High diagnostic rate of whole genome sequencing in primary ciliary dyskinesia

Holly A Black[1,2], Sophie Marion de Proce[1*], Jose L Campos[3*], Alison Meynert[3], Mihail Halachev[3], Joseph A Marsh[3], Robert A Hirst[4], Chris O'Callaghan[4], Scottish Genomes Partnership, Javier Santoyo-Lopez[5], Jennie Murray[2,3], Kenneth Macleod[6], Don S Urquhart[6,7], Stefan Unger[6,7], Timothy J Aitman[1^], Pleasantine Mill[3^]

[1] Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[2] South East of Scotland Genetics Service, Western General Hospital, Edinburgh, UK

[3] MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[4] Centre for PCD Diagnosis and Research, Department of Respiratory Sciences, University of Leicester, UK

[5] Edinburgh Genomics, Edinburgh, UK

[6] Department of Paediatric Respiratory and Sleep Medicine, Royal Hospital for Sick Children, Edinburgh, UK

[7] Department of Child Life and Health, University of Edinburgh, Edinburgh, UK

* These authors contributed equally to the manuscript

^ Joint senior authors

Summary: Whole genome sequencing (WGS) yielded a high genetic diagnostic rate (100%) in eight Scottish patients with clinically diagnosed primary ciliary dyskinesia (PCD) by detection of large structural variants, homology modelling and identification of a novel disease gene with a dominant mode of inheritance. Prioritised WGS may facilitate early genetic diagnosis in PCD.

Abstract

Aim: Primary ciliary dyskinesia (PCD) is a genetic disorder affecting motile cilia. Most cases are inherited recessively, due to variants in more than 50 genes that result in abnormal or absent motile cilia. This leads to chronic upper and lower airway disease, sub-fertility and laterality defects in some cases. Given overlapping clinical features and genetic heterogeneity, diagnosis can be difficult and often occurs late. Of those tested, an estimated 30% of genetically screened PCD patients still lack a molecular diagnosis. Here, we aimed to identify how readily a genetic diagnosis could be made in a clinically diagnosed population using whole genome sequencing (WGS) to facilitate identification of pathogenic variants in known genes as well as identify novel PCD candidate genes.

Methods: WGS was used to screen for variants causing PCD in 8 clinically diagnosed PCD patients, sequenced as trios where parental samples were available.

Results: Seven of the eight cases (87.5%) had homozygous or biallelic variants in *DNAH5*, *DNAAF4* or *DNAH11* that were classified as pathogenic or likely pathogenic. Three of the variants were deletions, ranging from 3kb to 13kb, for which WGS identified precise breakpoints, permitting confirmation by Sanger sequencing. WGS yielded a high genetic diagnostic rate from this clinically diagnosed population, in part through detection of structural variants as well as identification of a *de novo* variant in a novel PCD gene *TUBB4B*.

Conclusion: A molecular diagnosis allows for appropriate clinical management for cases and their families, including prediction of phenotypic features correlated to genotype. Here, WGS uplifted genetic diagnosis in cases of clinically diagnosed PCD by identifying structural variants and novel modes of inheritance in new candidate genes. Our study suggests that WGS could be a powerful part of the PCD diagnostic toolkit to increase the current molecular diagnostic yield from 70%. It provides important new insight into our understanding of fundamental biology of motile cilia as well as of variation in the non-coding genome in PCD.

Introduction

Primary ciliary dyskinesia (PCD, OMIM: PS244400) is a genetic disorder of motile cilia (1–3), highly structurally organised organelles that project from the surface of the cells. Their organised structure allows the cilia to beat in a coordinated manner to effectively propel fluids across the surface of the airways, brain ventricles and reproductive tracts. However, in PCD, these cilia have an abnormal structure or function, resulting in static cilia, cilia that beat in an uncoordinated manner, or a complete absence of cilia (2,3). This can result in a chronic respiratory disease, due to impaired mucociliary clearance, as well as hydrocephaly, laterality defects (e.g. situs inversus) and fertility defects in a subset of patients. PCD often presents at birth as unexplained neonatal respiratory distress. Affected children then present with variable clinical features including a daily wet cough, chronic respiratory tract infections, rhinitis, sinusitis and otitis media. Over time, these can lead to debilitating long-term complications, including bronchiectasis and hearing impairment (1–6).

However, PCD is a heterogeneous disorder that has significant phenotypic overlap with other genetic respiratory diseases, such as cystic fibrosis and primary immunodeficiency. The current clinical diagnostic work-up for suspected PCD cases requires a battery of highly specialised tests (7,8). This includes detection of low nasal nitric oxide (nNO), as well as high-speed video microscopy (HSVM) to assess cilia beating pattern and frequency and transmission electron microscopy (TEM) to assess the cilia ultrastructure from a nasal brush biopsy. Abnormalities identified in cilia structure and/or beating pattern or in the presence of clinical features such as chronic oto-sinopulmonary symptoms can be used to confirm a diagnosis of PCD (1,2,6,8,9). Access to diagnostic testing is limited to a small number of highly specialised units in the UK. Families must travel for testing, often over long distances and sometimes on multiple occasions. Even in those patients with a clinical diagnosis of PCD, 21% had normal ciliary ultrastructure (10). Therefore, additional sensitive and accessible diagnostic techniques acceptable to patients of all ages are required.

PCD is a genetically heterogeneous disorder. The majority of cases are autosomal recessively inherited, with over 50 genes currently associated with PCD (1,2,6,11,12). Mutations in these

genes account for approximately 70% of PCD cases screened, suggesting additional causal genes exist and/or more complicated structural variants (SVs) are being missed in known genes (13). The genes associated with PCD encode proteins involved in cilia structure, cilia assembly or regulatory complexes. There is often a correlation between the affected gene and the defects in cilia motility or ultrastructure observed at diagnosis (1,6,7,9). However, this is not always the case: it is estimated that 9-20% of PCD cases have normal or inconclusive ultrastructure by TEM (14). Similarly normal or non-diagnostic HSVM results complicate diagnosis for a significant portion of known PCD genes (15). nNO is also considered a useful screening test prior to referral for nasal brushing. While low nNO measurements may be indicative of PCD in patients lacking clear ultrastructural defects, its use, particularly in young children, has not been universally rolled out across PCD centres (16,17) and patients with mutations in at least some PCD genes show normal nNO levels (18,19). In such suspected cases, where a clinical diagnosis is unclear, a genetic diagnosis is an important means to confirm PCD (1,5,6,20). However, genetic testing often occurs late in the diagnostic work-up for PCD (7). Recent guidelines have recommended studies to investigate the utility of genetic testing in the PCD diagnostic pathway, particularly as next generation sequencing (NGS) is increasingly available (8,21,22).

In this study, we investigated the utility of whole genome sequencing (WGS) to detect disease-causing variants in children and young adults with a clinical diagnosis of PCD. We selected a non-endogamous population where PCD was clinically confirmed to test whether WGS could effectively improve the molecular diagnostic yield of PCD. WGS would provide unbiased genome coverage beyond exons of ~50 genes associated with PCD whilst simultaneously detecting a range of variants, from single nucleotide changes, small indels as well as complicated SVs, which may be missed by targeted sequencing approaches. Eight patients were recruited, alongside parental samples where available. This led to a genetic diagnosis for all eight cases, three of whom had a pathogenic deletion of 3 kb-13 kb and one patient had a de novo missense variant p.P259L in a novel candidate gene *TUBB4B*.

Materials and Methods

Patient cohort

Eight patients with a confirmed clinical diagnosis of PCD (50% female), aged 6 to 31 years, were recruited to the study. Signed and informed consent was obtained from the affected individual as well as relatives through approved protocols. Sample IDs were assigned and the key known only those within the research group. Clinical phenotypes are shown in **Table 1**. For details of blood samples, DNA extraction, and analysis of sample ethnicity see Supplementary Methods and **Supplementary Figure 1**. The study was approved by the London-West London and Gene Therapy Advisory Committee Research Ethics Committee (REC number 11/LO/0883).

Whole Genome Sequencing

DNA was sequenced by WGS at Edinburgh Genomics. Libraries were prepared using the Illumina TruSeq PCR-free protocol and sequenced on the Illumina HiSeq X platform. The average yield per sample was 136 Gb, with mean coverage of 36x (range 33.9-38.3).

Gene Panel, data analysis and variant classification

A virtual gene panel of 146 genes was created, which included the known 'green' 34 PCD genes plus 107 suspected ciliopathy genes that may have respiratory features based on the PCD PanelApp panel (v1.14) with five additional genes identified in the literature (*CFAP300*, *DNAH6*, *DNAJB13*, *STK36* and *TTC25)* (23–25). Analysis of variants was carried out as described in Supplementary Methods. Additional analyses were carried out across the *FOXJ1* locus for Case 3, in whom no diagnostic variants were identified using the virtual gene panel. A genome-wide expanded analysis identified a *de novo* missense mutation p.P259L (chr9:g.137242994:C>T (hg38)) in the gene *TUBB4B* only in the patient, and not present in either parent (26).

Modelling of whole exome sequencing data

A whole exome sequencing (WES)-like subset of the WGS data was obtained by extracting only the reads mapping to the regions in the TWIST Exome Capture Kit (using samtools v1.6) from the BAM file for each sample. WES CNV calling was performed by ExomeDepth (v1.1.15) as detailed in Supplementary Methods.

Homology modelling of DNAH11 and location of missense variants

A homology model of the C-terminal region of the DNAH11 motor domain (residues 3348-4504) was built using PHYRE2 (27,28). The effects of mutations in the C-terminal domain (CTD) (residues 4124-4504) on protein stability were modelled with FoldX (29). Sequences of human dynein genes were aligned with MUSCLE (30) and the sequence alignment was visualised with MView (31). Further details are given in the Supplementary Methods.

Results

Eight PCD patients were recruited to the study as six trios, one proband-mother duo and one singleton (**Table 1**). All eight cases had a confirmed clinical diagnosis of PCD, following a nasal brush biopsy. In all cases, nasal brushings were undertaken for clinical suspicion of PCD. Presentations and phenotype severity were varied and are summarised in **Table 1**.

Using a virtual PCD panel approach to initially screen the WGS, we confirmed genetic diagnosis for seven of the eight cases (**Table 2**). Three cases had biallelic pathogenic and/or likely pathogenic variants in the outer dynein arm heavy chain subunit gene *DNAH5*, which has previously been shown to contribute to the largest proportion of PCD cases among populations of European descent (1). Case 1 had a hemizygous nonsense variant (c.5281C>T, p.(Arg1761Ter)), inherited from the mother and previously reported in PCD cases (32,33), and an overlapping 13kb deletion on the other allele, inherited from the father (**Supplementary Figure 3**). Case 2 had two nonsense variants in *DNAH5*, c.3949C>T, p.(Gln1317Ter), which was previously reported (34), and the c.5281C>T, p.(Arg1761Ter) variant. ddPCR was used to phase the variants in Case 2 (**Supplementary Figure 4**). Case 4 had a single base pair deletion, resulting in a frameshift and premature termination codon (PTC) (c.10815del, p.(Pro3606Hisfs)), and a single base pair duplication, resulting in direct creation of a PTC (c.13458dup, p.(Asn4487Ter)), which were shown to be in trans through trio-based phasing. Both variants have been reported previously in PCD cases (33). All three cases with biallelic variants in *DNAH5* were shown to have absence or defects of outer dynein arms on TEM (**Figure 1**), with Case 1 and Case 2 also shown to have static cilia (**Table 1, Supplementary Video 1**). This is consistent with loss of DNAH5, which has been shown to result in outer dynein arm truncation and immotile cilia (33). All three share a similar clinical phenotype,

with chronic wet cough, sinus problems and recurrent chest infections. Hearing loss was also noted in cases 2 and 4.

Two cases had likely pathogenic variants in the cytoplasmic axonemal dynein assembly factor *DNAAF4* (*DYX1C1*), necessary for assembly and/or stability of the inner and outer dynein arms (**Table 2**) (35). Case 5 had a homozygous 3.5kb deletion encompassing exon 7 of *DNAAF4*, which was also heterozygous in the mother. There was no evidence of uniparental disomy and, given Case 5 is also reported to have an affected sibling, it is likely the father also carries the 3.5kb deletion, although the father's DNA was unavailable. This variant has previously been reported in PCD (35). Ultrastructural analysis was consistent with previous reports for *DNAAF4* mutations, with reports of absent dynein arms (35) (**Figure 1**) and immotile cilia shown on HSVM. Case 6 had a maternally-inherited nonsense variant (c.856G>T, p.(Glu286Ter)) and a paternally-inherited 3.1 kb deletion encompassing the last two exons of *DNAAF4*, predicted to disrupt the C-terminal tetratricopeptide-like helical domain region (TPR) by deleting the final 71 amino acids, as well as the 3'UTR. We predict this allele to be loss-of-function as the resulting transcript is predicted to be subject to nonsense-mediated decay. Case 6 has bronchiectasis with a recurrent need for antibiotics. He has no ear or hearing problems and situs solitus.

A further two cases had pathogenic or likely pathogenic variants in the outer dynein arm heavy chain subunit *DNAH11*. Case 7 had a homozygous missense variant in *DNAH11* (c.13373 C>T, p.(Pro4458Leu)) previously reported in PCD cases (36,37), at a low frequency in gnomAD and *in silico* predictions support pathogenicity (**Table 2**). There was sufficient evidence to classify the variant as likely pathogenic using ACMG guidelines (38,39). Case 8 had a heterozygous 2 bp deletion, resulting in a PTC (c.10221_10222del, p.(Cys3409Trpfs)), which was classified as pathogenic (**Table 2**). The second variant was a heterozygous missense (c.13288G>C, p.(Gly4430Arg)). This variant is at a low frequency in gnomAD and *in silico* predictions support pathogenicity, but it has not previously been reported in PCD cases. As the variant is in trans with the c.10221_10222del variant, shown by phasing with parental samples, there is sufficient evidence to classify the variant as likely pathogenic (**Table 2**). Both Case 7 and Case 8 were shown to have a normal cilia ultrastructure with a dysmotile phenotype (**Figure 1, Supplementary Video 2**), consistent with loss of DNAH11, which

localises to the proximal portion of respiratory cilia (40) . There is however phenotypic variability between Case 7 (situs inversus with dextrocardia and recurrent chest infections) and Case 8 (chronic wet cough and recurrent ear infections, with situs solitus).

The two missense variants in *DNAH11* both reside at the 3' C-terminal domain (CTD) of the axonemal dynein heavy chain. While there is no structure available for DNAH11, there are homologous structures of other dynein proteins. To investigate the effects of the *DNAH11* missense mutations on protein structure, we built a homology model of the DNAH11 motor domain. Both mutations occur within the CTD, on the outer side of the dynein motor dimer (**Figure 2A**) and are intriguingly very close to each other in three-dimensional space (**Figure 2B**), separated by only 4.2 Å.

Molecular modelling of the missense mutations using the program FoldX (29) predicts that the Gly4430Arg should be extremely disruptive to protein structure, with a ΔΔG of 10.9 kcal/mol. We also modelled all other CTD missense variants presented in the gnomAD v2.1 database (41) . Remarkably, out of 269 variants (**Supplementary Table 1**), Gly4430Arg has the highest ΔΔG and therefore is predicted to be the most damaging. Moreover, this position is fully conserved across all human dyneins (**Figure 2C**). This strongly suggests that Gly4430Arg is pathogenic due to its disruptive effects on protein structure.

In contrast to Gly4430Arg, Pro4458Leu is predicted to be relatively mild at a structural level, with a ΔΔG of 1.1 kcal/mol, making it only the 65th most damaging out of 269 variants (**Supplementary Table 1**). However, this residue is highly conserved, existing as a proline across all human dyneins except DNAH1, where it is an alanine. Thus, while Pro4458 is unlikely to cause a severe destabilisation of protein structure, its remarkable proximity to Gly4430 combined with a moderate structural perturbation and high conservation are supportive of pathogenicity.

To assess whether the three deletion variants identified in Cases 1, 5 and 6 would have been identified using WES, as opposed to WGS, the WGS data was subsetted to create a WES-like dataset for each sample (**Supplementary Figure 2**). CNV calling using ExomeDepth was able to identify the *DNAH5* variant c.5272-955_6197del p.(?) in both the proband and father for

Family 1. However, the deletion, which spans five exons and has a breakpoint within exon 37, was called as covering only 4 of the 5 affected exons in the proband. The single exon deletion of *DNAAF4* c.784-1037_894-2012del p.(?) was identified for the proband in Family 5, where it is homozygous, but not in the heterozygous mother. The *DNAAF4* deletion c.1048-149_*1048del p.(?) identified in Family 6 was not detected in the proband or father, despite both carrying the variant.

The phenotype of Case 3 is particularly relevant as ultrastructural analysis revealed ciliary agenesis (**Supplementary Video 3**), sometimes referred to as reduced generation of multiple motile cilia (RGMC), a specific subtype of PCD. In addition, Case 3 has shunted hydrocephalus, having undergone initial ventriculo-peritoneal (VP) shunt insertion neonatally, and a subsequent VP shunt revision as an adolescent. To date, few genes have been implicated in this RGMC phenotype by recessive inheritance, *CCNO* (cyclin O) and *MCIDAS* (multicilin) (42,43) but no pathogenic or potentially pathogenic variants were identified in either gene. Heterozygous de novo mutations in the master motile ciliogenesis transcriptional regulator *FOXJ1* were identified as the first autosomal dominant cause of a distinct PCD-like condition, associated with chronic respiratory disease, laterality defects and hydrocephalus (18). Similar cellular defects are observed with reduced apical docking of centrioles and fewer cilia, but a focused analysis revealed no identifiable pathogenic mutations in the *FOXJ1* locus. An expanded, whole genome analysis for SNP and indel candidates produced a very limited list of variants (**Supplementary Table 2**). This included a de novo missense mutation p.P259L (chr9:g.137242994:C>T (hg38)) in the gene *TUBB4B* only in the patient, and not present in either parent or found on gnomAD 4.0 or other publicly available databases. Whilst interpretation of pathogenicity from a single patient is limiting, as part of a large international collaboration, we were able to identify a further eleven patients with PCD carrying *TUBB4B* variants identified by next-generation sequencing (NGS) (26). This included five patients with PCD-only carrying the identical p.P259L (chr9:g.137242994:C>T) variant, one carrying a different missense p.P259S (chr9:g.137242993:C>T (hg38)) variant and one patient carried an in-frame ten amino acid duplication p.F242_R251dup (chr9: g.137242941_137242970dup (hg38)) (26). Moreover, we also identified a recurrent de novo *TUBB4B* variant four patients with a p.P358S (chr9:g.137243290:C>T (hg38)) variant, who presented with features of both PCD with Leber congenital amaurosis and sensorineural hearing loss. In-depth functional

277 analyses in cell and animal models confirmed pathogenicity and distinct dominant-negative
278 mechanisms of action to cause different disease presentations (26). In summary, our WGS
279 strategy was very effective at identifying a novel PCD disease gene with the first report of
280 dominant negative disease mechanisms.
281
282
283

284 **Discussion**

285 In this study, WGS of affected probands and family members led to a genetic diagnosis in all
286 eight PCD patients, representing 11 different mutations in 3 autosomal recessive PCD genes
287 and 1 de novo mutation in a novel autosomal dominant PCD candidate *TUBB4B*. In Case 3, a
288 patient with features of reduced generation of motile cilia and hydrocephalus, no pathogenic
289 variants would been identified in a panel-based approach, even with targeted sequencing of
290 potential candidates for known RGMC loci. The diagnostic rate of 100% in our small study is
291 higher than previous reports, in which 60-70% of cases received a genetic diagnosis based on
292 the known PCD gene panels (1,5,12,13,20), 75% on extended NGS panels (44) and 68-94% by
293 WES (45–48).
294
295 Next-generation sequencing (NGS) technologies continue to revolutionise rare genetics
296 research and clinical diagnostics, where the advantages of WES versus WGS are often fiercely
297 debated. Cheaper in terms of costs of sequencing, analysis and data storage, WES is generally
298 preferred as a front-line diagnostic tool. WES, however, has several issues in terms of
299 evenness of genome coverage and sequence bias, particular for copy number variations
300 (CNV). In comparison, several studies have found more accurate variant calls as well as even
301 and unbiased coverage of coding regions are generated by WGS (49,50). Our analysis was
302 focused on known and candidate PCD genes, screening simultaneously for single nucleotide
303 variants (SNVs), small indels and more complex SVs. In three cases, pathogenic deletions were
304 identified, ranging in size from 3kb to 13kb. Modelling of our WGS data to represent WES-like
305 data suggested that WES data would only have identified the deletion in one of these three
306 cases, in which the deletion was homozygous. Since our WES-like model has more uniform
307 coverage than true WES data, our WES-like model could be considered more reliable for CNV

calling than real-life WES sequence. However, we acknowledge that our model is based on 36X WGS data, whereas WES would be nearer 100X in practice but coverage does vary across commercial platforms. The ability to detect SVs in PCD is of importance and consistent with recent diagnostic guidelines for PCD, which highlighted that causal SVs and intronic mutations can be missed due to the large number and size of PCD genes (8,21). Further, WES would not have provided the precise breakpoint information that allowed confirmation of the three deletion variants by Sanger sequencing.

Where WGS was clearly advantageous was PCD disease gene discovery in patients without biallelic variants in known genes or in the one recent example of autosomal dominant inheritance (*FOXJ1*) or in the few cases of X-linked recessive inheritance (*RPGR, PIH1D3, OFD1) (12).* In Case 3, WES would have likely identified the de novo SNV in *TUBB4B,* as WES panels identified SNVs in the other 11 *TUBB4B* patients (26). Here, the clear advantage of WGS was to rule out potential non-coding alterations in the known RGMC PCD genes of known inheritance such as *FOXJ1* modes as to quickly prioritise novel candidates in our first proband. As such, a clear benefit of WGS, similar to WES, is that its findings are future-proof; the data generated can be re-screened for PCD-causing variants identified subsequent to genetic testing, if initial testing is inconclusive. However, only WGS will allow future analysis of non-coding genome for variations in regulatory elements such as of transcription factor binding sites that may underlie a subset of unsolved PCD cases.

Whatever the modality, increased genetic testing for PCD is critical. Currently genetic testing for PCD sits as an additional step to confirm diagnosis by both European and American guidelines (8,21,22). As shown here, and elsewhere, genetic testing can accurately diagnose PCD where standard clinical procedures are unavailable, impractical or inconclusive (45–47,51) Importantly, a delayed PCD diagnosis is associated with worse prognosis (52,53). A clinical diagnosis of PCD requires specialised, invasive testing, and frequently necessitates travel over considerable distances to specialist centres. These tests are unsuitable for critically ill neonates and for those unable to travel or unwilling to undergo invasive testing. Genetic testing for PCD may therefore have utility in the neonatal period or in infancy that may have been preferred to the diagnostic odyssey of the patients in this study, where the age of diagnosis was between 5 and 12 years. Suggested clinical criteria for genetic testing in PCD

could include term babies who become unwell at >12 hours of age with respiratory disease, lobar collapse, situs inversus and/or an unexpectedly high or prolonged oxygen requirement (54,55). Increasing availability, reducing costs and recognised clinical utility of NGS platforms such as WGS in the neonatal and paediatric intensive care unit setting would be consistent with such indications (50,51,54,56–59). Moreover, either WES or WGS would help rule out other confounding clinical presentations such as primary immune deficiency disorders (PIDs), with overlapping symptoms of frequent, often severe, airway infections as well as recurrent otitis media, and sinusitis (60–62). Both platforms are advantageous as they allow analysis of all potential causative genes, known and novel, thus faster to keep up with recently reported genes that may not yet be included on PCD-gene panels. Indeed such technologies may prove to be more cost- and time-effective as providing a molecular diagnosis test for PCD than conventional clinical gene panels (45,46,48,63). A prospective study would appear to be warranted, to define the optimum criteria for genetic testing and diagnostic yield in neonates and older infants and children with respiratory symptomatology to help expedite PCD diagnosis.

While earlier genetic testing for PCD is clearly a priority, PCD also remains underdiagnosed. PCD has an estimated incidence of 1 in 7,500 births, rising to 1 in 2,300 in endogamous populations (64,65). In North America, it is estimated that only 1,000 patients have a confirmed diagnosis of PCD as opposed to the predicted ~25,000- 50,000 these rates would suggest to be affected with PCD (66). Similar underdiagnosis is reported in the UK, where these prevalence estimates suggest there should be at least 8,900 people with PCD in the UK; less than a quarter of these are known to the NHS highly specialised PCD service (67). A large portion of these missing patients are likely adult patients within primary care or bronchiectasis clinics, as suggested by one study which found 12% of bronchiectasis patients had pathogenic variants in known motile ciliopathy genes (68). Increased genetic testing as a first-pass diagnostic test in these suspected cases of PCD could be a way of controlling access to more labour-intensive clinical and pathological diagnostic work-up in limited specialist centres.

PCD needs to enter the precision medicine era. A genetic diagnosis is key to improved patient prognosis as we better understand genotype-phenotype relations (69) and critically to being

trial-ready as much-needed genetic therapies come online (70,71). Increased genetic testing is also being combined with a curated worldwide database, similar to that of CFTR2 for Cystic Fibrosis (72), to enable a better understanding of these genotype-phenotype relations in PCD called CiliaVar (73). Such a database will significantly facilitate interpretation of variants, of which 21% has been suggested to be variants of unknown significance (VUS) (73). For example, in Cases 7 and 8 in our study, structural modelling of the CTD of *DNAH11* aided the assignment of pathogenicity of two missense variants, with one predicted to be detrimental to protein stability and the second, in close proximity to the first, shown to be highly conserved.

In conclusion, this study demonstrates the benefits of using WGS to obtain a genetic diagnosis for PCD, through its ability to detect SNVs and SVs simultaneously as well as detecting variants in genes outwith current gene panels. The detection of multi-kilobase deletions in three of the seven diagnosed cases highlights the need to detect SVs as part of the genetic testing for PCD. It also allowed rapid prioritisation of SNVs in novel candidate genes with dominant modes of inheritance. Practically and financially, WGS would likely sit behind current standard of care panel-based or WES diagnostic platforms. Given the high genetic diagnostic rate observed here and elsewhere (44,45,47,48), we suggest that genetic testing should be an early step in the current diagnostic pathway for PCD, particularly in cases where nasal brush biopsy is unavailable. By moving clinical and genetic diagnostic pathways in PCD earlier, ideally to early life, we could have a transformative long-term reduction in morbidity with access to specialist care and disease-modifying therapies commenced before permanent lung damage.

Acknowledgements

We thank the PCD families who participated in this study and the UK PCD Family Support Group for support; Dr Lee Murphy and colleagues at the Wellcome Trust Clinical Research Facility; Edinburgh Genomics for sequencing and analysis; and the Brompton Hospital PCD Diagnostic Service and South East Scotland Genetics Service for clinical support.

Support statement: The Scottish Genomes Partnership is funded by the Chief Scientist Office of the Scottish Government Health Directorates [SGP_1] and the MRC Whole Genome Sequencing for Health and Wealth Initiative (MC_PC_15080). We acknowledge support from the MRC (PM: MC_UU_00007_14, MR_Y015002_1); an MRC Career Development Award (MR_M02122X_1) and Lister Prize Fellowship to JAM; an NHS Research Scotland fellowship to SU; and an NRS/R+D fellowship from the NHS Lothian R&D office to DU.

Competing interests: We confirm that no competing interests.

Figures:

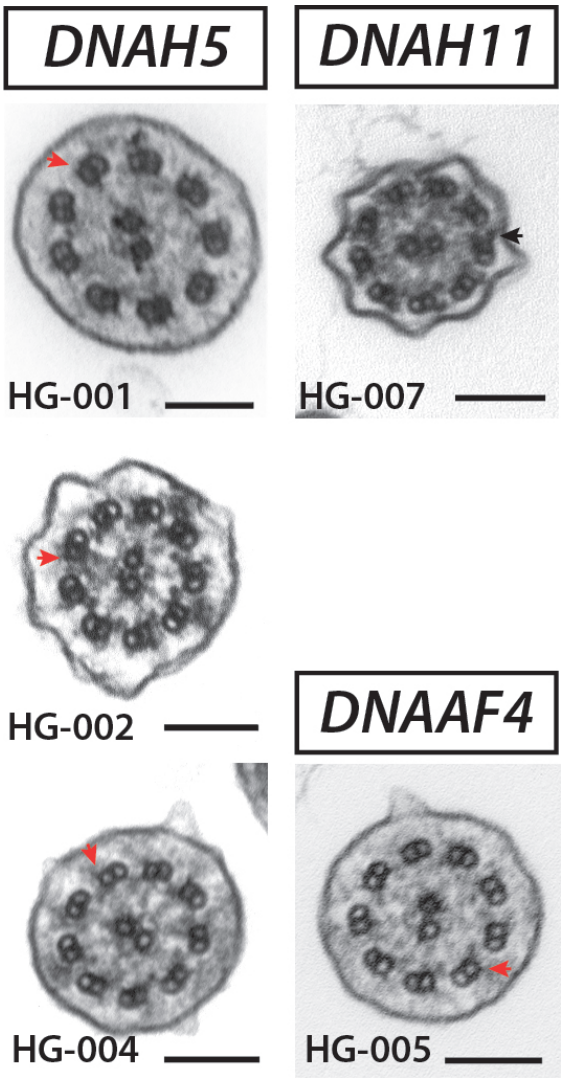


Figure 1: Ultrastructure analysis supports genetic diagnosis for outer arm dynein variants in PCD. (left) *DNAH5* variants disrupt outer dynein arms (red arrowhead = disrupted, black arrowheads normal) across axonemes of nasal brush samples from cases HG-001, HG-002 and HG-004. (right) In contrast, *DNAH11* variants do not uniformly disrupt outer dynein arms (black arrowheads) in cilia from nasal brush of case HG-007. Disruption of both inner and outer dynein arms (red arrowheads) is observed in *DNAAF4* variants, as shown for HG-005. Scale bars = 100 nm.

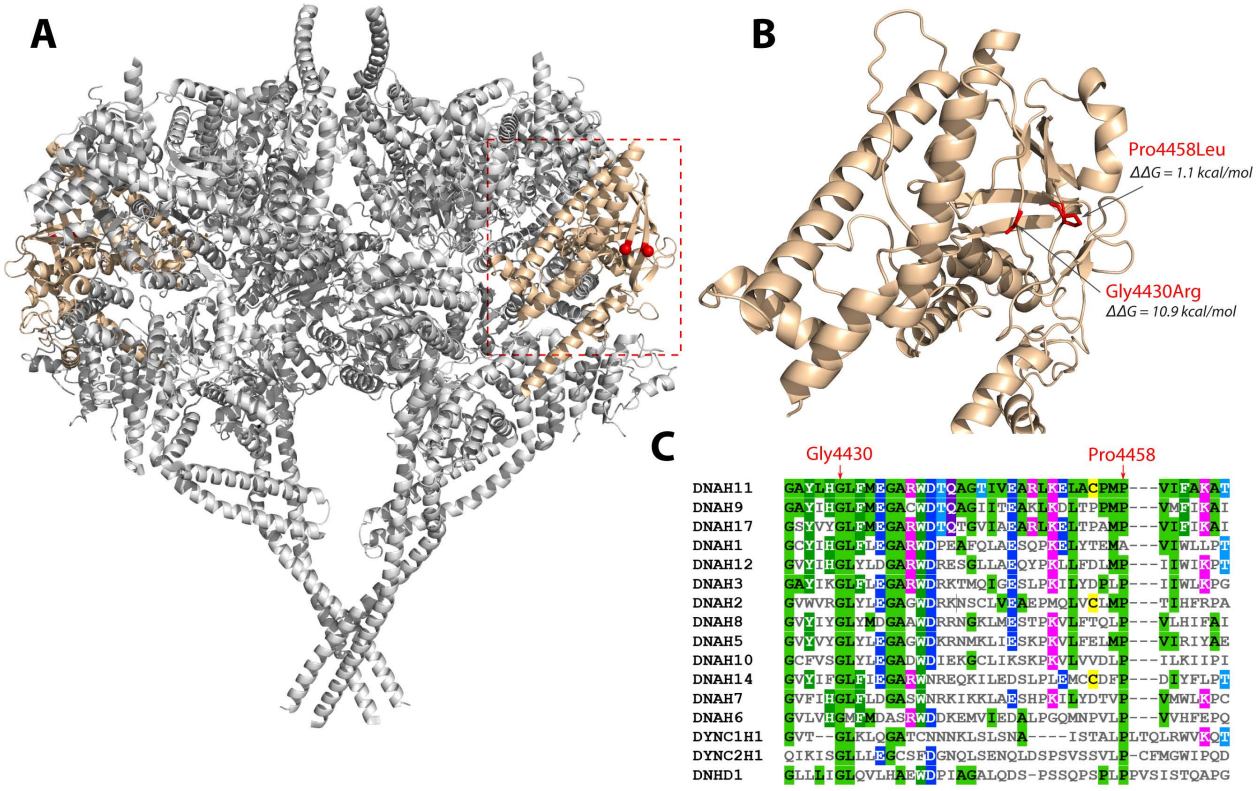


Figure 2: Structural and evolutionary analysis of DNAH11 missense mutations. (A) Structure of the human cytoplasmic dynein-1 dimer (PDB ID: 5NUG), with the location of the C-terminal domain coloured beige, and the equivalent sites of the DNAH11 mutations highlighted in red. (B) Homology model of the DNAH11 C-terminal domain with the sites of the missense mutations shown in red, along with the ΔΔG values calculated with FoldX. (C) Multiple sequence alignment of human dynein proteins around the region where the missense mutations occur.

480 **Table 1: Clinical details of PCD cases.**

Case number	Samples recruited	Sex	Ethnicity	Age at diagnosis	Age at recruitment	Ciliary ultrastructure	Cilia motility	nNO (ppb)	Current FEV1 (% predicted)	NRD	Chronic wet cough	Regular IV antibiotics	Recurrent infections	Rhinosinusitis	Bronchiectasis	Recurrent ear infection	Hearing loss	Situs inversus	Dextrocardia
1	Trio	F	EUR	11-15	16-20	Absent outer dynein arms	Static cilia	<5	83	U	Y	N	Y	Y	N	Y	N	N	N
2	Proband only	M	EUR	6-10	11-15	Outer dynein arm defect	Static cilia	8.5	90	Y	Y	Y	N	Y	Y	Y	Y	N	N
3	Trio	F	EUR	11-15	11-15	Ciliary agenesis	N/A	15.5	47	Y	Y	Y	Y	Y	Y	N	Y	N	N
4	Trio	M	EUR	0-5	6-10	Outer dynein arm defect	Static cilia	75	82	Y	Y	N	Y	Y	N	N	Y	N	N
5	Proband and Mother	F	EUR	11-15	31-35	Complete absence of dynein arm	Static cilia	U	U	N	Y	U	Y	Y	Y	Y	N	N	N
6	Trio	M	EUR	U	21-25	Unknown	Unknown	8.5	62	U	Y	Y	Y	Y	Y	N	N	N	N
7	Trio	M	SAS	6-10	6-10	Normal	Dysmotile	49.5	87	N	N	N	Y	Y	U	N	N	Y	Y
8	Trio	F	EUR	6-10	11-15	Normal	Dysmotile	22	75	U	Y	N	Y	Y	N	Y	N	N	N

Y= yes; N= no; U= unknown; F= female; M= male; NRD= neonatal respiratory distress; nNO= nasal nitric oxide; ppb= parts per billion; FEV1= forced expiratory volume in one second; IV= intravenous; EUR= European; SAS= South Asian

Table 2: Genetic variants identified in PCD cases. Variants were classified according to the ACMG/ACGS guidelines (38,39); the criteria met for each variant are shown.

Case number	Gene	Transcript	dbSNP ID	Variant(s)	Zygosity	gnomAD allele frequency (v4.0.0)	In silico predictions	ACMG/ACGS Classification	ClinVar Accession Number
1	*DNAH5*	NM_001369.3	N/A	c.5272-955_6197del p.(?)	Compound heterozygous	Absent[a]	N/A	P (PVS1, PM3, PP4)	SCV001334255
			rs148891849	c.5281C>T, p.(Arg1761Ter) (32–34)		$6x10^{-5}$	N/A	P (PVS1, PM2, PM3_str, PP4)	SCV001334256
2	*DNAH5*	NM_001369.3	rs1769508571	c.3949C>T, p.(Gln1317Ter)(74)	Compound heterozygous	$1.6x10^{-6}$	N/A	P (PVS1, PM2, PM3_str, PP4)	SCV001334257
			rs148891849	c.5281C>T, p.(Arg1761Ter)(32–34)		$6x10^{-5}$	N/A	P (PVS1, PM2, PM3_str, PP4)	SCV001334256
3	*TUBB4B*	NM_006088.6	N/A	c.776C>T, p.(Pro259Leu) (26)	Heterozygous	Absent	Grantham Score 98 AlphaMissense 0.998 (LP)	P (PS2, PM2, PP2, PP3)	SCV002770069.1
4	*DNAH5*	NM_001369.3	rs397515540	c.10815del, p.(Pro3606Hisfs) (33)	Compound heterozygous	$4.1x10^{-4}$	N/A	P (PVS1, PM2, PM3_str, PP4)	SCV001334258
			rs775696136	c.13458dup, p.(Asn4487Ter) (33)		$1.6x10^{-4}$	N/A	P (PVS1, PM2, PM3_str, PP4)	SCV001334259

5	*DNAAF4*	NM_130810.4	N/A	c.784-1037_894-2012del p.(?) (35)	Homozygous	$2.5x10^{-4}$ [b]	N/A	P (PVS1, PM3_str, PP4)	SCV001334260
6	*DNAAF4*	NM_130810.4	rs770136467	c.856G>T, p.(Glu286Ter)	Compound heterozygous	$2.6x10^{-5}$	N/A	P (PVS1, PM2, PP4)	SCV001334261
			N/A	c.1048-149_*1048del p.(?)		Absent[a]	N/A	P (PVS1, PM3, PP4)	SCV001334262
7	*DNAH11*	NM_001277115.2	rs72658835	c.13373C>T, p.(Pro4458Leu) (36,37,47,51)	Homozygous	$8.4x10^{-5}$	Grantham Score 98 REVEL 0.377 (Uncertain) AlphaMissense 0.625 (LP)	LP (PM2, PM3, PP3, PP4)	SCV001334263
8	*DNAH11*	NM_001277115.2	rs1450540788	c.10221_10222del, p.(Cys3409Trpfs)	Compound heterozygous	$1.5x10^{-5}$	N/A	P (PVS1, PM2, PP4)	SCV001334264
			N/A	c.13288G>C, p.(Gly4430Arg) (75)		$2x10^{-6}$	Grantham score 125 REVEL 0.641 (Damaging) AlphaMissense 0.813 (LP)	LP (PM1_sup, PM2, PM3_str, PP3, PP4)	SCV001334265

P= pathogenic; LP= likely pathogenic; N/A= not applicable. [a]absent from gnomAD SV and gnomAD CNV. [b]frequency from gnomAD SV.

REFERENCES

1. Lucas JS, Davis SD, Omran H, Shoemark A. Primary ciliary dyskinesia in the genomics age. Lancet Respir Med. 2020 Feb;8(2):202–16.

2. Leigh MW, Horani A, Kinghorn B, O'Connor MG, Zariwala MA, Knowles MR. Primary Ciliary Dyskinesia (PCD): A genetic disorder of motile cilia. Transl Sci Rare Dis. 2019 Jul 4;4(1–2):51–75.

3. Wallmeier J, Nielsen KG, Kuehni CE, Lucas JS, Leigh MW, Zariwala MA, et al. Motile ciliopathies. Nat Rev Dis Primers. 2020 Sep 17;6(1):77.

4. Damseh N, Quercia N, Rumman N, Dell SD, Kim RH. Primary ciliary dyskinesia: mechanisms and management. Appl Clin Genet. 2017 Sep 19;10:67–74.

5. Mirra V, Werner C, Santamaria F. Primary ciliary dyskinesia: an update on clinical aspects, genetics, diagnosis, and future treatment strategies. Front Pediatr. 2017 Jun 9;5:135.

6. Dunsky K, Menezes M, Ferkol TW. Advances in the diagnosis and treatment of primary ciliary dyskinesia: A review. JAMA Otolaryngol Head Neck Surg. 2021 Jun 17;

7. Halbeisen FS, Shoemark A, Barbato A, Boon M, Carr S, Crowley S, et al. Time trends in diagnostic testing for primary ciliary dyskinesia in Europe. Eur Respir J. 2019 Oct 24;54(4).

8. Shoemark A, Dell S, Shapiro A, Lucas JS. ERS and ATS diagnostic guidelines for primary ciliary dyskinesia: similarities and differences in approach to diagnosis. Eur Respir J. 2019 Sep 5;54(3).

9. Raidt J, Wallmeier J, Hjeij R, Onnebrink JG, Pennekamp P, Loges NT, et al. Ciliary beat pattern and frequency in genetic variants of primary ciliary dyskinesia. Eur Respir J. 2014 Dec;44(6):1579–88.

10. Jackson CL, Behan L, Collins SA, Goggin PM, Adam EC, Coles JL, et al. Accuracy of diagnostic testing in primary ciliary dyskinesia. Eur Respir J. 2016 Mar;47(3):837–48.

11. Horani A, Ferkol TW. Advances in the genetics of primary ciliary dyskinesia: clinical implications. Chest. 2018 Sep;154(3):645–52.

12. Legendre M, Zaragosi L-E, Mitchison HM. Motile cilia and airway disease. Semin Cell Dev Biol. 2021 Feb;110:19–33.

13. Knowles MR, Zariwala M, Leigh M. Primary Ciliary Dyskinesia. Clin Chest Med. 2016 Sep;37(3):449–61.

14. Goutaki M, Pedersen ESL. Phenotype-genotype associations in primary ciliary dyskinesia: where do we stand? Eur Respir J. 2021 Aug 5;58(2).

15. Shapiro AJ, Ferkol TW, Manion M, Leigh MW, Davis SD, Knowles MR. High-Speed Videomicroscopy Analysis Presents Limitations in Diagnosis of Primary Ciliary Dyskinesia. Am J Respir Crit Care Med. 2020 Jan 1;201(1):122–3.

16. Walker WT, Jackson CL, Lackie PM, Hogg C, Lucas JS. Nitric oxide in primary ciliary dyskinesia. Eur Respir J. 2012 Oct;40(4):1024–32.

17. Shapiro AJ, Josephson M, Rosenfeld M, Yilmaz O, Davis SD, Polineni D, et al. Accuracy of Nasal Nitric Oxide Measurement as a Diagnostic Test for Primary Ciliary Dyskinesia. A Systematic Review and Meta-analysis. Ann Am Thorac Soc. 2017 Jul;14(7):1184–96.

18. Wallmeier J, Frank D, Shoemark A, Nöthe-Menchen T, Cindric S, Olbrich H, et al. De Novo Mutations in FOXJ1 Result in a Motile Ciliopathy with Hydrocephalus and Randomization of Left/Right Body Asymmetry. Am J Hum Genet. 2019 Nov 7;105(5):1030–9.

19. Raidt J, Krenz H, Tebbe J, Große-Onnebrink J, Olbrich H, Loges NT, et al. Limitations of Nasal Nitric Oxide Measurement for Diagnosis of Primary Ciliary Dyskinesia with Normal Ultrastructure. Ann Am Thorac Soc. 2022 Aug;19(8):1275–84.

20. Kurkowiak M, Ziętkiewicz E, Witt M. Recent advances in primary ciliary dyskinesia genetics. J Med Genet. 2015 Jan;52(1):1–9.

21. Lucas JS, Barbato A, Collins SA, Goutaki M, Behan L, Caudri D, et al. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. Eur Respir J. 2017 Jan 4;49(1).

22. Shapiro AJ, Davis SD, Polineni D, Manion M, Rosenfeld M, Dell SD, et al. Diagnosis of primary ciliary dyskinesia. an official american thoracic society clinical practice guideline. Am J Respir Crit Care Med. 2018 Jun 15;197(12):e24–39.

23. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019 Nov;51(11):1560–5.

24. Primary ciliary disorders (Version 1.40) [Internet]. [cited 2024 Feb 17]. Available from: https://panelapp.genomicsengland.co.uk/panels/178/.

25. Home - OMIM - NCBI [Internet]. [cited 2024 Feb 17]. Available from: https://www.ncbi.nlm.nih.gov/omim

26. Mechaussier S, Dodd DO, Yeyati PL, McPhie F, Attard T, Shoemark A, et al. *TUBB4B* variants specifically impact ciliary function, causing a ciliopathic spectrum. medRxiv. 2022 Oct 21;20.

27. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845–58.

28. Zhang K, Foster HE, Rondelet A, Lacey SE, Bahi-Buisson N, Bird AW, et al. Cryo-EM Reveals How Human Cytoplasmic Dynein Is Auto-inhibited and Activated. Cell. 2017 Jun 15;169(7):1303-1314.e18.

29. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol. 2002 Jul 5;320(2):369–87.

30. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792–7.

31. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics. 1998;14(4):380–1.

32. Failly M, Bartoloni L, Letourneau A, Munoz A, Falconnet E, Rossier C, et al. Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia. J Med Genet. 2009 Apr;46(4):281–6.

33. Hornef N, Olbrich H, Horvath J, Zariwala MA, Fliegauf M, Loges NT, et al. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. Am J Respir Crit Care Med. 2006 Jul 15;174(2):120–6.

34. Hagen EM, Sicko RJ, Kay DM, Rigler SL, Dimopoulos A, Ahmad S, et al. Copy-number variant analysis of classic heterotaxy highlights the importance of body patterning pathways. Hum Genet. 2016 Dec;135(12):1355–64.

35. Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo JV, et al. DYX1C1 is required for axonemal dynein assembly and ciliary motility. Nat Genet. 2013 Sep;45(9):995–1003.

36. Knowles MR, Leigh MW, Carson JL, Davis SD, Dell SD, Ferkol TW, et al. Mutations of DNAH11 in patients with primary ciliary dyskinesia with normal ciliary ultrastructure. Thorax. 2012 May;67(5):433–41.

37. Guan W-J, Li J-C, Liu F, Zhou J, Liu Y-P, Ling C, et al. Next-generation sequencing for identifying genetic mutations in adults with bronchiectasis. J Thorac Dis. 2018 May;10(5):2618–30.

38. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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 May;17(5):405–24.

39. Ellard S. ACGS Best Practice Guidelines for Variant Classification inRare Disease 2020 [Internet]. 2020 [cited 2024 Feb 17]. Available from: https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf

40. Dougherty GW, Loges NT, Klinkenbusch JA, Olbrich H, Pennekamp P, Menchen T, et al. DNAH11 localization in the proximal region of respiratory cilia defines distinct outer dynein arm complexes. Am J Respir Cell Mol Biol. 2016 Aug;55(2):213–24.

41. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May 27;581(7809):434–43.

42. Boon M, Wallmeier J, Ma L, Loges NT, Jaspers M, Olbrich H, et al. MCIDAS mutations result in a mucociliary clearance disorder with reduced generation of multiple motile cilia. Nat Commun. 2014 Jul 22;5:4418.

43. Wallmeier J, Al-Mutairi DA, Chen C-T, Loges NT, Pennekamp P, Menchen T, et al. Mutations in CCNO result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia. Nat Genet. 2014 Jun;46(6):646–51.

44. Fassad MR, Patel MP, Shoemark A, Cullup T, Hayward J, Dixon M, et al. Clinical utility of NGS diagnosis and disease stratification in a multiethnic primary ciliary dyskinesia cohort. J Med Genet. 2020 May;57(5):322–30.

45. Gileles-Hillel A, Mor-Shaked H, Shoseyov D, Reiter J, Tsabari R, Hevroni A, et al. Whole-exome sequencing accuracy in the diagnosis of primary ciliary dyskinesia. ERJ Open Research. 2020 Oct;6(4).

46. Shamseldin HE, Al Mogarri I, Alqwaiee MM, Alharbi AS, Baqais K, AlSaadi M, et al. An exome-

first approach to aid in the diagnosis of primary ciliary dyskinesia. Hum Genet. 2020 Oct;139(10):1273–83.

47. Wheway G, Thomas NS, Carroll M, Coles J, Doherty R, Genomics England Research Consortium, et al. Whole genome sequencing in the diagnosis of primary ciliary dyskinesia. BMC Med Genomics. 2021 Sep 23;14(1):234.

48. Marshall CR, Scherer SW, Zariwala MA, Lau L, Paton TA, Stockley T, et al. Whole-Exome Sequencing and Targeted Copy Number Analysis in Primary Ciliary Dyskinesia. G3 (Bethesda). 2015 Jul 2;5(8):1775–81.

49. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci USA. 2015 Apr 28;112(17):5473–8.

50. Kingsmore SF, Cakici JA, Clark MM, Gaughran M, Feddock M, Batalov S, et al. A randomized, controlled trial of the analytic and diagnostic performance of singleton and trio, rapid genome and exome sequencing in ill infants. Am J Hum Genet. 2019 Oct 3;105(4):719–33.

51. Hao C, Guo R, Liu J, Hu X, Guo J, Yao Y, et al. Exome sequencing as the first-tier test for pediatric respiratory diseases: A single-center study. Hum Mutat. 2021 Jul;42(7):891–900.

52. Halbeisen FS, Pedersen ESL, Goutaki M, Spycher BD, Amirav I, Boon M, et al. Lung function from school age to adulthood in primary ciliary dyskinesia. Eur Respir J. 2022 Oct 20;60(4).

53. Kuehni CE, Frischer T, Strippoli MPF, Maurer E, Bush A, Nielsen KG, et al. Factors influencing age at diagnosis of primary ciliary dyskinesia in European children. Eur Respir J. 2010 Dec;36(6):1248–58.

54. Nogee LM, Ryan RM. Genetic testing for neonatal respiratory disease. Children (Basel). 2021 Mar 11;8(3).

55. Machogu E, Gaston B. Respiratory Distress in the Newborn with Primary Ciliary Dyskinesia. Children (Basel). 2021 Feb 18;8(2).

56. French CE, Delon I, Dolling H, Sanchis-Juan A, Shamardina O, Mégy K, et al. Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. Intensive Care Med. 2019 May;45(5):627–36.

57. NICUSeq Study Group, Krantz ID, Medne L, Weatherly JM, Wild KT, Biswas S, et al. Effect of Whole-Genome Sequencing on the Clinical Management of Acutely Ill Infants With Suspected Genetic Disease: A Randomized Clinical Trial. JAMA Pediatr. 2021 Dec 1;175(12):1218–26.

58. Farnaes L, Hildreth A, Sweeney NM, Clark MM, Chowdhury S, Nahas S, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. NPJ Genom Med. 2018 Apr 4;3:10.

59. Petrikin JE, Cakici JA, Clark MM, Willig LK, Sweeney NM, Farrow EG, et al. The NSIGHT1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. NPJ Genom Med. 2018 Feb 9;3:6.

60. Khalturina EO, Degtyareva ND, Bairashevskaia AV, Mulenkova AV, Degtyareva AV. Modern diagnostic capabilities of neonatal screening for primary immunodeficiencies in newborns. Clin Exp Pediatr. 2021 Oct;64(10):504–10.

61. Bousfiha A, Moundir A, Tangye SG, Picard C, Jeddane L, Al-Herz W, et al. The 2022 update of IUIS phenotypical classification for human inborn errors of immunity. J Clin Immunol. 2022 Oct 6;42(7):1508–20.

62. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun 24;184(13):3573-3587.

63. Platt CD, Zaman F, Bainter W, Stafstrom K, Almutairi A, Reigle M, et al. Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. J Allergy Clin Immunol. 2021 Feb;147(2):723–6.

64. O'Callaghan C, Chetcuti P, Moya E. High prevalence of primary ciliary dyskinesia in a British Asian population. Arch Dis Child. 2010 Jan;95(1):51–2.

65. Hannah WB, Seifert BA, Truty R, Zariwala MA, Ameel K, Zhao Y, et al. The global prevalence and ethnic heterogeneity of primary ciliary dyskinesia gene variants: a genetic database analysis. Lancet Respir Med. 2022 May;10(5):459–68.

66. O'Connor MG, Mosquera R, Metjian H, Marmor M. CHEST Reviews: Primary Ciliary Dyskinesia. CHEST Pulmonary. 2023;

67. NHS England Highly Specialised Services 2020/21 [Internet]. [cited 2024 Feb 17]. Available from: https://www.england.nhs.uk/wp-content/uploads/2023/11/PRN00328-highly-specialised-services-2020-2021.pdf

68. Shoemark A, Griffin H, Wheway G, Hogg C, Lucas JS, Genomics England Research Consortium, et al. Genome sequencing reveals underdiagnosis of primary ciliary dyskinesia in bronchiectasis. Eur Respir J. 2022 Nov 17;60(5).

69. Shoemark A, Rubbo B, Legendre M, Fassad MR, Haarman EG, Best S, et al. Topological data analysis reveals genotype-phenotype relationships in primary ciliary dyskinesia. Eur Respir J. 2021 Aug 5;58(2).

70. Raidt J, Maitre B, Pennekamp P, Altenburg J, Anagnostopoulou P, Armengot M, et al. The disease-specific clinical trial network for primary ciliary dyskinesia: PCD-CTN. ERJ Open Research. 2022 Jul;8(3).

71. ReCode doses first subject in Phase I primary ciliary dyskinesia trial [Internet]. [cited 2024 Feb 17]. Available from: https://www.clinicaltrialsarena.com/news/recode-primary-ciliary-dyskinesia-trial/

72. Welcome to CFTR2 | CFTR2 [Internet]. [cited 2024 Feb 17]. Available from: https://cftr2.org/

73. Mani R, Gomes M, González AR, Hogg C, Morris-Rosendahl D, Maitre B, et al. Development and first results of the BEAT-PCD international Primary Ciliary Dyskinesia gene variant database: CiliaVar. Paediatric respiratory epidemiology. European Respiratory Society; 2021. p. PA3458.

74. Aprea I, Nöthe-Menchen T, Dougherty GW, Raidt J, Loges NT, Kaiser T, et al. Motility of efferent duct cilia aids passage of sperm cells through the male reproductive system. Mol Hum Reprod. 2021 Feb 27;27(3).

75. Leung GKC, Mak CCY, Fung JLF, Wong WHS, Tsang MHY, Yu MHC, et al. Identifying the genetic causes for prenatally diagnosed structural congenital anomalies (SCAs) by whole-exome

670 sequencing (WES). BMC Med Genomics. 2018 Oct 25;11(1):93.

High diagnostic rate of whole genome sequencing in primary ciliary dyskinesia

Supplementary Material and Methods

Holly A Black[1,2], Sophie Marion de Proce[1*], Jose L Campos[3*], Alison Meynert[3], Mihail Halachev[3], Joseph A Marsh[3], Robert A Hirst[4], Chris O'Callaghan[4], Scottish Genomes Partnership, Javier Santoyo-Lopez[5], Jennie Murray[2,3], Kenneth Macleod[6], Don S Urquhart[6,7], Stefan Unger[6,7], Timothy J Aitman[1^], Pleasantine Mill[3^]

[1] Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[2] South East of Scotland Genetics Service, Western General Hospital, Edinburgh, UK

[3] MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[4] Centre for PCD Diagnosis and Research, Department of Respiratory Sciences, University of Leicester, UK

[5] Edinburgh Genomics, Edinburgh, UK

[6] Department of Paediatric Respiratory and Sleep Medicine, Royal Hospital for Sick Children, Edinburgh, UK

[7] Department of Child Life and Health, University of Edinburgh, Edinburgh, UK

* These authors contributed equally to the manuscript

[^] Joint senior authors

SUPPLEMENTARY MATERIALS

Supplementary Table 1: Ranked list of most disruptive reported variants (ΔΔG) in the C-terminal domain (CTD) of DNAH11 as predicted by FoldX. ΔΔG represents the change in free energy by mutation/design of proteins as predicted by the FoldX algorithm, where ΔΔG = ΔGfold(mutation) − ΔGfold(wild type).

Supplementary Table 2: Genome wide list of variants detected in patient HG-003 by Slivar.

Supplementary File 1: Gene panel used for variant filtering with G2P under a biallelic inheritance model

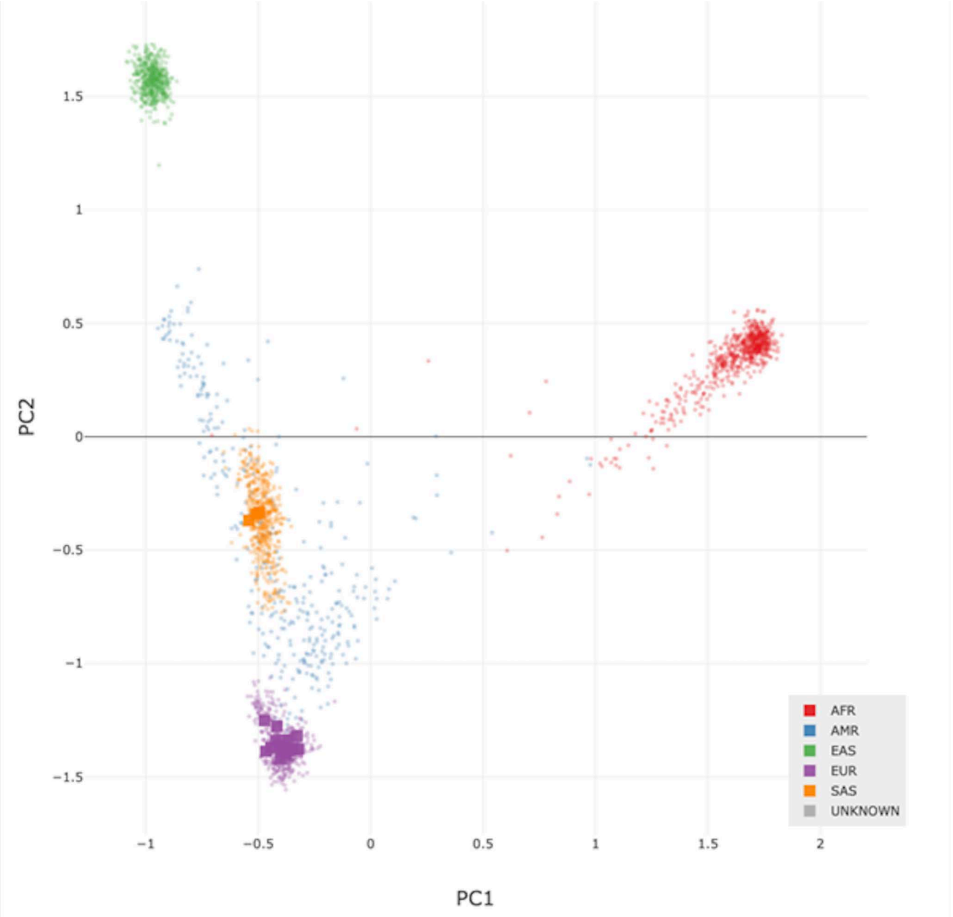
Supplementary File 2: Gene panel used for variant filtering with G2P under a monoallelic inheritance model

Supplementary File 3 Droplet digital PCR for variant phasing.

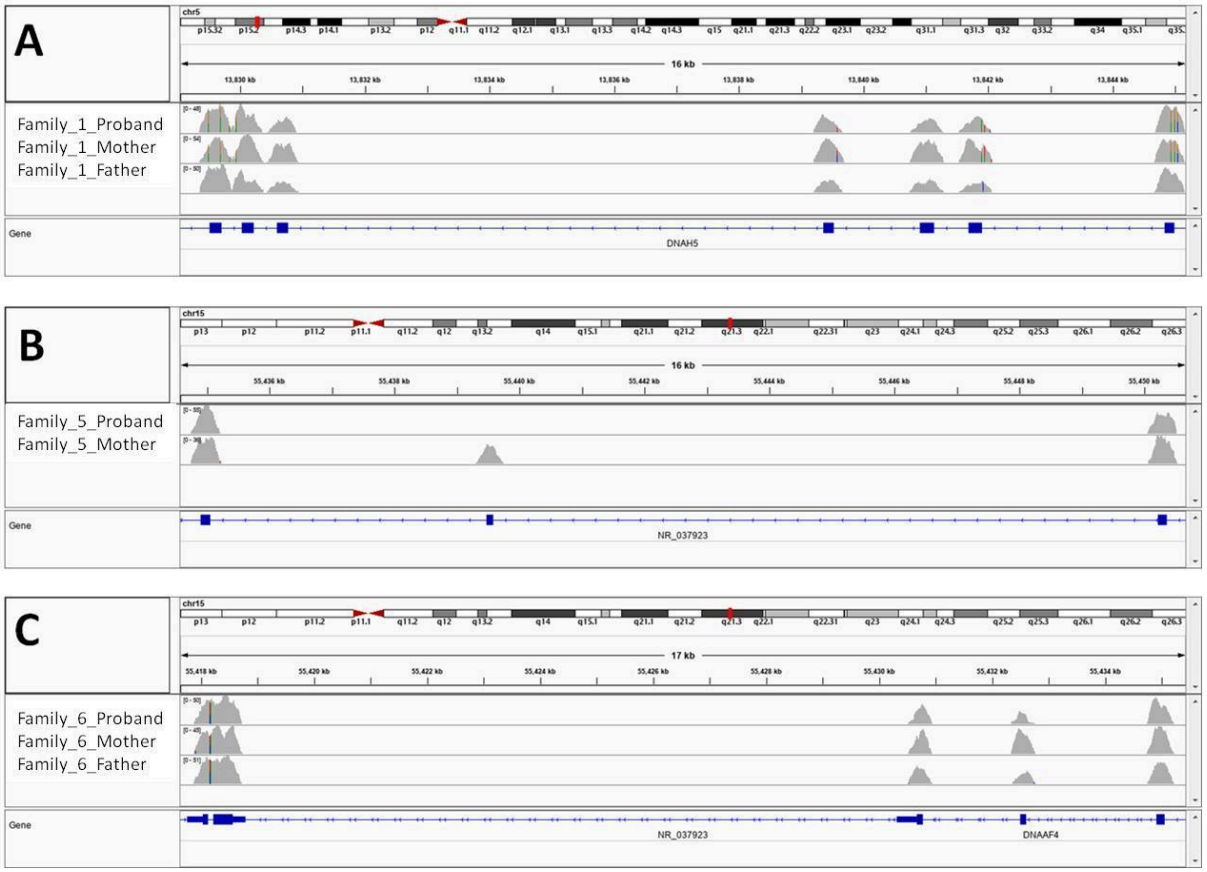
Supplementary Video 1: HSVM for cases 1, 2 and 4, which have a genetic diagnosis in *DNAH5*, showing static cilia

Supplementary Video 2: HSVM for cases 7 and 8, which have a genetic diagnosis in *DNAH11*, showing dysmotile cilia

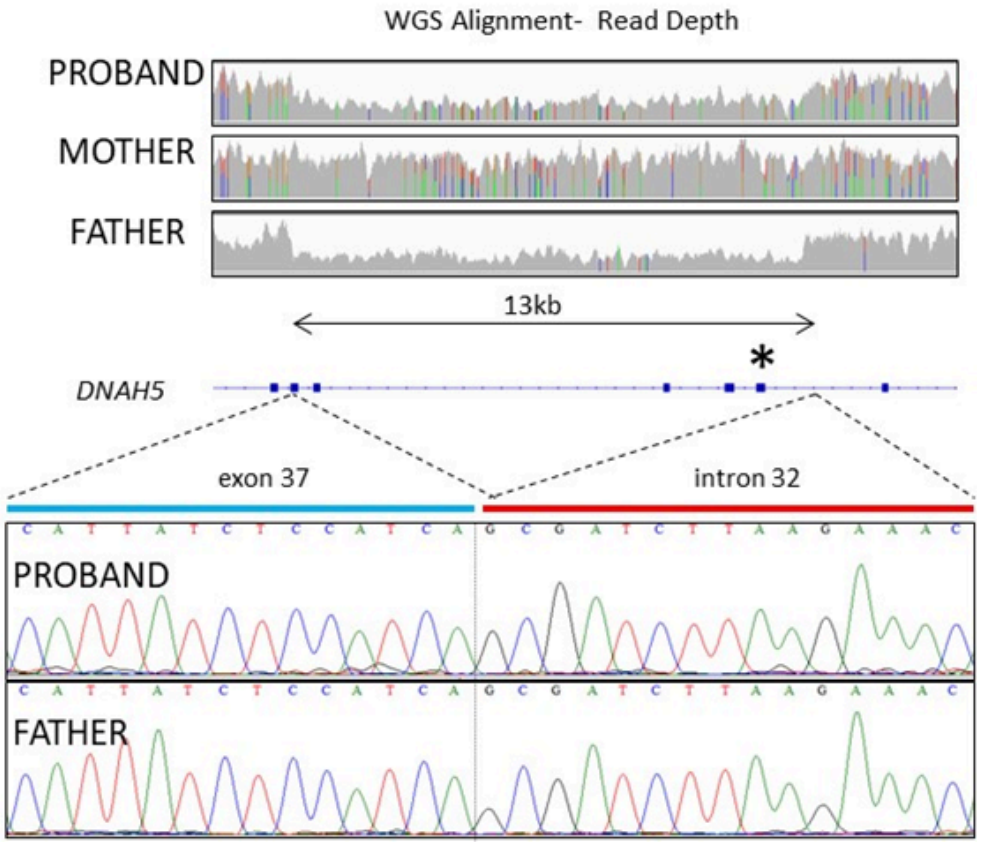
Supplementary Video 3: HSVM for case 3, showing ciliary agenesis



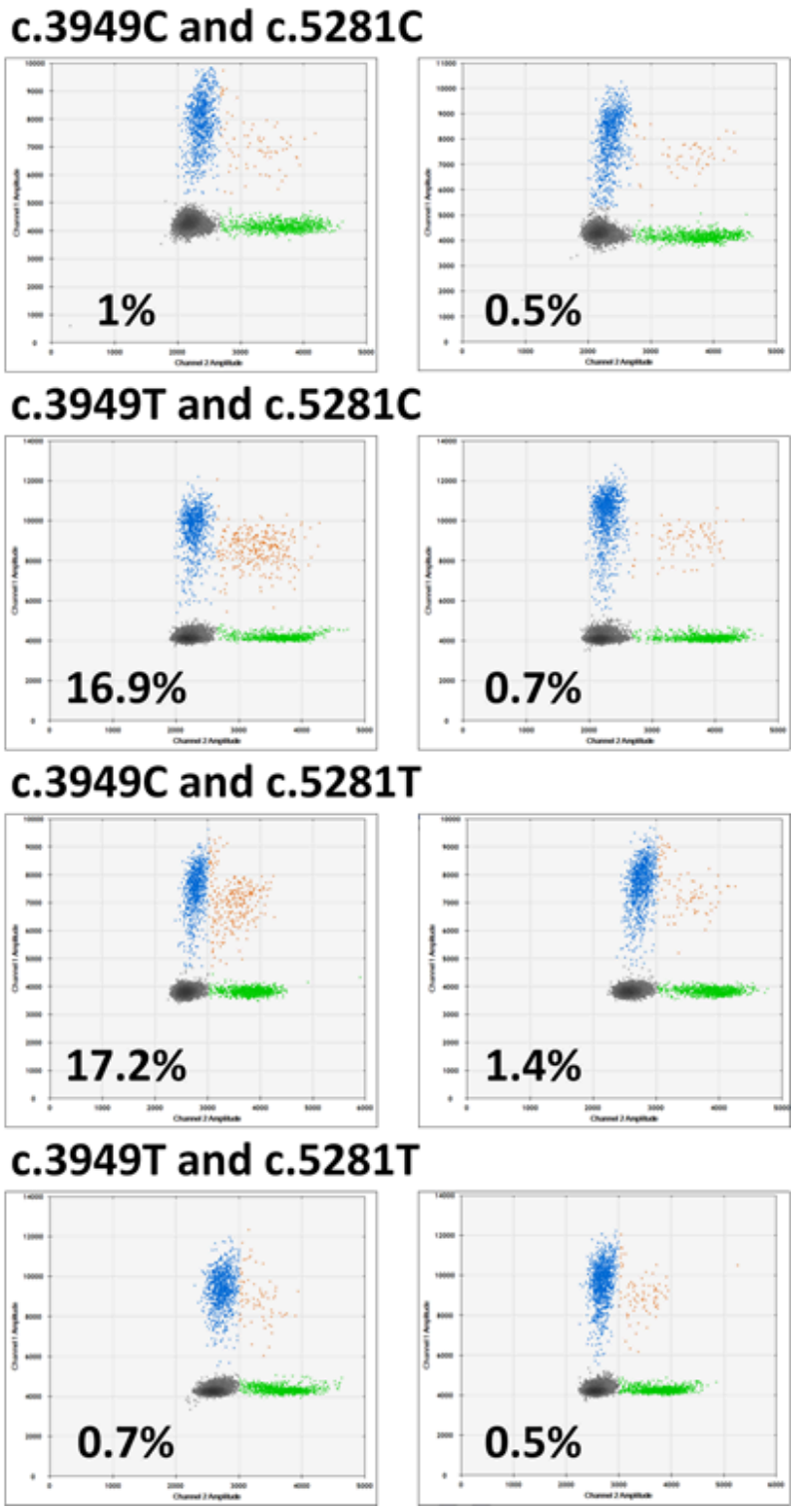
Supplementary Figure 1: Principal component analysis (PCA) of study samples compared to 1000 Genomes project samples. Peddy was used to predict ancestry of the samples used in the study by comparison with the 1000 Genomes samples. The 1000 Genomes samples (dots) are colour-coded by location. The samples in our study are represented by squares. All samples were predicted to be of European ancestry (purple squares), except three (orange squares), which were predicted to have South Asian ancestry. AFR= African; AMR= American; EAS= East Asian; EUR= European; SAS= South Asian.



Supplementary Figure 2: Alignments of the modelled WES data showing sufficient coverage to call the deletion variants identified in Cases 1, 5 and 6. A: Alignment for Family 1, showing coverage of exons 32-38 of *DNAH5.* B: Alignment for Family 5, showing coverage of exons 6 to 8 of *DNAAF4.* C: Alignment for Family 6, showing coverage exons 8-10 of *DNAAF4*.



Supplementary Figure 3: Sanger sequencing confirms 13kb deletion in *DNAH5* in Case 1. Alignments of the WGS data for Family 1 show a drop in read depth to approximately 50% of that of the surrounding regions across a 13kb region of *DNAH5* in the proband and the father. This spans from intron 32 to exon 37. PCR and Sanger sequencing across the breakpoints confirms this deletion. * indicates the position of the c.5281C>T nonsense variant, which is on the maternal haplotype and is therefore hemizygous in the proband.



Supplementary Figure 4: Drop phase results confirms the two *DNAH5* nonsense variants are on different haplotypes for Case 2. The c.3949 variant was assayed using FAM probes and the c.5281 variant was assayed using HEX probes. For each combination of alleles, a representative result from genomic DNA (left) and *Pac*I-digested DNA (right) is shown. Each figure plots the number of FAM-only positive (blue), HEX-only positive (green), FAM and HEX-positive (orange) and negative (grey) droplets. The linkage % is shown for each test. (Details see Supplementary File 3).

SUPPLEMENTARY METHODS

Patient cohort

Eight young people with PCD (50% female), aged 6 to 31 years (mean=15, SD= 7.9), were recruited to the study within the Department of Paediatric Respiratory and Sleep Medicine at the Royal Hospital for Sick Children, Edinburgh, and the South East Scotland Genetics service. All eight cases had a confirmed clinical diagnosis of PCD, following TEM and/or HSVM of a nasal brush biopsy sample. The screening and diagnostic testing was performed according to the PCD National Service protocols, with investigations including nNO, nasal brush biopsies analysed by HSVM for ciliary beat frequency and pattern and quantitative electron microscopy for ciliary ultrastructure. Clinical phenotypes are shown in **Table 1**. Blood samples were collected in EDTA tubes from the patients and parents, where available. DNA was extracted using the Chemagic DNA blood kit (Chemagen) or the Nucleon Bacc3 kit (GE Healthcare). Sample ethnicity was assessed using Peddy (v4.0.6) (1) (**Supplementary Figure 1**).

Gene Panel, sequence data analysis and variant classification

BCBio-Nextgen (0.9.7) was used for alignment and variant detection. This used bwa mem (0.7.13) to align reads to the hg38/GRCh38 reference genome (2), samblaster (0.1.22) to mark duplicate fragments (3) and GATK (3.4-0-g7e26428) for indel realignment and base recalibration (4). GATK HaplotypeCaller was used to calculate genotype likelihoods. Joint genotyping and quality control, including kinship estimates to confirm sample relatedness, were performed using in-house pipelines with GATK (4.0.2.1) following the GATK best practices. Variants were annotated using Ensembl variant effect predictor (VEP 90) (5).

A bespoke gene panel of 146 genes was created (**Supplementary Files 1 and 2**), based on the PCD PanelApp panel (v1.14) and five additional genes identified in the literature (*CFAP300*, *DNAH6*, *DNAJB13*, *STK36* and *TTC25*) (6–8). Variants (SNPs and small indels) were filtered to retain only those within genes on the gene panel and then further filtered to identify candidate variants by inheritance model (both biallelic and monoallelic), transcript consequence, and population allele frequency using the G2P plugin for VEP (9). Variants were assessed using Alamut (v2.13) (10) and classified using the ACMG variant

interpretation guidelines (11,12). Any variants classified as pathogenic or likely pathogenic were validated using Sanger sequencing and were submitted to ClinVar. Parental samples were used to determine the phase of compound heterozygous variants, except for Case 2, which used droplet digital PCR (ddPCR) for phasing (**Supplementary File 3**).

SVs were called using Manta (13) and Canvas (version 1.38) (14). The detection of SVs can be challenging and it is common practise to use complementary approaches to detect them (15). Manta detects SVs using discordant paired-end and split reads, whereas Canvas relies on changes in read coverage. We searched for any SVs present in our gene panel that were inherited from either parent. We also searched for de novo SVs for cases where both parents were available. SVs were confirmed *in silico* using SV-Plaudit (16), a tool for rapidly curating SV predictions, and/or using the Integrative Genomics Viewer tool (17). Candidate variants were confirmed in the laboratory by PCR and Sanger sequencing across the deletion breakpoints.

For Case 3, we did not find any diagnostic variants using our gene panel. As *FOXJ1* was only recently identified as a PCD gene and hence was not present on the panel, we searched for SNVs, indels and SVs in Case 3 in this gene, given the ciliary agenesis phenotype observed in this case is associated with *FOXJ1*. We also expanded our analysis for this case to a genome wide search for SNV and small indel candidates with Slivar (0.1.10), following the protocol for rare diseases (https://github.com/brentp/slivar/wiki/rare-disease Date accessed: January 2020). A small number of variants was detected (**Supplementary Table 2**) and none were identified that fitted the current modes of inheritance for PCD.

Modelling of whole exome sequencing data

A whole exome sequencing (WES)-like subset of the WGS data was obtained by extracting only the reads mapping to the regions in the TWIST Exome Capture Kit (using samtools v1.6) from the BAM file for each sample. The capture region fully covers the exons affected by copy number variants (CNVs) identified based on the WGS data (**Supplementary Figure 2**). The WES CNV calling was performed using ExomeDepth (v 1.1.15) separately on each individual from the three families in which a pathogenic CNV was identified (Families 1, 5

and 6). As controls, we used WES-like subset data from the WGS data for the other samples in this project (total of 21), from which we excluded any members of the family currently being evaluated. As the WGS data was aligned to GRCh38, we generated a custom reference dataset (exons.GRCh38) required by ExomeDepth, to replace the dataset currently distributed with the ExomeDepth package (exons.hg19); exons.GRCh38 is based on the latest CCDS release (r 22) available for the GRCh38 human genome reference (Available at: www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi?REQUEST=SHOW_STATISTICS).

Homology modelling of DNAH11 and location of missense variants

A homology model of the C-terminal region of the DNAH11 motor domain (residues 3348-4504) was built using PHYRE2 (18), based upon the cryo-electron microscopy structure of human cytoplasmic dynein-1 (PDB ID: 5NUG) (19). The effects of the mutations in the C-terminal domain (CTD) (residues 4124-4504) on protein stability were modelled with FoldX (20), using default parameters and calculated over 10 replicates. Sequences of human dynein genes were aligned with MUSCLE (21) and the sequence alignment was visualised with MView (22).

SUPPLEMENTARY REFERENCES

1. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet. 2017 Mar 2;100(3):406–13.

2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

3. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 2014 Sep 1;30(17):2503–5.

4. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

5. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016 Jun 6;17(1):122.

6. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019 Nov;51(11):1560–5.

7. Home - OMIM - NCBI [Internet]. [cited 2024 Feb 17]. Available from: https://www.ncbi.nlm.nih.gov/omim

8. Primary ciliary disorders (Version 1.40) [Internet]. [cited 2024 Feb 17]. Available from: https://panelapp.genomicsengland.co.uk/panels/178/

9. Thormann A, Halachev M, McLaren W, Moore DJ, Svinti V, Campbell A, et al. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. Nat Commun. 2019 May 30;10(1):2373.

10. Alamut™ Visual Plus - Variant Annotation and Analysis Software [Internet]. [cited 2024 Feb 18]. Available from: https://www.sophiagenetics.com/platform/alamut-visual-plus/

11. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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 May;17(5):405–24.

12. Ellard S. ACGS Best Practice Guidelines for Variant Classification inRare Disease 2020 [Internet]. 2020 [cited 2024 Feb 17]. Available from: https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf

13. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016 Apr 15;32(8):1220–2.

14. Ivakhno S, Roller E, Colombo C, Tedder P, Cox AJ. Canvas SPW: calling de novo copy number variants in pedigrees. Bioinformatics. 2018 Feb 1;34(3):516–8.

15. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun. 2019 Jul

19;10(1):3240.

16. Belyeu JR, Nicholas TJ, Pedersen BS, Sasani TA, Havrilla JM, Kravitz SN, et al. SV-plaudit: A cloud-based framework for manually curating thousands of structural variants. Gigascience. 2018 Jul 1;7(7).

17. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

18. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845–58.

19. Zhang K, Foster HE, Rondelet A, Lacey SE, Bahi-Buisson N, Bird AW, et al. Cryo-EM Reveals How Human Cytoplasmic Dynein Is Auto-inhibited and Activated. Cell. 2017 Jun 15;169(7):1303-1314.e18.

20. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol. 2002 Jul 5;320(2):369–87.

21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792–7.

22. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics. 1998;14(4):380–1.