Molecular diagnoses in the congenital malformations caused by ciliopathies cohort of the 100,000 Genomes Project

Sunayna Best,1,2 Jenny Lord,3,4 Matthew Roche,5 Christopher M Watson 6,7 James A Poulter 1, Roel P J Bevers,8 Alex Stuckey,8 Katarzyna Szymanska,1 Jamie M Ellingford 9,10 Jenny Carmichael,11 Helen Brittain,8 Carmel Toomes,1 Chris Inglehearn,1 Colin A Johnson,1 Gabrielle Wheway 12 Genomics England Research Consortium

ABSTRACT
Background Primary ciliopathies represent a group of inherited disorders due to defects in the primary cilium, the ‘cell’s antenna’. The 100,000 Genomes Project was launched in 2012 by Genomics England (GEL), recruiting National Health Service (NHS) patients with eligible rare diseases and cancer. Sequence data were linked to Human Phenotype Ontology (HPO) terms entered by recruiting clinicians.

Methods Eighty-three prescreened probands were recruited to the 100,000 Genomes Project suspected to have congenital malformations caused by ciliopathies in the following disease categories: Bardet-Biedl syndrome (n=45), Joubert syndrome (n=14) and ‘Rare Multisystem Ciliopathy Disorders’ (n=24). We implemented a bespoke variant filtering and analysis strategy to improve molecular diagnostic rates for these participants.

Results We determined a research molecular diagnosis for n=43/83 (51.8%) probands. This is 19.3% higher than previously reported by GEL (n=27/83 (32.5%)). A high proportion of diagnoses are due to variants in non-ciliopathy disease genes (n=19/43, 44.2%) which may reflect difficulties in clinical recognition of ciliopathies. n=11/83 probands (13.3%) had at least one causative variant outside the tiers 1 and 2 variant prioritisation categories (GEL’s automated triaging procedure), which would not be reviewed in standard 100,000 Genomes Project diagnostic strategies. These include four structural variants and three predicted to cause non-canalicular splicing defects. Two unrelated participants have biallelic likely pathogenic variants in LRRK45S, a putative novel ciliopathy disease gene.

Conclusion These data illustrate the power of linking large-scale genome sequence to phenotype information. They demonstrate the value of research collaborations in order to maximise interpretation of genomic data.

INTRODUCTION
Ciliopathies represent a group of inherited genetic disorders that arise as a result of defects in the primary cilium, the ‘cell’s antenna’,1 or motile cilia, organelles responsible for the movement of fluid over the surface of cells.2 They encompass a range of severe developmental and degenerative diseases that are individually rare but collectively common, affecting an estimated 15.8 million people worldwide including an estimated 133,000 people in the UK. Cilia have also been implicated in conditions such as diabetes, cancer, congenital heart disease and osteoarthritis.3-5 As cilia have a near-ubiquitous anatomical distribution, genetic defects affecting the structure or function of cilia cause a range of conditions that can affect multiple organs. Ciliopathies are typically classified into: retinal ciliopathies that exclusively or predominantly affect the eye; renal ciliopathies, which include autosomal dominant polycystic kidney disease affecting around 1:500 people; skeletal ciliopathies that cause a diverse range of skeletal dysplasias and cranio-facial dysmorphology;5 metabolic or ‘obesity’ ciliopathies;2 neurodevelopmental ciliopathies;6 and the respiratory motile ciliopathies.11 It is estimated that around 1000 genes contribute to ciliogenesis and cilium function,12-15 and ciliopathies are highly genetically heterogeneous.16-17 Approximately one-third of the around 270 genes implicated in inherited retinal dystrophies are cilia genes,18 whereas roughly 20 genes have been associated with renal ciliopathies (PKD OMIM phenotypic type PS173900; nephronophthisis OMIM PS256100). The short-rib polydactyly syndromes, which encompass most of the skeletal ciliopathies, have 22 known genetic causes (OMIM PS208500). There are 24 known genetic causes of the metabolic/obesity ciliopathy Bardet-Biedl syndrome (BBS) (OMIM PS209900). In this same series, Alström syndrome is unusual, because it is a single gene ciliopathy (caused by pathogenic variants in ALMS1). There is extensive genetic overlap between neurodevelopmental ciliopathies Joubert syndrome (JBTS) and Meckel-Gruber syndrome (MKS), with 37 known JBTS genes (OMIM PS213300) and 13 MKS genes (OMIM PS249000), many of which also cause JBTS. Several MKS and JBTS disease genes also overlap with the nine genes known to cause complex multigorgan ciliopathy orofacial digital syndrome (OFD) (OMIM PS311200). OFD is considered by some to be a skeletal ciliopathy, involving malformations of the face, mouth and digits, while OFD type 1, which specifically includes...
Diagnostics

polycystic kidney disease, may be considered a renal ciliopathy. In total, at least 220 different genes have been shown to cause a single (or multiple) ciliopathy when mutated.

The number of identified ciliopathy disease genes has advanced rapidly since the early to mid-2010s following the ubiquitous implementation of next-generation sequencing (NGS) technologies. Using targeted gene panel, or whole exome sequencing (WES) approaches, genetic diagnosis rates for syndromic primary (non-motile) ciliopathies are typically 40%–70% and for motile (respiratory ciliopathies) are approximately 70% (studies summarised in online supplemental table 1). A recent large whole genome sequencing (WGS) study in 125 families with ciliopathies achieved an 87% diagnosis rate, and a further increase was achieved following the inclusion of structural variant (SV) analysis and RNA sequencing in carefully phenotyped cohorts.19

The 100,000 Genomes project is a hybrid clinical/research initiative, launched in 2012 and overseen by Genomics England Ltd (GEL), a company set up and wholly owned by the UK Government Department of Health and Social Care.20 The project aimed to sequence 100,000 genomes from 70,000 individuals with rare diseases and cancer. Rare disease patients’ genomes were sequenced alongside their family members in a trio testing approach. Cancer patients’ germline and somatic genomes were sequenced from matched tumour and normal tissue. Genome sequence data were linked to clinical data from longitudinal patient records and Human Phenotype Ontology (HPO) terms entered by recruiting clinicians. Participants consented to receive a diagnosis for the specific condition they were recruited to the project for and to allow access to their fully anonymised genome sequence data and phenotype information for approved academic and commercial researchers. Recruitment to 190 different rare disease domains took place between 2016 and 2018 across 85 NHS Trusts, coordinated by 13 Genomic Medicine Centres (GMCs). In the data release used in this study (Main Programme Release 11 (17 December 2020)), data were available for 88,918 individuals: 71,682 in the rare diseases arm of the 100,000 Genomes Project and 17,236 in the cancer arm. In the rare diseases arm, 33,329 participants were entered as probands and 38,352 as relatives.

GEL also developed PanelApp (available from https://panelapp.genomicsengland.co.uk), a crowdsourcing tool for sharing and evaluation of gene panels by the scientific community.21 Virtual gene panels were applied to WGS data to facilitate focused analysis, returning variants in selected genes on curated lists with convincing evidence of an association with the disease(s) of interest. Not only does this shorten the list of variants to analyse, but it also reduces the risk of unwanted incidental findings.

As part of the effort to integrate NGS into standard of care (SOC) testing in the UK’s National Health Service (NHS), ciliopathy patients who had previously undergone existing SOC testing (typically gene panel testing) were recruited to the 100,000 Genomes Project to undergo WGS.22 Patients recruited under congenital malformations caused by ciliopathies (CMC) categories (subdivided into BBS, JBTS and rare multisystem ciliopathy disorders (RMCD) or respiratory ciliopathies) accounted for just under 1% of the total rare disease cohort. There were no dedicated recruitment categories for retinal ciliopathies, renal ciliopathies or skeletal ciliopathies, and these were recruited under subcategories of ophthalmological disorders, renal and urinary tract disorders or other categories, and so there are likely to be many further ciliopathy participants in the rare disease cohort. In this study, we aimed to optimise strategies to improve molecular diagnostic rates for probands recruited to the CMC category within the 100,000 Genomes Project.

MATERIALS AND METHODS

Participant selection and phenotypic classification

Participants recruited under CMC categories were extracted from the GEL Main Programme Release 11 (17 December 2020) using the user interface ‘LabKey’ within the GEL secure research environment. All data analysis was conducted within the GEL Research Environment. We exported anonymised data for publication through the Airlock system, after review by the GEL Airlock Review Committee. HPO terms recorded for each participant by their recruiting clinicians were assessed within the research environment prior to genetic analysis to determine the most likely clinical diagnosis for each proband based on phenotypic features alone. For selected cases, further clinical information was obtained through the ‘Participant Explorer’ interface.

Variant filtering and analysis

The GEL data processing pipeline, which includes an automated variant triaging algorithm to classify variants into a series of ‘Tiered’ categories (as defined by the Genomics England Rare Disease Tiering Process), has been described previously.22 Variants were tiered against ‘green’ genes listed in PanelApp panels selected according to entered HPO terms. PanelApp provides a traffic light system for genes: ‘green’ genes are diagnostic grade, ‘amber’ genes are borderline and ‘red’ genes have a low level of evidence. In instances where tiered variants did not indicate the cause of disease, untiered single nucleotide variants (SNVs) including heterozygous variants were extracted from participant genomes using a custom Python script (‘find_variants_by_gene_and_consequence.py’; available at https://github.com/JLord86/Extract_variants). The script extracts variants in diagnostic grade ‘green’ genes from provided PanelApp panels and candidate genes with the variant effect predictor (VEP) annotations stop gained, splice_acceptor, splice_donor, frameshift, missense and splice_region (if the variant was within either the terminal 1–3 bases of the exon or terminal 3–8 bases of the intron).

The script was first run using the RMCD Super Panel V.4.91 (available from https://panelapp.genomicsengland.co.uk/panels/728/) (green genes recorded in online supplemental table 2) and ciliopathy candidate genes from several sources. These include all ‘red’ and ‘amber’ genes from the PanelApp RMCD panel, genes of interest highlighted by local research teams and all genes on the curated SYSCILIA gold standard (SCGsv1) (online supplemental table 3). If a single potentially pathogenic heterozygous SNV in a recessive gene was identified through this strategy, manual inspection of the whole gene locus was undertaken using the Integrative Genomics Browser (IGV)23 to determine if a potential SV could be identified as the second biallelic variant. SVs were considered potentially causative if present in >30% of reads.

For those cases that remained unsolved, untiered SNVs were then extracted using further panels compatible with the participant’s phenotype. These included: the Retinal Disorders panel V.2.172 for those with retinal dystrophy only (available from https://panelapp.genomicsengland.co.uk/panels/3077), the Developmental Disorders Genotype-to-Phenotype database (DDG2P) panel V.2.21 for those with multisystem developmental disorders (https://panelapp.genomicsengland.co.uk/panels/484), the Laterality Disorders and Isomerism panel V.1.21 for those with a laterality defect (https://panelapp.genomicsengland.co.uk/panels/549) and the Broad Renal Super
For all remaining unsolved participants, variants potentially affecting splicing (SpliceAI delta scores >0.5) in diagnostic grade ‘green’ genes from the PanelApp RMCD panel were extracted with a further custom Python script ('find_variants_by_gene_and_SpliceAI_score.py'); available at https://github.com/JLord86/Extract_variants. Finally, the find_variants_by_gene_and_SpliceAI_score.py Python script was run again using the DDG2P panel V.2.21 for all remaining unsolved participants.

Bespoke research variant analysis pipeline

All data analysis was conducted within the secure online Research Environment including interrogation of BAM, VCF, SV and HPO information files. The Ensembl VEP was used to obtain variant information for interpretation of variant pathogenicity. Information about associations between genes and disease phenotypes was obtained from the OMIM database (https://www.omim.org). The mode of inheritance was defined according to the literature and OMIM for each gene. Variant evidence was reviewed using ACMG/AMP guidelines for clinical variant interpretation, and each variant of interest was assigned a pathogenicity score according to current (Association for Clinical Genomic Science (ACGS) guidelines)

The research analysis workflow comprised steps to filter genomic data (figure 1A), assess putative pathogenic variants (figure 1B), then classify and assign diagnostic confidence (figure 1C).

Variant classification and diagnostic confidence

To benchmark our ability to appropriately classify and interpret identified variants, first-pass analysis was blinded to previous results, and then verified against the GEL reported findings in the GMC exit questionnaires. These were completed by regional NHS GMCs for each analysed participant. Recruiting clinicians were contacted through the GEL secure airlock system for notification of a research molecular diagnoses, if they did not have a consistent completed GMC exit questionnaire. Additional clinical data were requested, where required, using the ‘contact the clinician’ form. All diagnoses identified through this blinded research strategy were termed ‘research molecular diagnoses’. The interpretation of these findings was subdivided into ‘confident’, ‘probable’ or ‘possible’ according to the ACMG classification for each variant, the inheritance pattern of the identified condition and the match to the proband’s phenotypic features (summarised in figure 1C).

RESULTS

Congenital malformations caused by ciliopathies cohort

A total of 83 probands were identified in the CMC cohort. This was subdivided into 45 in the BBS category, 14 in the JBTS category and 24 in the RMCD category. Fifteen participants were recruited as singleton cases, and for 68 individuals at least one additional family member underwent WGS. Including probands and relatives, genomic data were available for 211 individuals.

HPO term analysis

Analysis of HPO terms for the 83 probands shows that for 51 cases, phenotypes were consistent with their disease recruitment category. The remaining 32 probands lack recorded phenotypes suggestive of a syndromic ciliopathy (table 1). This suggests that participants were either frequently misdiagnosed as having ciliopathies or HPO terms were not entered accurately.

Tiered variants

Thirty-eight tier 1 variants were identified in 28 different genes among 29 different probands in the CMC cohort. Two hundred and sixteen tier 2 variants were identified in 142 different genes among 53 different probands. A total of 8777 tier 3 variants were identified in 5220 different genes among all 83 probands. No SVs had been tiered.

GEL reported molecular diagnoses

GMC exit questionnaires were completed for 67/83 (80.7%) patients by Release 11 (released 17 December 2020) (table 1). Twenty-three participants (27.7%) had GMC exit questionnaires reporting causative tier 1 or tier 2 variants, with one case partially solved and 22 fully solved. Four GMC exit questionnaires reported variants of uncertain significance (VUS) (figure 2A).

We identified that one of the cases previously reported as solved was a false positive. The GMC questionnaire reported compound heterozygous ALMS1 variants in participant #32 including an untiered heterozygous exon 11 deletion. The deletion was not visible using the IGV or detectable in the patients...
<table>
<thead>
<tr>
<th>Research number</th>
<th>Recruitment category</th>
<th>Most likely clinical diagnosis based on HPO terms</th>
<th>Does recruitment category match most likely clinical diagnosis?</th>
<th>GEL GMC exit report</th>
<th>Research molecular diagnosis</th>
<th>Gene</th>
<th>Is identified diagnosis a ciliopathy?</th>
<th>Diagnostic confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Sol</td>
<td>CHARGE Syn</td>
<td>CHD7</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>2</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Sol</td>
<td>Alström Syn</td>
<td>ALMS1</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>3</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>BBS + RP</td>
<td>ARL6 + IMPS2</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>4</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>RP</td>
<td>RPGR</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>5</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Sol</td>
<td>Retinal cil, possibly syndromic</td>
<td>CEP290</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>6</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Sol</td>
<td>JBTS</td>
<td>KIAAS86</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>7</td>
<td>RMCD</td>
<td>OFD-like cil</td>
<td>Yes</td>
<td>Sol</td>
<td>OFD1, PKD1, CYBB1</td>
<td>OFD1 Conf, PKD1 + CYBB1</td>
<td>Yes (OFD1)</td>
<td>Conf</td>
</tr>
<tr>
<td>8</td>
<td>BBS</td>
<td>Isol RD</td>
<td>No</td>
<td>Sol</td>
<td>RP</td>
<td>PRPP8</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>9</td>
<td>RMCD</td>
<td>JBTS-like MS cil</td>
<td>Yes</td>
<td>Uns</td>
<td>Seckel Syn</td>
<td>CEP192</td>
<td>No</td>
<td>Poss</td>
</tr>
<tr>
<td>10</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Sol</td>
<td>JBTS</td>
<td>CEP290</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>11</td>
<td>RMCD</td>
<td>Jeune-like cil</td>
<td>Yes</td>
<td>Unr</td>
<td>Feingold Syn</td>
<td>MYCN</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>12</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>JBTS</td>
<td>ARMC9</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>13</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>Tubulinopathy</td>
<td>TUBA1A</td>
<td>No</td>
<td>Poss</td>
</tr>
<tr>
<td>14</td>
<td>RMCD</td>
<td>Jeune-like cil</td>
<td>Yes</td>
<td>Unr</td>
<td>Jeune Syn</td>
<td>WDR19</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>15</td>
<td>BBS</td>
<td>Isol RD</td>
<td>No</td>
<td>Unr</td>
<td>RP</td>
<td>RHO</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>16</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>VUS</td>
<td>STAG1 syndromic ID syn</td>
<td>STAG1</td>
<td>No</td>
<td>Prob</td>
</tr>
<tr>
<td>17</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>BBS</td>
<td>BBS1</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>18</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>Neurodevelopmental disorder</td>
<td>RERE</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>19</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>Alström Syn</td>
<td>ALMS1</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>20</td>
<td>BBS</td>
<td>Isol eye cond (not RD)</td>
<td>No</td>
<td>Sol</td>
<td>BBS</td>
<td>BBS2</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>21</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>Poretti-Boltshauser Syn+ArboledaTham Syn</td>
<td>LAMA1, KAT6A</td>
<td>No LAMA1 Prob, KAT6A Poss</td>
<td>Conf</td>
</tr>
<tr>
<td>22</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>BBS</td>
<td>MKKS</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>23</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Sol</td>
<td>JBTS</td>
<td>CEP290</td>
<td>Yes</td>
<td>Prob</td>
</tr>
<tr>
<td>24</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>Smith Magenis Syn</td>
<td>RA11</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>26</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>Cone-rod dystrophy</td>
<td>PROM1</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>27</td>
<td>JBTS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>Luscan-Lumish Syn</td>
<td>SETD2</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>28</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>Optic Atrophy</td>
<td>ORA1</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>29</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Sol</td>
<td>Alström Syn</td>
<td>ALMS1</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>30</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>Chung-Jansen Syn</td>
<td>PHP</td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>31</td>
<td>BBS</td>
<td>Isol RD</td>
<td>No</td>
<td>Sol</td>
<td>Cone-rod dystrophy</td>
<td>RAB28</td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>32</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>None: Unsolved</td>
<td>ALMS1</td>
<td>N/a False+ve</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>Van Esch-O’Driscoll Syn</td>
<td>POLA1</td>
<td>No</td>
<td>Poss</td>
</tr>
<tr>
<td>35</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>JBTS</td>
<td>C3PP1</td>
<td>Yes</td>
<td>Prob</td>
</tr>
<tr>
<td>42</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>JBTS</td>
<td>PIBF1</td>
<td>Yes</td>
<td>Prob</td>
</tr>
<tr>
<td>43</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>BBS</td>
<td>Isol polydactyly</td>
<td>No</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>RMCD</td>
<td>MKS/JBTS-like MS cil</td>
<td>Yes</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>RMCD</td>
<td>BBS-like MS cil</td>
<td>Yes</td>
<td>Uns</td>
<td>Candidate cil</td>
<td>LRRC45</td>
<td>Candidate</td>
<td>Poss</td>
</tr>
<tr>
<td>49</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>RMCD</td>
<td>DM</td>
<td>DM</td>
<td>Unr</td>
<td>Uns</td>
<td>Uns</td>
<td>Uns</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Diagnostics

Identification of research molecular diagnoses

Our bespoke variant-to-diagnosis pipeline shows that 43 of the 83 probands (51.8%) have a research molecular diagnosis that is compatible with their phenotypic features (table 1). Individual variant information, including data taken into consideration in performing ACMG classification, is recorded in online supplemental table 4. Twenty-eight of the 83 participants (33.7%) are classified as having a confident diagnosis, 5/83 (6%) a probable diagnosis and 10/83 (12%) only a possible diagnosis (figure 2B). Overall, 34/83 participants (41%) had a research molecular diagnosis that fully accounted for their entered phenotypic features and 9/83 (10.8%) that partially accounted for their entered features (online supplemental table 4). No phenotypic features were entered for proband #75, but the possible molecular diagnosis of BBS matches their BBS recruitment category. Diagnoses according to recruitment category are shown in figure 2C.

Table 1

<table>
<thead>
<tr>
<th>Research number</th>
<th>Recruitment category</th>
<th>Most likely clinical diagnosis based on HPO terms</th>
<th>Does recruitment category match most likely clinical diagnosis?</th>
<th>GEL GMC exit report</th>
<th>Research molecular diagnosis</th>
<th>Gene</th>
<th>Is identified diagnosis a ciliopathy?</th>
<th>Diagnostic confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>RMCD</td>
<td>JBTS-like MS cil</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS BBS9</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>53</td>
<td>RMCD</td>
<td>Isol GI disorder</td>
<td>No</td>
<td>Unr</td>
<td>BBS</td>
<td></td>
<td>Unr</td>
<td>Uns</td>
</tr>
<tr>
<td>54</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>ALMS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>55</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>56</td>
<td>BBS</td>
<td>Isol eye cond (not RD)</td>
<td>No</td>
<td>VUS</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>57</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>58</td>
<td>RMCD</td>
<td>JBTS-like MS cil</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>59</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>60</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>61</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>WT1-related disorder</td>
<td></td>
<td>No</td>
<td>Conf</td>
</tr>
<tr>
<td>62</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>BBS</td>
<td></td>
<td>Unr</td>
<td>Uns</td>
</tr>
<tr>
<td>63</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>BBS</td>
<td></td>
<td>Unr</td>
<td>Uns</td>
</tr>
<tr>
<td>64</td>
<td>RMCD</td>
<td>JBTS-like MS cil</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS</td>
<td></td>
<td>Unr</td>
<td>Uns</td>
</tr>
<tr>
<td>65</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>66</td>
<td>RMCD</td>
<td>BBS-like MS cil</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>67</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>VUS</td>
<td>Alström Syn</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>68</td>
<td>JBTS</td>
<td>JBTS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>69</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Sol</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>70</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Uns</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>71</td>
<td>RMCD</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>72</td>
<td>BBS</td>
<td>Isol GI disorder</td>
<td>No</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>73</td>
<td>RMCD</td>
<td>Isol GI disorder</td>
<td>No</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>74</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>75</td>
<td>BBS</td>
<td>DM DM</td>
<td>No</td>
<td>Uns</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
<tr>
<td>76</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>VUS</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>77</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>78</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>79</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>80</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>81</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Poss</td>
</tr>
<tr>
<td>82</td>
<td>BBS</td>
<td>Non-cil MS cond</td>
<td>No</td>
<td>Unr</td>
<td>Attenuated mucopolysaccharidosis 1</td>
<td></td>
<td>No</td>
<td>Prob</td>
</tr>
<tr>
<td>83</td>
<td>BBS</td>
<td>BBS</td>
<td>Yes</td>
<td>Unr</td>
<td>BBS1</td>
<td></td>
<td>Yes</td>
<td>Conf</td>
</tr>
</tbody>
</table>

Table includes the recruitment category, designated ‘most likely’ clinical diagnosis based on entered HPO terms alone, GEL GMC exit questionnaire reporting outcome, research molecular diagnosis (determined by genotype), responsible gene, whether the identified diagnosis is a ciliopathy and diagnostic confidence. Note: individual variant information, including data taken into consideration in forming ACMG classifications, can be found in online supplemental table 4.

BBS, Bardet-Biedl syndrome; Cil, ciliopathy; Cond, condition; Conf, confident; DM, data missing; GEL, Genomics England; GI, gastrointestinal; GMC, Genomic Medicine Centres; HPO, Human Phenotype Ontology; Isol, isolated; JBTS, Joubert syndrome; MKS, Meckel Gruber syndrome; MS, multisystemic; PKD, polycystic kidney disease; Poss, possible; Prob, probable; RD, retinal dystrophy; RMCD, rare multisystem ciliopathy disorders; RP, retinitis pigmentosa; Sol, solved; Syn, syndrome; Unr, unreported; Uns, unsolved.

VCF file; following correspondence with the GEL helpdesk, the variant was confirmed to be a false positive.

Our bespoke variant-to-diagnosis pipeline shows that 43 of the 83 probands (51.8%) have a research molecular diagnosis that is compatible with their phenotypic features (table 1). Individual variant information, including data taken into consideration in performing ACMG classification, is recorded in online supplemental table 4. Twenty-eight of the 83 participants (33.7%) are classified as having a confident diagnosis, 5/83 (6%) a probable diagnosis and 10/83 (12%) only a possible diagnosis (figure 2B). Overall, 34/83 participants (41%) had a research molecular diagnosis that fully accounted for their entered phenotypic features and 9/83 (10.8%) that partially accounted for their entered features (online supplemental table 4). No phenotypic features were entered for proband #75, but the possible molecular diagnosis of BBS matches their BBS recruitment category. Diagnoses according to recruitment category are shown in figure 2C.

Seventeen of the 43 research molecular diagnoses (39.5%) can be considered novel findings. Fourteen diagnoses are new findings in probands with no completed GMC exit questionnaire (unreported) and three are in probands with negative GMC outcome questionnaires (reported as ‘unsolved’). Interestingly, a significant proportion of research molecular diagnoses have been made in non-ciliopathy genes. Only 23 of the 43 potentially solved participants (53.5%) have variants in genes known to be causative of ciliopathy syndromes. The remaining 19/43 potentially solved probands (44.2%) have variants identified in non-ciliopathy genes.
was discovered via IGV inspection to be an indel (92 nucleotide
ALMS1

Interestingly, a variant annotated by GEL as a tier 2

SNVs identified through our bespoke filtering pipeline. Inter-

4. BBS, Bardet-Biedl syndrome; RMCD, rare multisystem ciliopathy disorder.

Three of the untiered

variations are SVs (IGV captures shown in figure

Figure 2

Comparison of diagnostic reporting outcomes between gel

GMC exit reports (A) and research diagnostic outcomes (B) for the 83

probands in the CMC cohort. (C) Research molecular diagnoses according
to recruitment category. Genes with identified potentially causative variants
are grouped according to whether they are known to be associated
ciliopathies or not. A ‘+’ is used where participants had potentially
causal variants in more than one gene contributing to their clinical
features (additional gene(s) are included in brackets). Diagnostic confidence
for each research molecular diagnosis is shown in table 1. Detailed variant
information, including whether the gene variant(s) is thought to be a
full or partial match to phenotype, is provided in online supplemental table
4. BBS, Bardet-Biedl syndrome; GMC, congenital malformations caused
by ciliopathies; GEL, Genomics England; GMC, Genomic Medicine Centre;
J B TS, Joubert syndrome; RMCD, rare multisystem ciliopathy disorder.

Research molecular diagnoses made outside GEL tiers 1 and 2
Thirty-two of the 83 probands (38.5%) have research molecular diagnoses
made from tier 1 and 2 variants only. The remaining
11/83 probands (13.3%) with research molecular diagnoses have
at least one variant outside of tiers 1 and 2 (variant information
provided in online supplemental table 4). These diagnoses would
have been missed by the standard 100,000 Genomes Project
diagnostic pipeline, which routinely inspects only tier 1 and 2
variants. Five tier 3 variants and 12 untiered variants contribute
to the diagnoses for these 11 participants. Three of the untiered
variants are SVs (IGV captures shown in figure 3); the other nine
are SNVs identified through our bespoke filtering pipeline.
Interestingly, a variant annotated by GEL as a tier 2 ALMS1 missense
was discovered via IGV inspection to be an indel (92 nucleotide
deletion and 31 nucleotide insertion) leading to a splice acceptor change (participant #29, shown in figure 3A).

SpliceAI analysis of variants filtered using our pipeline iden-
tified three untiered ciliopathy gene variants predicted to cause
splice donor site losses. One is a homozygous synonymous variant in
ARL6 in proband #3, entered with suspected BBS
(NM_001278293.3:c.534A>G, NP_001265222.1:p.Gln178=)
(online supplemental table 4). The overall allele frequency (AF)
on gnomAD is 0.000007960 with zero homozygotes.29 The
100,000 Genomes Project AF is 0.00049985 for participants
called on GrCh37 (one heterozygote) and 0.0000571872 for participants
called on GrCh38 (three heterozygotes and three homozygotes). On further analysis, the two further homozygous
individuals were identified as affected siblings of proband #3. The
heterozygous individuals are the parents of proband #3 plus one
unrelated participant. This variant has previously been published
in association with BBS and proven to cause aberrant splicing in
vitro by minigene assay.29 The other two are at +3 and +5 positions
in probands #75 (BBS4 NM_033028.5:c.642+3>A+T) and
#41 (CSPP1 NM_001382391.1:c.2968+5G>A). Clinical
material was not available for testing to validate splicing effects at
the molecular level. Therefore, both have been classified as VUSs.

Putative novel disease genes
Participant #48, entered to the RMCD category and deter-
mined most likely to have BBS based on entered HPO terms,
has two separate homozygous, protein-truncating variants in
candidate ciliopathy genes. Proband #48 has a sibling who was
separately entered to the 100,000 Genomes Project in the intel-
lectual disability category, without additional features suggestive
of a syndromic ciliopathy. Further phenotypic analysis using the
Participant Explorer tool revealed that participant #48 also has
clinical features suggestive of a motile ciliopathy. Specific clinical
features cannot be provided to protect participant anonymity.
There is a recorded history of parental consanguinity in this
family.

The first variant of interest identified in participant #48 is a
homozygous frameshift variant in LRRCA5 (GrCh38 chromosome
17: 82028260 C<GTG; NM_144999.4:c.1074_1075insTG,
NP_659436.1:p.Leu359CysTer19). This was also found to
be homozygous in the proband’s sibling from the intellectual
disability category. Segregation analysis is consistent with auto-
sonal recessive inheritance; both parents are confirmed hetero-
ygotes. According to the Illumina Region of Homozygosity
(ROH) caller, this LRRCA5 variant is in a 1.359.569 base pair ROH
(GrCh38 chromosome 17: 81841582–83201151) containing
797 homozygous and zero heterozygous variants (ROH score
19.92) in the proband and an 1.364.960 base pair ROH (GrCh38
chromosome 17: 81841582–83206342) containing 728 homo-
ygous and zero heterozygous variants (ROH score 18.2) in the
sibling. The second variant of interest is a homozygous stop gain variant in
CAPF45 (CCDC19) (GrCh38 chromosome 1: 159887996 G>A;
NM_012337.3:c.433C>T; NP_064692.2:p.Arg145Ter) (online supplemental table 4). Segregation analysis
showed again that the parents are both heterozygotes but the
sibling in the intellectual disability category is homozygous for
the reference allele. This CAPF45 variant is in a 8142476bp
ROH (GrCh38 chromosome 1: 158386429–166528905)
containing 3821 homozygous and zero heterozygous variants
(ROH score 95.53), not present in the sibling.

Next, we searched for other biallelic, potentially causative
variants in either LRRCA5 or CAPF45 across the entire rare
disease 100,000 genomes dataset to gain independent replication
of causality. No additional potentially pathogenic variants were identified for CFAP45. However, we identified a second proband with LRRC45 variants within the cone-rod dystrophy recruitment category and with an ‘unsolved’ GMC exit questionnaire. We identified a heterozygous LRRC45 start loss variant: NM_144999.4:c.1A>T, NP_659436.1:p.Met1? (absent from gnomAD, GEL 100K MAF 1.271 \times 10^{-5}), and a heterozygous splice acceptor variant: NM_144999.4:c.1126–1G>A (gnomAD allele frequency 8.059 \times 10^{-6}, GEL 100K MAF 2.542 \times 10^{-5}). The proband was entered as a singleton participant, so parental sequence is not available in the 100,000 Genomes Project or on clinician request to establish phase. LRRC45 therefore remains a putative novel disease gene accounting for the phenotype in these individuals.

**DISCUSSION**

**Diagnosis rate for participants in the CMC cohort of the 100,000 Genomes Project**

This study provides a research molecular diagnosis from WGS data for just over half of the participants in the CMC cohort of the 100,000 Genomes Project (43/83, 51.8%), 33 of which are classified as confident or probable (39.8%). Our overall diagnosis rate is 19.3% higher than the 27/83 (32.5%) with GEL reported findings in GMC exit questionnaires (23/83 reported as solved plus 4/83 with VUSs). It is likely that at least nine of the novel research molecular diagnoses would eventually be made and reported by GEL given that they contain only tier 1 or 2 variants (participants #11, #12, #13, #14, #15, #21, #27, #72 and #75). In identifying and alerting clinical teams, we are providing benefit to participants who have, in some cases, been waiting years for identification of a molecular diagnosis (recruitment to the 100,000 Genomes Project ended in 2018).

There are 11 participants with research molecular diagnoses with at least one variant outside of tiers 1 and 2, which would be missed by the standard diagnostic strategy of inspecting only those variants. Therefore, the added diagnostic value of undertaking analyses outside tiers 1 and 2 is at least 11/83 (13.3%). This highlights the value of research collaborations to investigate unsolved cases and improved diagnosis rates from accessible genomic data.

Unfortunately, major challenges remain in returning research identified diagnoses to recruiting clinicians to ensure they are successfully fed back to participants, which is being addressed with collaborators at GEL. Improved communication between recruiting clinicians and researchers would facilitate better interpretation of variants, but a lack of an automated system for researcher/clinician contact introduces a significant bottleneck, and the long time between recruitment and research identified molecular diagnosis has meant that some recruiting clinicians no longer work in the NHS trust and GMC where they recruited patients to the project, and there is no mechanism of forwarding emails in cases such as this. Recruiting clinician collaboration is
hugely valuable to provide additional clinical information where required, as well as contacting patients to ask for consent to publication of more detailed clinical data. Furthermore, they can obtain relevant tissue samples to validate variant effects, particularly useful for novel splice variants and SVs.

**Conditions identified**

Among probands in the CMC cohort with research molecular diagnoses, a surprisingly high proportion have causative variants in non-ciliopathy genes (19/43, 44.2%). This suggests that there are likely to be significant numbers of participants with ciliopathies recruited to other rare disease categories. This misdiagnosis rate may be because primary ciliopathies can be difficult to recognise clinically due to the great diversity of possible disease features. More specific ‘hard’ phenotypic features can assist healthcare professionals to the likelihood of a ciliopathy syndrome, but these are not always present.

The best example is the molar tooth sign, which is the pathognomonic sign for JBTS-related conditions with no differential diagnoses. This is reflected in the highest correlation between recruitment category and identified molecular diagnosis rate being for the JBTS group: 6/14 (42.9%) were recruited as suspected JBTS, and then confirmed to have JBTS at the molecular level. Ten of the 14 patients recruited with suspected JBTS had the HPO term ‘Molar Tooth Sign on MRI’ entered by the recruiting clinician, including all six that were solved at the molecular level.

Another reason for the high proportion of non-ciliopathy diagnoses could be limitations or difficulties in choosing appropriate recruitment categories for participants of the 100,000 Genomes Project. Categories may have been selected for convenience or lack of awareness of alternative, potentially more appropriate options. The RMCD category may have been treated as a ‘catch-all’ group for participants with constellations of multisystemic features, not obviously recognisable as a specific syndrome. This is reflected by this group having the lowest diagnosis rate of the three included in the CMC cohort: 9/24 (37.5%) have a research-identified molecular diagnosis, but only two are ciliopathies.

An important outcome to explore further is the relatively high number of participants recruited in the BBS category, found to have variants causative of isolated eye disorders (n=4). It is unclear if recruiting clinicians suspected BBS due to the presence of non-ocular features or whether the participants were inappropriately included in the BBS category. This problem clearly demonstrates the importance of accurate and comprehensive phenotyping to refine the interpretation of sequence variants.

**Mutational mechanism of causative variants**

Sixty-four individual, potentially causative variants, have been identified in this research study (online supplemental table 4). Of the variants detected, at least four would not have been detectable or accurately described by WES or gene panel, as they are SVs including significant intronic regions (figure 3). Ideally, all SVs of interest should be confirmed by long-range PCR and either third generation nanopore or Sanger sequencing, but DNA samples from these cases could not be obtained from referring clinicians. A recent study of NHS rare diseases patients undergoing WGS, reported 102 large deletions and six complex SVs from 1103 distinct causal variants (9.8% SVs). Our identified rate of SVs is slightly lower at 4/64 (6.3%). It seems likely that further SVs are responsible for a proportion of the unsolved participants in the CMC cohort, but strategies to detect them are not yet well established.

WGS, particularly PCR-free WGS, offers great advantages in SV analysis over WES, due to even coverage of the whole genome permitting reliable identification of SVs, but we are yet to fully take advantage of these methodologies. The GEL dataset is being used to improve the way we analyse SVs, with a gnomAD-type database of all SVs in GEL with allele frequencies in the cohort having been developed by Jing Yu in Oxford to permit exclusion of SVs from analysis in a patient if that SV appears above a particular minor allele frequency (MAF) in the GEL dataset. PCR-free WGS adds the further benefit of improved coverage of GC rich regions of the genome that are not efficiently amplified in PCR. As many promoter regions are GC rich, this provides an advantage for identifying regulatory region variants.

A further benefit of WGS over WES or gene panel testing is the opportunity to analyse intronic regions. We used the in silico tool SpliceAI to find variants predicted to cause novel splicing effects and identified three variants outside the canonical splice sites predicted to cause splice donor site defects. No novel splicing variants were identified in genes from the DDG2P gene panel using our SpliceAI script in unsolved participants of the cohort. However, given the diversity of diagnoses, it is highly likely that further causative splicing variants could be found in non-ciliopathy genes. As well as splice variant identification, intronic WGS data can also be interrogated for regulatory region variants implicated in human disease, using resources such as the UTRannotator tool to annotate high-impact 5' untranslated region variants either creating new upstream opening reading frames (ORFs) or disrupting existing upstream ORFs.

Despite the many advantages of WGS over WES, WES remains a popular sequencing strategy as it involves sequencing of only around 2% of the genome, significantly lowering costs of sequencing, permitting sequencing to greater depth on a limited budget, lowering demands on data storage, increasing analysis times and reducing workload for clinical scientists and researchers to process and interpret the significantly smaller number of identified variants. Furthermore, coding region variants are more straightforward to classify, making analysis of WES data more straightforward than analysis of WGS data.

**Candidate gene analysis**

A list of 302 candidate ciliopathy genes (online supplemental table 3) was used in conjunction with our custom variant filtering pipeline in pursuit of diagnosis for probands unsolved through tiered variant analysis. One proband, participant #48, has two homozygous, protein-truncating variants in the candidate ciliopathy genes LLRC45, a protein associated with distal appendages of the basal body that contributes to early steps of axoneme extension during ciliogenesis, and CFAP45, a coiled coil domain protein and expressed in nasopharyngeal epithelium and trachea.

There are various possibilities regarding the potential contribution of these variants to the clinical features of proband #48 and their sibling in the intellectual disability category. The two siblings share neurodevelopmental delay and intellectual disability. Proband #48 also has additional features in keeping with both syndromic primary and motile ciliopathies. CFAP45 has been recently published as a motile ciliopathy gene, so it is possible that the homozygous nonsense CFAP45 variant present in participant #48 but not their sibling could account for the clinical motile ciliopathy features in participant #48, with the LLRC45 variants accounting for the neurodevelopmental delay and intellectual disability in both siblings.
Given the phenotypic heterogeneity in ciliopathies even within families with the same variant, another hypothesis is that the two siblings have different presentations of a condition caused by their shared homozygous LRRC45 frame-shift variant. The putative loss of function (pLoF) gnomAD tolerant to loss of function. Analysis can be iterative such that the data can be ‘opened up’ beyond the first virtual gene panel without the need for serial testing. Results from this study demonstrate the value of this approach, given the high proportion of participants with non-ciliopathy diagnoses. The NHS Genomic Medicine service, introduced in 2018 as a follow on from the 100,000 Genomes Project, provides a curated National Genomic Test Directory including WES and WGS where appropriate. This will embed genomic testing into mainstream care and standardise testing across the country.

Determining the underlying genotype for a patient’s phenotype allows provision of accurate information about their condition, including potential current and future associated features for which screening or treatment may be available. An example of this in action is participant #61, recruited in the RMCD category. An untiered heterozygous missense variant in WT1 was identified through our filtering that is in the RMCD category. An untiered heterozygous missense variant in WT1 was identified through our filtering that is in the RMCD category. This diagnosis, which was successfully fed back to the recruiting clinician, is considered especially important given the associated risk of Wilms’ tumour and the recommendation for regular screening to facilitate early detection and treatment.

Lack of a genetic diagnosis can lead to inappropriate management of conditions and delays in accessing specialised services such as the multidisciplinary service for BBS and Alström syndrome in Birmingham Children’s Hospital and Great Ormond Street Hospital in the UK. Without greater awareness and higher diagnosis rates of ciliopathies, it may continue to be difficult to secure funding for additional specialist services for rare ciliopathies.

**Perspective on the future of genetic diagnosis**

This study prompts reconsideration of approaches to genetic diagnostics, particularly traditional forward genetics in comparison with reverse phenotyping. Classically, clinicians have suggested a possible underlying diagnosis based on the collection of clinical features observed, then the lab have tested for variants in gene(s) associated with that suspected diagnosis. This study demonstrates the utility of a reverse genetics strategy, by going ‘backwards’ from variants that are assessed as pathogenic at the molecular level, to determine if they could match with the patient’s features and the disease’s inheritance pattern. As the cost and availability of large-scale sequencing tests including WES and WGS continues to fall, this reverse phenotyping strategy is becoming increasingly integrated into NHS genetic diagnostics.

With this, the current bottleneck is clinical interpretation of variants. To realise the potential of WES and WGS, investment into dedicated time and resourcing for specialist variant interpretation is essential, as is careful and comprehensive phenotyping and strong communication between clinical scientists, clinical geneticists, mainstream clinicians and researchers. Improved integration of SV and splice variant analysis tools, such as SpliceAI, will be essential to maximise the diagnostic potential of WGS data beyond coding variants in exons of virtual panels of genes. The 19.3% genetic diagnosis uplift achieved in our study demonstrates what can be achieved with additional time and resources invested into WGS analysis. Now that this variant filtering and analysis pipeline has been established, we anticipate that this additional analysis can be achieved within days or weeks rather than months.

Clearly, large-scale genomic studies such as the 100,000 Genomes Project offer huge opportunities to improve diagnostics, understanding of disease mechanisms and identification of novel drug targets. The current challenge is to improve our strategies to analyse sequence data to provide the maximum benefit for patients and the scientific community.
Open access  This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Christopher M Watson http://orcid.org/0000-0003-2371-1844
James A Poulter http://orcid.org/0000-0003-2048-5693
Jamie M Ellifigton http://orcid.org/0000-0003-1137-9768

Gabrielle Wheway http://orcid.org/0000-0002-0494-0783

REFERENCES

1 Singh V, Reiter JF. The primary cilium as the cell’s antenna: signaling at a sensory organelle. Science 2006;313:629–33.


Contributors
Conceptualisation: SB, JL, CT, CFJ, CAJ, GW; Data curation: SB, JL, RPJB, AS, KS, JAP, CB, GW; Funding acquisition: SB, JL, CT, CFJ, CAJ, GW; Investigation: SB, JL, MR, RPJB, AS, JME; Project administration: SB, Genomics England Research Consortium, GW; Resources: SB, JL, MR, RPJB, AS, KS, JME, HC, JB, Genomics England Research Consortium; Supervision: CT, CFJ, CAJ, GW; Validation: JC, HB; Writing – original draft: SB, GW; Writing – review and editing: all authors; Guarantors: CAJ, GW.

Funding
SB acknowledges support from the Wellcome Trust 4Ward North Clinical PhD Academy (ref. 203914/2162). GW acknowledges support from Wellcome Trust Seed Award (ref. 204378/2162). CAJ acknowledges support from MRC project grants MR/M000532/1 and MR/T017503/1. JL is supported by an NIH Research Professorship awarded to Professor Diana Baralle (DB NIHR RP-grants MR/M000532/1 and MR/T017503/1. JL is supported by an NIHR Research Seed Award (ref. 204378/Z/16/Z). CAJ acknowledges support from MRC project funding.

Competing interests
Disclosure: HB, RPJB and AS are employed by Genomics England, UK. GW is employed by Illumina. The other authors declare no conflict of interest.

Patient consent for publication
Not applicable.

Ethics approval
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 Research (ref. GEL) Data Access Committee was provided by the GEL Access Review Committee, and research project RR185 ‘Study of cilia and ciliopathy 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 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 (ERGOM54400).

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
Data may be obtained from a third party and are not publicly available.

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


