y, Haginoya K, Kobayashi Y, Yoshizawa C, Hisano M, Nakashima M, Saitsu H, Takeda S, Matsumoto N. Molecular genetic analysis of 30 families with Joubert syndrome. Clin Genet. 2016;90(6):526-35.
13. Otto EA, Tory K, Attanasio M, Zhou W, Chaki M, Paruchuri Y, Wise EL, Wolf MT, Utsch B, Becker C, Nurnberg G, Nurnberg P, Nayir A, Saunier S, Antignac C, Hildebrandt F. Hypomorphic mutations in meckelin (MKS3/TMEM67) cause nephronophthisis with liver fibrosis (NPHP11). J Med Genet. 2009;46(10):663-70.
14. Hartill V, Szymanska K, Sharif SM, Wheway G, Johnson CA. Meckel-Gruber Syndrome: An Update on Diagnosis, Clinical Management, and Research Advances. Front Pediatr. 2017;5(244).
15. Iannicelli M, Brancati F, Mougou-Zerelli S, Mazzotta A, Thomas S, Elkhartoufi N, Travaglini L, Gomes C, Ardissino GL, Bertini E, Boltshauser E, Castorina P, D'Arrigo S, Fischetto R, Leroy B, Loget P, Bonniere M, Starck L, Tantau J, Gentilin B, Majore S, Swistun D, Flori E, Lalatta F, Pantaleoni C, Penzien J, Grammatico P, Dallapiccola B, Gleeson JG, Attie-Bitach T, Valente EM. Novel TMEM67 mutations and genotype-phenotype correlates in meckelin-related ciliopathies. Hum Mutat. 2010;31(5):E1319-31.
16. Brancati F, Iannicelli M, Travaglini L, Mazzotta A, Bertini E, Boltshauser E, D'Arrigo S, Emma F, Fazzi E, Gallizzi R, Gentile M, Loncarevic D, Mejaski-Bosnjak V, Pantaleoni C, Rigoli L, Salpietro CD, Signorini S, Stringini GR, Verloes A, Zabloka D, Dallapiccola B, Gleeson JG, Valente EM. MKS3/TMEM67 mutations are a major cause of COACH Syndrome, a Joubert Syndrome related disorder with liver involvement. Hum Mutat. 2009;30(2):E432-42.
17. Doherty D, Parisi MA, Finn LS, Gunay-Aygun M, Al-Mateen M, Bates D, Clericuzio C, Demir H, Dorschner M, van Essen AJ, Gahl WA, Gentile M, Gorden NT, Hikida A, Knutzen D, Ozyurek H, Phelps I, Rosenthal P, Verloes A, Weigand H, Chance PF, Dobyns WB, Glass IA. Mutations in 3 genes (MKS3, CC2D2A and RPGRIP1L) cause COACH syndrome (Joubert syndrome with congenital hepatic fibrosis). J Med Genet. 2010;47(1):8-21.
18. Travaglini L, Brancati F, Attie-Bitach T, Audollent S, Bertini E, Kaplan J, Perrault I, Iannicelli M, Mancuso B, Rigoli L, Rozet JM, Swistun D, Tolentino J, Dallapiccola B, Gleeson JG, Valente EM, Zankl A, Leventer R, Grattan-Smith P, Janecke A, D'Hooghe M, Sznajer Y, Van Coster R, Demerleir L, Dias K, Moco C, Moreira A, Kim CA, Maegawa G, Petkovic D, Abdel-Salam GM, Abdel-Aleem A, Zaki MS, Marti I, Quijano-Roy S, Sigaudy S, de Lonlay P, Romano S, Touraine R, Koenig M, Lagier-Tourenne C, Messer J, Collignon P, Wolf N, Philippi H, Kitsiou Tzeli S, Halldorsson S, Johannsdottir J, Ludvigsson P, Phadke SR, Udani V, Stuart B, Magee A, Lev D, Michelson M, Ben-Zeev B, Fischetto R, Benedicenti F, Stanzial F, Borgatti R, Accorsi P, Battaglia S, Fazzi E, Giordano L, Pinelli L, Boccone L, Bigoni S, Ferlini A, Donati MA, Caridi G, Divizia MT, Faravelli F, Ghiggeri G, Pessagno A, Briguglio M,

Briuglia S, Salpietro CD, Tortorella G, Adami A, Castorina P, Lalatta F, Marra G, Riva D, Scelsa B, Spaccini L, Uziel G, Del Giudice E, Laverda AM, Ludwig K, Permunian A, Suppiej A, Signorini S, Uggetti C, Battini R, Di Giacomo M, Cilio MR, Di Sabato ML, Leuzzi V, Parisi P, Pollazzon M, Silengo M, De Vescovi R, Greco D, Romano C, Cazzagon M, Simonati A, Al-Tawari AA, Bastaki L, Mégarbané A, Sabolic Avramovska V, de Jong MM, Stromme P, Koul R, Rajab A, Azam M, Barbot C, Martorell Sampol L, Rodriguez B, Pascual-Castroviejo I, Teber S, Anlar B, Comu S, Karaca E, Kayserili H, Yüksel A, Akcakus M, Al Gazali L, Sztriha L, Nicholl D, Woods CG, Bennett C, Hurst J, Sheridan E, Barnicoat A, Hennekam R, Lees M, Blair E, Bernes S, Sanchez H, Clark AE, DeMarco E, Donahue C, Sherr E, Hahn J, Sanger TD, Gallager TE, Dobyns WB, Daugherty C, Krishnamoorthy KS, Sarco D, Walsh CA, McKanna T, Milisa J, Chung WK, De Vivo DC, Raynes H, Schubert R, Seward A, Brooks DG, Goldstein A, Caldwell J, Finsecke E, Maria BL, Holden K, Cruse RP, Swoboda KJ, Viskochil D. Expanding CEP290 mutational spectrum in ciliopathies. Am J Med Genet A. 2009;149a(10):2173-80.
19. Kumaran N, Pennesi ME, Yang P, Trzupek KM, Schlechter C, Moore AT, Weleber RG, Michaelides M. Leber Congenital Amaurosis / Early-Onset Severe Retinal Dystrophy Overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
20. Coppieters F, Lefever S, Leroy BP, De Baere E. *CEP290*, a gene with many faces: mutation overview and presentation of *CEP290*. Human Mutation. 2010;31(10):1097-108.
21. Adams M, Simms RJ, Abdelhamed Z, Dawe HR, Szymanska K, Logan CV, Wheway G, Pitt E, Gull K, Knowles MA, Blair E, Cross SH, Sayer JA, Johnson CA. A meckelin-filamin A interaction mediates ciliogenesis. Human molecular genetics. 2012;21(6):1272-86.
22. Chang B, Khanna H, Hawes N, Jimeno D, He S, Lillo C, Parapuram SK, Cheng H, Scott A, Hurd RE, Sayer JA, Otto EA, Attanasio M, O'Toole JF, Jin G, Shou C, Hildebrandt F, Williams DS, Heckenlively JR, Swaroop A. In-frame deletion in a novel centrosomal/ciliary protein CEP290/NPHP6 perturbs its interaction with RPGR and results in early-onset retinal degeneration in the rd16 mouse. Hum Mol Genet. 2006;15(11):1847-57.
23. Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, Alfadhel M, Lewis RA, Eyaid W, Banin E, Dollfus H, Beales PL, Badano JL, Katsanis N. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nature Genetics. 2008;40:443.
24. Brancati F, Camerota L, Colao E, Vega-Warner V, Zhao X, Zhang R, Bottillo I, Castori M, Caglioti A, Sangiuolo F, Novelli G, Perrotti N, Otto EA. Biallelic variants in the ciliary gene TMEM67 cause RHYNS syndrome. European journal of human genetics : EJHG. 2018;26(9):1266-71.
25. Schmidts M, Arts HH, Bongers EM, Yap Z, Oud MM, Antony D, Duijkers L, Emes RD, Stalker J, Yntema JB, Plagnol V, Hoischen A, Gilissen C, Forsythe E, Lausch E, Veltman JA, Roeleveld N, Superti-Furga A, Kutkowska-Kazmierczak A, Kamsteeg EJ, Elçioğlu N, van Maarle MC, Graul-Neumann LM, Devriendt K, Smithson SF, Wellesley D, Verbeek NE, Hennekam RC, Kayserili H, Scambler PJ, Beales PL, Knoers NV, Roepman R, Mitchison HM. Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement. J Med Genet. 2013;50(5):309-23.
26. Baujat G, Huber C, El Hokayem J, Caumes R, Do Ngoc Thanh C, David A, Delezoide AL, Dieux-Coeslier A, Estournet B, Francannet C, Kayirangwa H, Lacaille F, Le Bourgeois M,

Martinovic J, Salomon R, Sigaudy S, Malan V, Munnich A, Le Merrer M, Le Quan Sang KH, Cormier-Daire V. Asphyxiating thoracic dysplasia: clinical and molecular review of 39 families. J Med Genet. 2013;50(2):91-8.
27. Huber C, Wu S, Kim AS, Sigaudy S, Sarukhanov A, Serre V, Baujat G, Le Quan Sang KH, Rimoin DL, Cohn DH, Munnich A, Krakow D, Cormier-Daire V. WDR34 mutations that cause short-rib polydactyly syndrome type III/severe asphyxiating thoracic dysplasia reveal a role for the NF-κB pathway in cilia. Am J Hum Genet. 2013;93(5):926-31.
28. Schmidts M, Vodopiutz J, Christou-Savina S, Cortés CR, McInerney-Leo AM, Emes RD, Arts HH, Tüysüz B, D'Silva J, Leo PJ, Giles TC, Oud MM, Harris JA, Koopmans M, Marshall M, Elçioglu N, Kuechler A, Bockenhauer D, Moore AT, Wilson LC, Janecke AR, Hurles ME, Emmet W, Gardiner B, Streubel B, Dopita B, Zankl A, Kayserili H, Scambler PJ, Brown MA, Beales PL, Wicking C, Duncan EL, Mitchison HM. Mutations in the gene encoding IFT dynein complex component WDR34 cause Jeune asphyxiating thoracic dystrophy. Am J Hum Genet. 2013;93(5):932-44.
29. Halbritter J, Porath JD, Diaz KA, Braun DA, Kohl S, Chaki M, Allen SJ, Soliman NA, Hildebrandt F, Otto EA. Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. Hum Genet. 2013;132(8):865-84.
30. Stokman M, Lilien M, Knoers N. Nephronophthisis. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle. Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
31. Parisi MA, Bennett CL, Eckert ML, Dobyns WB, Gleeson JG, Shaw DW, McDonald R, Eddy A, Chance PF, Glass IA. The NPHP1 gene deletion associated with juvenile nephronophthisis is present in a subset of individuals with Joubert syndrome. Am J Hum Genet. 2004;75(1):82-91.
32. Bruel AL, Franco B, Duffourd Y, Thevenon J, Jego L, Lopez E, Deleuze JF, Doummar D, Giles RH, Johnson CA, Huynen MA, Chevrier V, Burglen L, Morleo M, Desguerres I, Pierquin G, Doray B, Gilbert-Dussardier B, Reversade B, Steichen-Gersdorf E, Baumann C, Panigrahi I, Fargeot-Espaliat A, Dieux A, David A, Goldenberg A, Bongers E, Gaillard D, Argente J, Aral B, Gigot N, St-Onge J, Birnbaum D, Phadke SR, Cormier-Daire V, Eguether T, Pazour GJ, Herranz-Pérez V, Goldstein JS, Pasquier L, Loget P, Saunier S, Mégarbané A, Rosnet O, Leroux MR, Wallingford JB, Blacque OE, Nachury MV, Attie-Bitach T, Rivière JB, Faivre L, Thauvin-Robinet C. Fifteen years of research on oral-facial-digital syndromes: from 1 to 16 causal genes. J Med Genet. 2017;54(6):371-80.
33. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualiza- tion and exploration. Brief Bioinform. 2013;14:178-92.
34. Kent WJ. BLAT-the BLAST-like Alignment Tool. Genome Res. 2002;12:656-64.
35. Delvallée C, Nicaise S, Antin M, Leuvrey AS, Nourisson E, Leitch CC, et al. A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. Clin Genet. 2021;99:318-324.