EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

Genome sequencing reveals underdiagnosis of primary ciliary dyskinesia in bronchiectasis

Amelia Shoemark, Helen Griffin, Gabrielle Wheway, Claire Hogg, Jane S Lucas, Genomics England Research Consortium, Carme Camps, Jenny Taylor, Mary Carroll, Michael R Loebinger, James D Chalmers, Deborah Morris-Rosendahl, Hannah M Mitchison, Anthony De Soyza

Please cite this article as: Shoemark A, Griffin H, Wheway G, *et al*. Genome sequencing reveals underdiagnosis of primary ciliary dyskinesia in bronchiectasis. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.00176-2022).

Genome sequencing reveals underdiagnosis of primary ciliary dyskinesia in bronchiectasis

Amelia Shoemark[1,2]*, Helen Griffin[3]*, Gabrielle Wheway[4], Claire Hogg[2], Jane S Lucas[5,6], Genomics England Research Consortium[α], Carme Camps[7], Jenny Taylor[7], Mary Carroll[5], Michael R Loebinger[2], James D Chalmers[1], Deborah Morris-Rosendahl[8], Hannah M Mitchison[9]*, Anthony De Soyza[10]*

1. Respiratory Research Group, Molecular and cellular Medicine, University of Dundee, Dundee, DD1 9SY UK

2. Royal Brompton Hospital, London, SW3 6NP, UK and NHLI, Imperial College London

3. Primary Immunodeficiency Group, Newcastle University Translational and Clinical Research Institute, Newcastle upon Tyne, NE2 4HH, UK

4. Human Development and Health, Faculty of Medicine, University of Southampton

5. Primary Ciliary Dyskinesia Centre, University Hospital Southampton NHS Foundation Trust, Southampton SO17 1BJ, UK

6. Clinical and Experimental Sciences Academic Unit, University of Southampton Faculty of Medicine, Southampton SO17 1BJ, UK.

7. Genomics England, and William Harvey Research Institute, Queen Mary University of London, Dawson Hall, Charterhouse Square
London, EC1M 6BQ, UK

8. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

9. Clinical Genetics and Genomics, Royal Brompton Hospital, Guy's and St. Thomas' NHS Foundation Trust and NHLI, Imperial College London, UK

10. Genetics and Genomic Medicine Department, University College London, UCL Great Ormond Street Institute of Child Health, London WC1N 1EH, UK.

11. 9. Newcastle University and NIHR Biomedical Research Centre for Ageing, Freeman Hospital, Newcastle upon Tyne, NE7 7DN, UK

* Authors contributed equally to this manuscript

α The Genomics England Research Consortium Brown D, Ambrose, J. C. 1 ; Arumugam, P.1 ; Bevers, R.1 ; Bleda, M. 1 ; Boardman-Pretty, F. 1,2 ; Boustred, C. R. 1 ; Brittain, H.1 ; Caulfield, M. J.1,2 ; Chan, G. C. 1 ; Fowler, T. 1 ; Giess A. 1 ; Hamblin, A.1 ; Henderson, S.1,2; Hubbard,

T. J. P. 1 ; Jackson, R. 1 ; Jones, L. J. 1,2; Kasperaviciute, D. 1,2 ; Kayikci, M. 1 ; Kousathanas, A. 1 ; Lahnstein, L. 1 ; Leigh, S. E. A. 1 ; Leong, I. U. S. 1 ; Lopez, F. J. 1 ; Maleady-Crowe, F. 1 ; McEntagart, M.1 ; Minneci F. 1 ; Moutsianas, L. 1,2 ; Mueller, M. 1,2 ; Murugaesu, N. 1 ; Need, A. C. 1,2 ; O'Donovan P. 1 ; Odhams, C. A. 1 ; Patch, C. 1,2 ; Perez-Gil, D. 1 ; Pereira, M. B.1 ; Pullinger, J. 1 ; Rahim, T. 1 ; Rendon, A. 1 ; Rogers, T. 1 ; Savage, K. 1 ; Sawant, K. 1 ; Scott, R. H. 1 ; Siddiq, A. 1 ; Sieghart, A. 1 ; Smith, S. C. 1 ; Sosinsky, A. 1,2 ; Stuckey, A. 1 ; Tanguy M. 1 ; Taylor Tavares, A. L.1 ; Thomas, E. R. A. 1,2 ; Thompson, S. R. 1 ; Tucci, A. 1,2 ; Welland, M. J. 1 ; Williams, E. 1 ; Witkowska, K. 1,2 ; Wood, S. M. 1,2 . 1. Genomics England, London, UK 2. William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.

Take home message: Primary Ciliary Dyskinesia is underdiagnosed as a cause of idiopathic bronchiectasis. Whole genome sequencing reveals variants in motile ciliopathy genes

Abstract (236/250)

Background: Bronchiectasis can result from infectious, genetic, immunological and allergic causes. 60-80% cases are idiopathic, but a well-recognised genetic cause is the motile ciliopathy, primary ciliary dyskinesia (PCD). Diagnosis of PCD has management implications including addressing co-morbidities, implementing genetic and fertility counselling and future access to PCD-specific treatments. Diagnostic testing can be complex, however PCD genetic testing is rapidly moving from research into clinical diagnostics and would confirm the cause of bronchiectasis.

Methods: This observational study used genetic data from severe bronchiectasis patients recruited to the UK 100,000 Genomes Project and patients referred for gene panel testing within a tertiary respiratory hospital, Patients referred for genetic testing due to clinical suspicion of PCD were excluded from both analyses. Data was accessed from the British Thoracic Society audit, to investigate whether motile ciliopathies are underdiagnosed in people with bronchiectasis in the UK..

Results: Pathogenic or likely pathogenic variants were identified in motile ciliopathy genes in 17/142 (12%) individuals by whole genome sequencing. Similarly in a single centre with access to pathological diagnostic facilities, 5-10% patients received a PCD diagnosis by gene panel, often linked to normal/inconclusive nasal nitric oxide and cilia functional test results. In 4,898 audited patients with bronchiectasis, <2% were tested for PCD and <1% received genetic testing.

Conclusions: PCD is underdiagnosed as a cause of bronchiectasis. Increased uptake of genetic testing may help to identify bronchiectasis due to motile ciliopathies and ensure appropriate management.

Introduction

Bronchiectasis is both a clinical disease and a radiological appearance that has multiple causes and can be associated with a range of conditions[1].It can result from infectious, genetic, immunological or allergic causes but the majority of cases are of unknown cause and termed 'idiopathic' [2]. Guidelines recommend investigation of aetiology since it can alter management [3, 4]. A targeted approach to aetiology can reduce the number of idiopathic cases reported [5]Genetic causes include rare *CFTR* genotypes, channelopathies, immunodeficiencies, and primary ciliary dyskinesia (PCD) [2]. The estimated PCD prevalence among adults with bronchiectasis is 1-13% [6-11]. Testing for PCD is suggested in patients with supporting clinical features, including a history of neonatal distress, symptoms from childhood, recurrent otitis media, rhinosinusitis, or infertility [3]. Patients with adult-onset bronchiectasis arising from PCD are described as younger than their idiopathic bronchiectasis counterparts, having moderate impairment of lung function and higher rates of chronic infection with *Pseudomonas aeruginosa* [11-13]. Due to these risk factors and the multi-system nature of PCD, in the UK, diagnosed PCD patients have access to a specific and multidisciplinary team approach to management in specialist PCD referral centres. Recently the first randomised control trial for evidence-based medicine in PCD was completed and specific therapies which target genetic defects are in development [12, 13]. A diagnosis therefore is becoming increasingly important as it impacts on clinical care.

Mutations in at least fifty different genes cause PCD and its spectrum of associated motile ciliopathies [14]. Diagnosing PCD is complex due to the requirement for a multi-test approach requiring specialist expertise and equipment [15-17]. Therefore, the risk of late or missed diagnosis is high. Late diagnosis is associated with poorer prognosis [18]. In England, PCD

testing is available at 3 specialists centres and includes evaluation by nasal nitric oxide (NO) measurement, high-speed video microscopy, immunofluorescence, tissue culture at air liquid interface and transmission electron microscopy(TEM) [15, 19]. Genetic testing for PCD has now moved from a research-based test into clinical practice [14].

In this study we investigated the contribution of primary motile ciliopathies in bronchiectasis using three datasets: whole genome sequencing (WGS) of patients recruited to the UK 100,000 Genomes Project, clinical gene panel sequencing of patients within a large tertiary PCD and bronchiectasis centre, and PCD diagnostic data from the British Thoracic Society national audit. Specifically, the aim was to identify motile ciliopathies in patients in which diagnosis was not strongly suspected. Our analysis demonstrates that despite comprehensive national PCD testing facilities, motile ciliopathies remain underdiagnosed in people with idiopathic bronchiectasis.

Methods

The 100,000 Genomes Project

The UK 100,000 Genomes Project, overseen by Genomics England Ltd (GEL; https://www.genomicsengland.co.uk), was initiated in 2013 to sequence 100,000 genomes from NHS patients and family members in the UK affected by rare diseases or cancer [20].

142 people were recruited as 'non-CF bronchiectasis' (with 107 additional family members in a total cohort of 249 individuals). All participants provided written informed consent. Inclusion criteria were severe disease (<30 FEV_1% predicted), or individual <50yr old with involvement of >1 lobe or suspicion of an inherited cause (including ciliopathies); full recruitment inclusion/exclusion criteria are shown in the online supplement. Importantly, participants with strong clinical suspicion of PCD were recruited to the 100,000 Genomes Project as a separate group and were excluded from the present study [21].

WGS was carried out using Illumina short-read sequencing. GEL developed standardised data analysis pipelines (detailed in the supplementary methods) to filter and tier variants most likely to be clinically relevant.

We carried out expert curation of all variants tiered by GEL in non-CF bronchiectasis patients and applied additional complementary variant analysis pipelines. These interrogated an expanded panel of 91 genes associated with motile ciliopathies (**Table E3**) for exonic or splice donor/acceptor single nucleotide, short insertion/deletion, copy number (CNVs) and structural variants that were predicted to be protein altering. Recently developed variant annotation tools

(SpliceAI [22], UTRannotator [23]), were also applied to screen for additional, potentially protein-altering variants. Pathogenic and likely pathogenic variants in known disease genes for motile ciliopathies were categorised using the ACMG/AMP guidelines [24]. A summary of methodology and participant numbers can be found in **Figure E1**.

Here we report individuals with genetic variants associated with motile ciliopathies only; other potentially disease-causing variants identified, for example those in *CFTR*, will be reported elsewhere.

Royal Brompton Hospital Clinical Genetics and Genomics Laboratory bronchiectasis panel audit results

In 2017 the Royal Brompton Hospital set up clinical genetic diagnostic testing for patients with respiratory disease, including targeted analysis of genes for bronchiectasis and PCD, as part of a custom "Respigene" gene panel (Agilent Technologies, Inc.). Sequencing was performed on an Illumina NextSeq550, and reads were mapped to human genome reference (GRCh38). Variants were classified for pathogenicity according to ACMG/AMP guidelines [24]. An in-house CNV caller was and all likely pathogenic and pathogenic SNVs and CNVs were confirmed by Sanger sequencing or digital droplet PCR (ddPCR), respectively. The 52-gene bronchiectasis panel consisted of *CFTR*, *SCNN1A*, *B*, *D* and *G* genes and 47 PCD genes (**Tables E2, E3**).

Results of all clinical referrals for the bronchiectasis panel between 2017-2020 were included. Patients with a strong clinical diagnosis of PCD were referred only for testing of a PCD gene sub-panel and were excluded from this study. PCD diagnostic investigations (nasal NO, high

speed video microscopy, immunofluorescence and TEM) were performed as described previously [25-27].

British Thoracic Society audit

The 2017 British Thoracic Society (BTS) bronchiectasis audit was carried out across 105 hospitals with 4,845 records. The audit focused on diagnosis and management of adult bronchiectasis. Audits applied to patients who had a follow-up or review outpatient appointment for bronchiectasis between 1 Oct – 30 Nov 2017. Data were collected from patient notes. Participants were asked to enter all eligible cases, or where this was not possible due to large numbers, to take care to avoid bias in case selection [28]. For the purpose of this analysis, cases in which the answer was yes to two or more of the following questions were considered to have severe disease: 'advanced disease / considering transplantation', 'recurrent exacerbations (>3 per year)'; 'deteriorating bronchiectasis with declining lung function', 'pseudomonas isolated 2 or more times in the past 12 months'.

Results

WGS of patients with severe and familial bronchiectasis reveals mutations in genes associated with motile ciliopathies.

17 of 142 (12%) individuals with a clinical diagnosis of bronchiectasis screened by WGS as part of the 100,000 Genomes Project had pathogenic or likely pathogenic variants in a motile

ciliopathy gene as listed in **Tables E2 and E3**. Results for these individuals are shown in **Table 1**. The mean age of those with bronchiectasis with variants suggestive of an inherited motile ciliopathy was 45, median age 46.5, range 21-75. The male:female ratio was 6:11, in keeping with a female predominance in patients with bronchiectasis [29]. These 17 patients were recruited across seven genomic medicine centres, including the three associated with specialist PCD diagnostic centres.

Two patients from consanguineous families were found to be homozygous for the relevant pathogenic variants. All other families were not knowingly consanguineous, and all patients had compound heterozygous variants except for 1 homozygous and 2 hemizygous X-linked patients. All individuals with causal variants in motile ciliopathy genes were reported to have sinusitis and recurrent respiratory infections. Where distribution of bronchiectasis was noted (10 cases), this was generally bilateral (9/10). Where age of onset had been recorded (5 cases), this was always childhood onset (5/5). Three patients had dextrocardia, 2 had hydrocephalus, 2 had hearing impairment, 1 had bilateral otitis media.

Mutations in 13 different motile ciliopathy genes were recorded amongst the affected individuals from this cohort (**Table 1**). Genetic diagnoses included 10 cases with pathogenic or likely pathogenic variants (also 3 variants of unknown significance, VUS) identified in several known PCD genes: *CCDC39* (2 cases), *DNAI1* (2 cases), *DNAI2, DNAH5* (2 cases), *DNAH11, RSPH1* and *RSPH4A*. Two cases shared the same single de novo dominant *FOXJ1* variant initially classified as a Tier 3 VUS, until *FOXJ1* was subsequently confirmed as a novel PCD gene associated with dominant inheritance through the identification of additional patients and further experimental analysis described elsewhere [30]. Two other affected individuals carried likely

diagnostic X-linked PCD-causing variants in genes associated with additional clinical phenotypes [31, 32]: first, an *OFD1* variant c.3G>A identified as likely to affect the start codon and protein translation, however this remains a Tier 3 VUS without further experimental evidence since the parental genotypes were not available to confirm familial segregation, and furthermore since PCD-linked *OFD1* mutations tend to be located towards the 3' end of the gene [33]; secondly, an *RPGR* variant c.602A>G creating a predicted missense amino acid substitution that also remains a VUS without experimental validation or parental genotypes available. Finally, autosomal recessive variants classified as pathogenic or likely pathogenic were also identified in *CEP164*, *CFAP53* and *NEK10* in three affected individuals. All three genes have previously been connected to motile ciliopathy phenotypes, with *NEK10* and *CEP164* mutations directly linked to causing bronchiectasis in humans [34-36]. *CFAP53* mutations were previously associated with situs inversus but only mild respiratory symptoms (recurrent cough, sinusitis) [37, 38].
Additional detailed analyses identified nine more cases with variants in ciliary genes, but of less certain significance (**Table E5**).

Of note, mutations in 6 of the 13 reported genes in **Table 1** are associated with non-classical PCD clinical diagnostic findings of normal TEM and / or normal NO (*FOXJ1, NEK10, OFD1, RPGR, DNAH11* and *RSPH1)*. Nasal NO and nasal brushing data were not available for the 100,000 Genome cohort and therefore we sought to replicate the findings through audit of genetic testing in bronchiectasis patients in a tertiary respiratory hospital.

Gene panel testing of bronchiectasis patients referred to a tertiary care centre reveals mutations in genes associated with motile ciliopathies

56 patients with idiopathic bronchiectasis were referred to the Royal Brompton Hospital for diagnostic genetic testing (cases referred specifically for PCD genetic testing were excluded from this study). Four (7%) received a definite PCD genetic diagnosis, with two pathogenic or likely pathogenic variants identified in known PCD genes (*CCDC103, CCDC40, DNAH11*) (**Table 2**). There were a further three potential diagnoses, two with a likely pathogenic variant plus a second variant in the same gene classified as a VUS (*DNAH11, GAS2L2*) and one apparently homozygous for an exon duplication (*DNAL1*). This increases the total number of cases in which bi-allelic mutations were identified to 12.5%, similar to the frequency of PCD gene variants seen in the 100,000 Genomes patient cohort. In a further 4 patients, a single heterozygous pathogenic/ likely pathogenic variant was identified but no second variant, precluding definitive diagnosis.

33/56 patients had cilia function tests prior to referral for genotyping. Two of the 4 definite genetic PCD diagnoses had normal nasal NO (>77nl/min). Normal functional tests associated with *CCDC103 p.(His154Pro)* variants are in keeping with previous descriptions [39]. Another patient was homozygous for a variant in *CCDC40* (c. 940-1G>C) affecting an essential splice acceptor site. *CCDC40* causal variants normally confer microtubular disorganisation and absent inner dynein arms (absence in IF of GAS8 and DNALI1) [40, 41]. However, this complex case had unusual HSVM, TEM and IF with some features not being the classical phenotype, as there was microtubular disorganisation and absent GAS8 but with the inner arms present when tested by TEM and IF [40]. Furthermore, the individual has a brother who is a heterozygous carrier of the *CCDC40* splice variant, who does not have respiratory symptoms but has dextrocardia. These results make interpretation of the variant difficult, however as demonstrated by phenotypes of patients carrying the missense *CCDC103 H154P* variant compared to those with a loss of

function mutation in *CCDC103*, different mutations in the same gene do not always present functional cilia defects in the same way.

Functional analyses were also available on 6/7 individuals with a potential PCD genetic diagnosis. One case with a single likely pathogenic *DNAH11* variant, c.3020T>G, and no second pathogenic allele had a typical HSVM pattern for *DNAH11* defects, making the diagnosis highly likely. Another case with a single frameshift deletion of *DNAAF1* (exon 2-3) and no second allele had atypical findings on HSVM for a *DNAAF1* defect. Strikingly, this was the only individual in this cohort with low nasal NO.

BTS bronchiectasis audit data suggest access to testing in the UK may limit diagnosis

4,898 adults with bronchiectasis from 89 centres were included in the BTS audit. Only 95 were tested for PCD (1.9%). 47 had nasal NO, 45 TEM and 45 HSVM measured. 23 patients had nasal NO only, which is known to be normal in several motile ciliopathies. Evaluation by all 3 tests, as would be appropriate according to the ERS guidelines [15], was performed in 22 patients (0.4%). 597 people had a severe disease phenotype. Testing was more likely in this group and conducted in 15 people (2.5%) of which 6 (1%) received full testing. Given that the 100,000 Genome Project recruitment criteria included a category for <50 years and PCD patients tend to be younger than their idiopathic counterparts, we analysed the data according to age. 56/534 (11%) of those under 50 were referred for testing and 12/534 (2%) received full testing. These findings, taken together with the results of the 100,000 Genomes Project, suggest that there is insufficient testing for PCD in patients with bronchiectasis to identify the majority of affected patients.

Discussion

This multi-centre study is one of the first to analyse WGS in bronchiectasis. Our study highlights under-diagnosis of PCD. We identified motile ciliopathy associated genes in 12% of idiopathic bronchiectasis patients recruited for WGS due to severe, familial disease or <50years of age.

The WGS Project had the option to recruit patients under a PCD phenotype category, this infers that clinicians recruiting to the non-CF bronchiectasis category had not clinically diagnosed PCD in their patients but did have suspicion of an inherited cause. Hence this level of diagnosis in a large portion of people recruited as non-CF bronchiectasis suggests either there are barriers to accessing PCD testing and/or clinicians struggle to ascertain which cases have features suggestive of PCD. We have to question whether these truly are all cases where PCD should not have been suspected given for instance that three had dextrocardia, upper airway symptoms and bronchiectaisis.

To identify if access to PCD testing was a barrier to diagnosis, we audited data from the specialist respiratory genetics service at the Royal Brompton Hospital which runs alongside a PCD diagnostic service. This identified that 7% patients referred for genetic testing with a clinical diagnosis of idiopathic bronchiectasis had pathogenic mutations in motile ciliopathy genes. Including those with a single pathogenic mutation with an abnormal functional test and/or variant of unknown significance, this rises to 12.5%. ‘An additional 7% were found to be heterozygous for a single pathogenic variant in a known PCD gene. This could represent genuine PCD gene mutation carriers, found to be at more risk of bronchiectasis, or alternatively, these may simply be cases where a second pathogenic variant was not identified by the current analysis. Typically,

the gene panel-based genetic diagnosis of PCD is based upon a sequencing strategy that covers only the coding regions, the canonical splice sites and immediate flanking intronic sequence of the known genes, hence it is postulated that in at least some of these patients, the second pathogenic variant may be in the non-coding regions of the relevant genes. There may also be more complex structural variants missed by the current computational analyses.' Future work to achieve a higher diagnostic rate for bronchiectasis in this cohort could likely benefit from a more detailed interrogation of promotor and intronic regions of the relevant genes, that were not studied here, as well as further functional experiments to determine whether some of the variants identified in this study can provide a likely diagnosis.

Many of the patients with bronchiectasis in the Royal Brompton Hospital analysis were referred for PCD testing before genotyping, but the functional testing had given equivocal results. Normal nasal NO and normal TEM were present in most cases where tested. Normal NO and TEM have been described previously in some of the genes reported [34, 39, 42-44]. Importantly, our data suggest nasal NO testing alone is not sufficient to exclude a diagnosis of PCD in bronchiectasis [45]. We suggest that NO is not a screening test and should be used as part of a diagnostic testing algorithm alongside other testing modalities such as genetics and nasal brushing.

Amongst the WGS and gene panel genetic diagnoses, several affected individuals carried causal variants in PCD genes that confer classic cilia structure and function defects. However, a number also carried variants in PCD genes linked to less classic defects (*FOXJ1, OFD1, RPGR, RSPH1, CCDC103, GAS2L2*). In three such cases (two also reported elsewhere [46]), variants were revealed in *CEP164, CFAP53* and *NEK10,* genes currently linked to motile ciliopathy rather than clinically defined PCD. Previously, bronchiectasis in addition to syndromic features but no cilia

functional testing was reported for *CEP164* mutation patients; *NEK10* was reported to cause bronchiectasis in patients with normal nasal NO levels, normal nasal ciliary ultrastructure and negligible ciliary beating abnormalities; and *CFAP53* mutation patients also had normal NO and negligible reduction in airway ciliary beat frequency [34, 35, 37]. Patients with these milder respiratory features could therefore escape detection during standard PCD clinical evaluation.

Both the predefined cohorts (100,000 Genomes and Royal Brompton Hospital) were biased towards selection of more severe and familial disease and the prevalence in an unselected cohort may be less. A future unbiased study of all cases of bronchiectasis will define the rate of PCD in an unselected cohort. There is possible greater genetic heterogeneity than has been considered. In bronchiectasis, our cases of incomplete diagnoses imply that particular attention to sequencing of non-classical PCD genes and ciliary gene variants located outside of exonic coding regions, with potentially less clear-cut effects on cilia motility, may be warranted. Both WGS and gene panel testing were successful at identifying undiagnosed motile ciliopathies. Using a panel of known PCD genes may be a cost effective first step for referring patients with features of a motile ciliopathy, severe or familial bronchiectasis. Our data showing a significant contribution of ciliopathies within bronchiectasis cohorts supports the need for a change in policy for genetics testing in bronchiectasis, as is now reflected in the National Test Directory guidelines for the UK NHS Genomic Medicine Service which includes the clinical indication "Respiratory ciliopathies including non-CF bronchiectasis".

The BTS audit data shows only 0.4% people with bronchiectasis in the UK have guideline-recommended testing for PCD, despite the presence of a network of 3 specialist diagnostic services [47]. Better access will not resolve all the issues: the genetic cause is not identified in up

to 30% of well-defined PCD patients, therefore this may be an underestimate of the true prevalence of motile ciliopathy defects in a bronchiectasis cohort, and the true number of ciliopathy cases could account for up to 16% of the cohort [48-50].

We conclude that PCD is an underdiagnosed cause of severe adult bronchiectasis and that people with bronchiectasis who are young or have severe or familial disease should be tested for motile ciliopathies.

Acknowledgements

Study authors participate in the Bronch UK Consortium (www.bronch.ac.uk) who developed the inclusion groups within bronchiectasis for the 100,000 Genomes Project. Study authors and data contributors participate in the BEAT-PCD clinical research collaboration, supported by the European Respiratory Society. We acknowledge the BEAT-PCD Workpackage 2 team, for gene list input (WP2 Leads, Marie Legendre, Sorbonne Université, Paris, France and Suzanne Crowley, Oslo University Hospital, Norway). Study authors participate in the EMBARC clinical research collaboration supported by the European Respiratory Society. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. The British Thoracic Society Audit 2017 data was provided by the British Thoracic Society (BTS). The contribution of all centres contributing to the BTS audit is

gratefully acknowledged. This publication makes use of data provided by the British Thoracic Society Clinical Audit Programme which has no responsibility or liability for the accuracy, currency or correctness of this publication.

Author contributions:

Study conception and design: AS, HG, GW, CH, JSL, JDC, DM-R, HMM, ADeS

Data collection: AS, CH, JSL, MC, ML, DM-R, ADeS

Data analysis: AS, HG, GW, DB, CC, JDC, DM-R, HMM

Writing of the manuscript: AS, HG, GW, CH, JSL, JDC, DM-R, HMM, ADeS

Revising critically for important intellectual content and final approval: all authors

Sources of support: HMM acknowledges funding from the NIHR Biomedical Research Centre at Great Ormond Street Hospital and the Ministry of Higher Education in Egypt. The National PCD Centre in Southampton is commissioned and funded by NHS England; PCD research is supported by NIHR Southampton Biomedical Research Centre, NIHR Clinical Research Facility, National Institute for Health Research (RfPB PB-PG-1215-20014; and 200470) and The AAIR Charity (Reg. No. 1129698).

Conflict of interest statement: AS has received grants from AstraZeneca, GW is currently employed by Illumina Inc, ML has received consultancy or speaker fees from Insmed, Astra Zeneca, Parion, Grifols, Armata. JDC has received grants or contracts from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Gilead Sciences, Novartis and Insmed; and received consultancy or speaker fees from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Insmed, Janssen, Novartis and Zambon. JSL reports grants from NHS England, NIHR, AIR Charity. HM is a Trustee of Ciliopathy Alliance. ADS reports grants from AZ, GSK, Pfizer; consulting fees and lecture honoraria from Gilead, GSK, AstraZeneca, LifeArc, 30T; participation on advisory board at Bayer; receipt of equipment from GSK. All other authors have nothing to disclose.

Table 1. Genotypes of people recruited to the UK 100,000 Genomes Project as non-CF bronchiectasis carrying variants suggestive of PCD

Family code	Gene	Genomic location (GRCh38)	Mat/Pat origin	cDNA change	Protein change	Tier	ACMG/ AMP class	GEL/literature reported	dbSNP [Reference]
1	*CCDC39 (NM_181426.2)*	3:180659582T>C 3:180616872TTAC>TA	Mat Pat	c.610-2A>G c.2357_2359delinsT	Splice variant p.(Ser786fs)	Tier 1 Tier 1	5 5	GMC exit questionnaire GMC exit questionnaire	rs756235547 [51] rs587778821 [51]
2	*CCDC39 (NM_181426.2)*	3:180651496delT 3:180616604CTG>C	Unknown phase	c.1072del c.2497_2498del	p.(Thr358fs) p.(Gln833fs)	Tier 1 Tier 1	5 5	Tiered only Tiered only	rs587778822 [51] rs1007345781 [52]
3	*DNAI1 (NM_012144.4)*	9:34514436G>A 9:34514436G>A	Mat Pat	c.1612G>A c.1612G>A	p.(Ala538Thr) p.(Ala538Thr)	Tier 2 Tier 2	4 4	GMC exit questionnaire GMC exit questionnaire	rs368248592 [53] rs368248592 [53]
4	*DNAI1 (NM_012144.4)*	9:34459054G>GT 9:34513112G>A	Unknown phase	c.48+2dup c.1490G>A	Splice variant p.(Gly497Asp)	N.T. N.T.	5 3	RIPD Form RIPD Form	rs397515363 [54] rs376252276 [54]
5	*DNAI2 (NM_023036.6)*	17:74309345G>A 17:74309345G>A	Mat Pat	c.1304G>A c.1304G>A	p.(Trp435*) p.(Trp435*)	Tier 1 Tier 1	5 5	Tiered only Tiered only	rs752924362 [54] rs752924362 [54]
6	*DNAH5 (NM_001369.2)*	5:13708286A>C 5:13870868G>A	Unknown phase	c.13175T>G c.3733C>T	p.(Phe4392Cys) p.(Arg1245Cys)	Tier 2 Tier 2	3 3	Tiered only Tiered only	rs145400611 [NR] rs149609746 [NR]
7	*DNAH5 (NM_001369.2)*	5:13753290delT 5:13780960C>A	Mat Pat	c.10815del c.8821-1G>T	p.(Pro3606fs) Splice variant	Tier 1 Tier 1	5 5	Tiered only Tiered only	rs397515540 [55] rs1060501454 [NR]
8	*DNAH11 (NM_001277115.2)*	7:21620016C>T 7:21899361C>T	Unknown phase	c.4438C>T c.13075C>T	p.(Arg1480*) p.(Arg4359*)	Tier 1 Tier 1	5 5	Tiered only Tiered only	rs72657321 [44] rs774903187 [44]
9	*FOXJ1 (NM_001454.4)*	17:76137652delG	De novo, dominant	c.967del	p.(Glu323fs)	Tier 3	5	RIPD Form	rs1598372791 [30]
10	*FOXJ1 (NM_001454.4)*	17:76137652delG	De novo, dominant	c.967del	p.(Glu323fs)	Tier 3	5	GMC exit questionnaire (VUS*)	rs1598372791 [30]
11	*OFD1 (NM_003611.3)*	X:13735074G>A	Parental unknown	c.3G>A	p.(Met1?)	Tier 3	3	RIPD Form	rs778840618 [NR]
12	*RPGR (NM_000328.3)*	X:38317333T>C	Parental unknown	c.602A>G	p.(His201Arg)	Tier 2	3	Tiered only	NR
13	*RSPH1 (NM_080860.4)*	21:42486463T>G 21:42486463T>G	Mat Pat	c.275-2A>C c.275-2A>C	Splice variant Splice variant	Tier 1 Tier 1	5 5	Tiered only Tiered only	rs151107532 [56] rs151107532 [56]
14	*RSPH4A (NM_001010892.3)*	6:116630553G>GTT 6:116628100C>T	Mat Pat	c.1916+2_1916+3insTT c.1393C>T	Splice variant p.(Arg465*)	Tier 3 Tier 3	4 5	RIPD Form RIPD Form	NR rs755782051 [57]
15	*CEP164 (NM_014956.5)*	11:117387204C>T 11:117411859C>T	Mat Pat	c.1726C>T c.4228C>T	p.(Arg576*) p.(Gln1410*)	Tier 3 Tier 3	5 4	RIPD Form RIPD Form	rs145646425 [58] rs147398904 [NR]
16	*CFAP53 (NM_145020.5)*	18:50262051G>A 18:50242969C>A	Unknown phase	c.238C>T c.1144G>T	p.(Arg80*) p.(Glu382*)	Tier 3 Tier 3	5 5	RIPD Form RIPD Form	rs374898373 [NR] rs200321140 [NR]
17	*NEK10 (NM_152534.4)*	3:27352882T>C 3:27304746C>A	Unknown phase	c.1A>G c.1028+1G>T	p.(Met1?) Splice variant	Tier 3 Tier 3	4 4	RIPD Form RIPD Form	rs1363654282 [NR] rs1323610713 [NR]

*Initially classified as a VUS, done prior to the finding of additional patients and further studies describing *FOXJ1* as a new ciliopathy gene [30]. Mat, maternal; Pat, paternal; NR, not reported. RIPD (Researcher Identified Potential Diagnosis) is the notification submitted by researchers from within the GEL Research Environment for potential mutations that are not Tier 1 or Tier 2 and have not already been reported in a GMC (NHS Genomics Medical Centre) exit questionnaire. ACMG/AMP classification and GEL tiering criteria are outlined in methods section.

Table 2. Genotypes of people referred with non-CF bronchiectasis to the Royal Brompton Hospital carrying variants suggestive of PCD

Family code	Confirmed diagnosis	Gene	cDNA change	Protein change	ACMG/ AMP class	dbSNP [Reference]	Ciliary function studies			
							nNO (nl/min)	HSVM (CBF in Hz)	TEM	IF
RBH-1	Yes	*CCDC103 (NM_181426.2)*	c.461A>C c.461A>C	p.(His154Pro) p.(His154Pro)	5 5	rs145457535 [59] rs145457535 [59]	371	Slow and stiff (7.4)	Normal	DNAH5, GAS8, RSPH9 present
RBH-2	Yes	*CCDC103 (NM_181426.2)*	c.461A>C c.461A>C	p.(His154Pro) p.(His154Pro)	5 5	rs145457535 [59] rs145457535 [59]	NA	NA	NA	NA
RBH-3	Yes	*CCDC40 (NM_017950.4)*	c.940-1G>C c.940-1G>C	Splice variant Splice variant	4 4	NR NR	77	Normal in areas, reduced bending and amplitude in areas (10.6)	Microtubular disorganisation but normal IDAs	GAS8 inconclusive DNAH5, DNALI1, RSPH9 present
RBH-4	Yes	*DNAH11 (NM_001277115.2)*	c.2569C>T c.2569C>T	p.(Arg857*) p.(Arg857*)	5 5	rs72655998 [44] rs72655998 [44]	NA	NA	NA	NA
RBH-5	Potential diagnosis	*DNAH11 (NM_001277115.2)*	c.4669C>T c.8072A>G	p.(Arg1557*) p.(Gln2691Arg)	4 3	rs759040005 [NR] rs183682756 [NR]	NA	Static and stiff areas (9.3)	Normal	NA
RBH-6	Potential diagnosis	*DNAL1 (NM_031427.4)*	Exon 5 dup Exon 5 dup	Exon 5 dup Exon 5 dup	4 4	NR NR	NA	Mixed non-specific findings	Normal	DNAH5, GAS8, RSPH9 present
RBH-7	Potential diagnosis	*GAS2L2 (NM_139285.4)*	c.887_890del c.307G>A	p.(Val296Glyfs*13) p.(Ala103Thr)	4 3	rs587633197 [42] NR	77	Normal beat pattern (10.6)	Normal	NA
RBH-8	No – no 2nd pathogenic variant identified	*CCDC103 (NM_181426.2)*	c.461A>C NA	p.(His154Pro) NA	4 NA	rs145457535 [59] NA	118	Reduced beat amplitude and mucus impeded (9.7)	Normal	NA
RBH-9	No – no 2nd pathogenic variant identified	*DNAAF1 (NM_178452.6)*	Del exons 2-3 NA	p.? NA	4 NA	NR NA	47	Reduced beat amplitude (8.1)	Partial absence of outer dynein arms	DNAH5 partial absence DNALI1, RSPH9, GAS8 present
RBH-10	No – no 2nd pathogenic variant identified	*DNAAF1 (NM_178452.6)*	c.882G>A NA	p.(Trp294*) NA	4 NA	NR NA	150	Reduced beat amplitude, one twisting area (12.1)	Normal	DNAH5, GAS8, RSPH9, RSPH1, RSPH4A present
RBH-11	No – no 2nd pathogenic variant identified	*DNAH11 (NM_001277115.2)*	c.3020T>G NA	p.(Leu1007*) NA	4 NA	rs1480698078 [60] NA	86	Hyperfrequent (16.4)	Normal	NA

nNO, nasal nitric oxide level; HSVM, high speed video microscopy; CBF, ciliary beat frequency; TEM, transmission electron microscopy; IF, immunofluorescence; IDA, inner dynein arm. ACMG/AMP classification as outlined in methods section.

References

1. Aliberti S, Goeminne PC, O'Donnell AE*, et al.* Criteria and definitions for the radiological and clinical diagnosis of bronchiectasis in adults for use in clinical trials: international consensus recommendations. *The Lancet Respiratory Medicine* 2022: 10(3): 298-306.
2. Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *The Lancet* 2018: 392(10150): 880-890.
3. Polverino E, Goeminne PC, McDonnell MJ*, et al.* European Respiratory Society guidelines for the management of adult bronchiectasis. *European Respiratory Journal* 2017: 50(3).
4. Hill AT, Sullivan AL, Chalmers JD*, et al.* British Thoracic Society Guideline for bronchiectasis in adults. *Thorax* 2019: 74(Suppl 1): 1-69.
5. Araújo D, Shteinberg M, Aliberti S*, et al.* Standardised classification of the aetiology of bronchiectasis using an objective algorithm. *European Respiratory Journal* 2017: 50(6): 1701289.
6. Contarini M, Shoemark A, Rademacher J*, et al.* Why, when and how to investigate primary ciliary dyskinesia in adult patients with bronchiectasis. *Multidisciplinary Respiratory Medicine* 2018: 13.
7. Shoemark A, Ozerovitch L, Wilson R. Aetiology in adult patients with bronchiectasis. *Respir Med* 2007: 101(6): 1163-1170.
8. Lonni S, Chalmers JD, Goeminne PC*, et al.* Etiology of Non-Cystic Fibrosis Bronchiectasis in Adults and Its Correlation to Disease Severity. *Annals of the American Thoracic Society* 2015: 12(12): 1764-1770.
9. McCallum GB, Binks MJ. The Epidemiology of Chronic Suppurative Lung Disease and Bronchiectasis in Children and Adolescents. *Frontiers in pediatrics* 2017: 5: 27.
10. Chalmers JD, Chang AB, Chotirmall SH*, et al.* Bronchiectasis. *Nature reviews Disease primers* 2018: 4(1): 45.
11. Gokdemir Y, Hamzah A, Erdem E*, et al.* Quality of Life in Children with Non-Cystic-Fibrosis Bronchiectasis. *Respiration* 2014: 88(1): 46-51.
12. Kobbernagel HE, Buchvald FF, Haarman EG*, et al.* Efficacy and safety of azithromycin maintenance therapy in primary ciliary dyskinesia (BESTCILIA): a multicentre, double-blind, randomised, placebo-controlled phase 3 trial. *The Lancet Respiratory medicine* 2020: 8(5): 493-505.
13. Paff T, Omran H, Nielsen KG*, et al.* Current and Future Treatments in Primary Ciliary Dyskinesia. *Int J Mol Sci* 2021: 22(18).
14. Lucas JS, Davis SD, Omran H*, et al.* Primary ciliary dyskinesia in the genomics age. *The Lancet Respiratory Medicine* 2020: 8(2).
15. Lucas JS, Barbato A, Collins SA*, et al.* European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *Eur Respir J* 2017: 49(1).
16. Shapiro AJ, Davis SD, Polineni D*, et al.* Diagnosis of Primary Ciliary Dyskinesia. An Official American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med* 2018: 197(12): e24-e39.
17. Shoemark A, Rubbo B, Haarman E*, et al.* The controversies and difficulties of diagnosing primary ciliary dyskinesia. *American Journal of Respiratory and Critical Care Medicine* 2020: 201(1).
18. Shah A, Shoemark A, MacNeill SJ*, et al.* A longitudinal study characterising a large adult primary ciliary dyskinesia population. *Eur Respir J* 2016: 48(2): 441-450.
19. Lucas JS, Burgess A, Mitchison HM*, et al.* Diagnosis and management of primary ciliary dyskinesia. *Archives of disease in childhood* 2014: 99(9): 850-856.
20. The National Genomics Research and Healthcare Knowledgebase v5, Genomics England. doi:10.6084/m9.figshare.4530893.v5. . 2019.
21. Wheway G, Mitchison HM. Opportunities and Challenges for Molecular Understanding of Ciliopathies-The 100,000 Genomes Project. *Frontiers in genetics* 2019: 10: 127.
22. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF*, et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 2019: 176(3): 535-548 e524.
23. Zhang X, Wakeling M, Ware J*, et al.* Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics* 2021: 37(8): 1171-1173.
24. Richards S, Aziz N, Bale S*, 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. *Genetics in medicine : official journal of the American College of Medical Genetics* 2015: 17(5): 405-424.
25. Rubbo B, Shoemark A, Jackson CL*, et al.* Accuracy of High-Speed Video Analysis to Diagnose Primary Ciliary Dyskinesia. *Chest* 2019: 155(5).
26. Shoemark A, Dixon M, Corrin B*, et al.* Twenty-year review of quantitative transmission electron microscopy for the diagnosis of primary ciliary dyskinesia. *Journal of Clinical Pathology* 2012: 65(3).
27. Shoemark A, Frost E, Dixon M*, et al.* Accuracy of Immunofluorescence in the Diagnosis of Primary Ciliary Dyskinesia. *Am J Respir Crit Care Med* 2017.
28. Hill AT, Sullivan AL, Chalmers JD*, et al.* British Thoracic Society guideline for bronchiectasis in adults. *BMJ open respiratory research* 2018: 5(1): e000348.
29. Vidaillac C, Yong VFL, Jaggi TK*, et al.* Gender differences in bronchiectasis: a real issue? *Breathe (Sheff)* 2018: 14(2): 108-121.
30. Wallmeier J, Frank D, Shoemark A*, et al.* De Novo Mutations in FOXJ1 Result in a Motile Ciliopathy with Hydrocephalus and Randomization of Left/Right Body Asymmetry. *American Journal of Human Genetics* 2019: 105(5).
31. Moore A, Escudier E, Roger G*, et al.* RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet* 2006: 43(4): 326-333.
32. Budny B, Chen W, Omran H*, et al.* A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. *Human genetics* 2006: 120(2): 171-178.
33. Bukowy-Bieryllo Z, Rabiasz A, Dabrowski M*, et al.* Truncating mutations in exons 20 and 21 of OFD1 can cause primary ciliary dyskinesia without associated syndromic symptoms. *J Med Genet* 2019: 56(11): 769-777.
34. Chivukula RR, Montoro DT, Leung HM*, et al.* A human ciliopathy reveals essential functions for NEK10 in airway mucociliary clearance. *Nature Medicine* 2020: 26(2): 244-251.
35. Shamseldin HE, Al Mogarri I, Alqwaiee MM*, et al.* An exome-first approach to aid in the diagnosis of primary ciliary dyskinesia. *Hum Genet* 2020: 139(10): 1273-1283.
36. Shamseldin HE, Shaheen R, Ewida N*, et al.* The morbid genome of ciliopathies: an update. *Genetics in medicine : official journal of the American College of Medical Genetics* 2020: 22(6): 1051-1060.
37. Narasimhan V, Hjeij R, Vij S*, et al.* Mutations in CCDC11, which encodes a coiled-coil containing ciliary protein, causes situs inversus due to dysmotility of monocilia in the left-right organizer. *Hum Mutat* 2015: 36(3): 307-318.
38. Perles Z, Cinnamon Y, Ta-Shma A*, et al.* A human laterality disorder associated with recessive CCDC11 mutation. *J Med Genet* 2012: 49(6): 386-390.
39. Shoemark A, Moya E, Hirst RA*, et al.* High prevalence of CCDC103 p.His154Pro mutation causing primary ciliary dyskinesia disrupts protein oligomerisation and is associated with normal diagnostic investigations. *Thorax* 2018: 73(2).
40. Antony D, Becker-Heck A, Zariwala MA*, et al.* Mutations in CCDC39 and CCDC40 are the major cause of primary ciliary dyskinesia with axonemal disorganization and absent inner dynein arms. *Hum Mutat* 2013: 34(3): 462-472.
41. Becker-Heck A, Zohn IE, Okabe N*, et al.* The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nature genetics* 2011: 43(1): 79-84.
42. Bustamante-Marin XM, Yin WN, Sears PR*, et al.* Lack of GAS2L2 Causes PCD by Impairing Cilia Orientation and Mucociliary Clearance. *Am J Hum Genet* 2019: 104(2): 229-245.
43. Bukowy-Bieryłło Z, Ziętkiewicz E, Loges NT*, et al.* RPGR mutations might cause reduced orientation of respiratory cilia. *Pediatr Pulmonol* 2013: 48(4): 352-363.
44. Knowles MR, Leigh MW, Carson JL*, et al.* Mutations of DNAH11 in patients with primary ciliary dyskinesia with normal ciliary ultrastructure. *Thorax* 2012: 67(5): 433-441.
45. Shapiro AJ, Davis SD, Leigh MW*, et al.* Limitations of Nasal Nitric Oxide Testing in Primary Ciliary Dyskinesia. *American Journal of Respiratory and Critical Care Medicine* 2020: 202(3): 476-477.
46. Wheway G, Thomas NS, Carroll M*, et al.* Whole genome sequencing in the diagnosis of primary ciliary dyskinesia. *BMC Med Genomics* 2021: 14(1): 234.
47. Lucas JS, Chetcuti P, Copeland F*, et al.* Overcoming challenges in the management of primary ciliary dyskinesia: The UK model. *Paediatric respiratory reviews* 2014: 15(2): 142-145.

48. Paff T, Kooi IE, Moutaouakil Y*, et al.* Diagnostic yield of a targeted gene panel in primary ciliary dyskinesia patients. *Hum Mutat* 2018: 39(5): 653-665.
49. Fassad MR, Patel MP, Shoemark A*, et al.* Clinical utility of NGS diagnosis and disease stratification in a multiethnic primary ciliary dyskinesia cohort. *J Med Genet* 2020: 57(5): 322-330.
50. Gileles-Hillel A, Mor-Shaked H, Shoseyov D*, et al.* Whole-exome sequencing accuracy in the diagnosis of primary ciliary dyskinesia. *ERJ Open Res* 2020: 6(4).
51. Merveille AC, Davis EE, Becker-Heck A*, et al.* CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nature genetics* 2011: 43(1): 72-78.
52. Zariwala MA, Gee HY, Kurkowiak M*, et al.* ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. *Am J Hum Genet* 2013: 93(2): 336-345.
53. Zariwala MA, Leigh MW, Ceppa F*, et al.* Mutations of DNAI1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation. *Am J Respir Crit Care Med* 2006: 174(8): 858-866.
54. Knowles MR, Leigh MW, Ostrowski LE*, et al.* Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. *Am J Hum Genet* 2013: 92(1): 99-106.
55. Hornef N, Olbrich H, Horvath J*, et al.* DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am J Respir Crit Care Med* 2006: 174(2): 120-126.
56. Kott E, Legendre M, Copin B*, et al.* Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects. *Am J Hum Genet* 2013: 93(3): 561-570.
57. Frommer A, Hjeij R, Loges NT*, et al.* Immunofluorescence Analysis and Diagnosis of Primary Ciliary Dyskinesia with Radial Spoke Defects. *Am J Respir Cell Mol Biol* 2015: 53(4): 563-573.
58. Chaki M, Airik R, Ghosh AK*, et al.* Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. *Cell* 2012: 150(3): 533-548.
59. Panizzi JR, Becker-Heck A, Castleman VH*, et al.* CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nature genetics* 2012: 44(6): 714-719.
60. Xia H, Huang X, Deng S*, et al.* DNAH11 compound heterozygous variants cause heterotaxy and congenital heart disease. *PLoS One* 2021: 16(6): e0252786.

Genome sequencing reveals underdiagnosis of primary ciliary dyskinesia in idiopathic bronchiectasis

Online Data Supplement

Supplementary Table E1. PanelApp Non-CF bronchiectasis gene panel. Genes considered to be clinically diagnostic for non-CF bronchiectasis.

Gene	Mode of inheritance
CFTR	BIALLELIC, autosomal or pseudoautosomal
PIK3CD	MONOALLELIC, autosomal or pseudoautosomal, NOT imprinted
SCNN1A	MONOALLELIC, autosomal or pseudoautosomal, imprinted status unknown
SCNN1B	BOTH monoallelic and biallelic, autosomal or pseudoautosomal
SCNN1G	MONOALLELIC, autosomal or pseudoautosomal, NOT imprinted

Supplementary Table E2. PanelApp primary ciliary disorders gene panel. Genes considered to be clinically diagnostic for primary ciliary disorders.

Gene	Mode of inheritance	Phenotype, OMIM number
ARMC4	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 23, 615451
C21orf59	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 26, 615500
CCDC103	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 17, 614679
CCDC114	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 20, 615067
CCDC151	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 30, 616037
CCDC39	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 14, 613807
CCDC40	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 15, 613808
CCDC65	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 27, 615504
CCNO	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 29;
DNAAF1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 13, 613193
DNAAF2	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 10, 612518
DNAAF3	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 2, 606763
DNAAF4	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 25, 615482 Dyslexia, susceptibility to, 1, 127700
DNAAF5	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 18, 614874
DNAH11	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 7, with or without situs inversus, 611884
DNAH5	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 3, with or without situs inversus, 608644
DNAI1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 1, with or without situs inversus, 244400
DNAI2	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 9, with or without situs inversus, 612444
DNAL1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 16, 614017
DRC1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 21, 615294
GAS8	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 33, 616726
HYDIN	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 5, 608647
LRRC6	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 19, 614935
MCIDAS	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 42, 618695
PIH1D3	X-LINKED: hemizygous mutation in males, biallelic mutations in females	Ciliary dyskinesia, primary, 36, X-linked, 300991
RPGR	X-LINKED: hemizygous mutation in males,	Ciliopathies

	biallelic mutations in females	
RSPH1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 24, 615481
RSPH3	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 32, 616481
RSPH4A	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 11, 612649
RSPH9	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 12, 612650
SPAG1	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 28, 615505
ZMYND10	BIALLELIC, autosomal or pseudoautosomal	Ciliary dyskinesia, primary, 22, 615444

Supplementary Table E3. Additional 91 genes screened in 100,000 Genomes Project cohort during expanded GeCIP analysis.

AK7	*CFAP52*	*DNAH2*	*LRRC56*	*RPSA*
AKAP4	*CFAP57*	*DNAH5*	*MCIDAS*	*RSPH1*
ARMC2	*CFAP58*	*DNAH6*	*MED13L*	*RSPH3*
BRWD1	*CFAP65*	*DNAH8*	*MEGF8*	*RSPH4A*
CCDC103	*CFAP69*	*DNAH9*	*MMP21*	*RSPH9*
CFAP53	*CFAP70*	*DNAI1*	*MNS1*	*SPAG1*
CCDC39	*CFAP91*	*DNAI2*	*NEK10*	*SPAG17*
CCDC40	*CFC1*	*DNAJB13*	*NEK8*	*SPAG6*
CCDC65	*CRELD1*	*DNAL1*	*NME7*	*SPEF2*
CCNO	*DNAAF1*	*DRC1*	*NME8*	*STK36*
CEP135	*DNAAF11*	*DZIP1*	*NODAL*	*TTC12*
CFAP221	*DNAAF2*	*ENKUR*	*ODAD1*	*TTC21A*
CFAP251	*DNAAF3*	*FOXJ1*	*ODAD2*	*TTC29*
CFAP298	*DNAAF4*	*FSIP2*	*ODAD3*	*ZIC3*
CFAP300	*DNAAF5*	*GAS2L2*	*ODAD4*	*ZMYND10*
CFAP43	*DNAAF6*	*GAS8*	*OFD1*	
CFAP44	*DNAH1*	*GATA6*	*PKD1L1*	
CFAP45	*DNAH11*	*GDF1*	*QRICH2*	
CFAP47	*DNAH17*	*HYDIN*	*RPGR*	

Supplementary Table E4. 52 genes screened in RBH patients referred with non-CF bronchiectasis

ARMC4	*CFAP300*	*DNAI2*	*PIH1D3*	*STK36*
CCDC39	*CFTR*	*DNAJB13*	*RPGR*	*TAPT1*
CCDC40	*DNAAF1*	*DNAL1*	*RSPH1 (TSGA2)*	*TTC25*
CCDC65	*DNAAF2*	*DRC1*	*RSPH3*	*ZMYND10*
CCDC103	*DNAAF3*	*FOXJ1*	*RSPH4A (RSHL3)*	

CCDC114	*DNAAF4*	*GAS2L2*	*RSPH9*	
CCDC151	*DNAAF5*	*GAS8*	*SCNN1A*	
CCNO	*DNAH11*	*HYDIN*	*SCNN1B*	
CEP164	*DNAH5*	*LRRC56*	*SCNN1D*	
CFAP46	*DNAH6*	*LRRC6*	*SCNN1G*	
CFAP53	*DNAH9*	*MCIDAS*	*SPAG1*	
CFAP298	*DNAI1*	*NME8*	*SPEF1*	

Supplementary Table E5. Genotypes of people recruited to the UK 100,000 Genomes Project as non-CF bronchiectasis carrying additional variants which may be consistent with PCD

Family code	Gene	Genomic location (GRCh38)	Mat/Pat origin	cDNA change	SpliceAI delta score	Protein change	Tier	ACMG/ AMP class	dbSNP [Reference]
18	*DNAH5*	5:13753290delA	Pat	c.10815del	N/A	p.(Pro3606fs)	Not tiered	5	rs397515540 [1]
	(NM_001369.2)	5:13717534T>C	Mat	c.12500-14A>G	0.91 AG	Splice variant	Not tiered	3	NR
19	*DNAH8**	6:38737847A>G	Shared sibling	c.991A>G	N/A	p.(Thr331Ala)	Tier3	3	rs143492695 [NR]
	(NM_001206927.2)	6:38915197A>G	Not sibling	c.9964-4A>G	0.88 AG	Splice variant	Tier3	3	rs199513328 [NR]
20	*DNAAF5*	7:761741G>A	Parental	c.1471-12G>A	0.91 AG	Splice variant	Not tiered	3	rs370817681 [NR]
	(NM_017802.4)	No 2nd variant found	unknown	-	-	-	-	-	-
21	*DNAAF5*	7:774754C>T	Parental	c.2083-252C>T	0.90 DG	Splice variant	Not tiered	3	NR
	(NM_017802.4)	No 2nd variant found	unknown	-	-	-	-	-	-
22	*DNAH8**	6:38823689G>A	Parental	c.3847+1G>A	1.00 DL	Splice variant	Tier3	3	NR
	(NM_001206927.2)	No 2nd variant found	unknown	-	-	-	-	-	-
23	*DNAH17**	17:78566999C>T	Parental	c.1452G>A	0.82 DG	p.(Ser484Ser)	Tier3	3	rs763109316 [NR]
	(NM_173628.4)	No 2nd variant found	unknown	-	-	-	-	-	-
24	*DNAI1*	9:34459054G>GT	Shared sibling	c.48+2dup	0.92 DL	Splice variant	Tier3	5	rs397515363 [53]
	(NM_012144.4)	No 2nd variant found	-	-	-	-	-	-	-
25	*DNAI1*	9:34490011G>C	Parental	c.389-1G>C	0.68 AG	Splice variant	Not tiered	3	rs200488444 [NR]
	(NM_012144.4)	No 2nd variant found	unknown	-	-	-	-	-	-
	SPEF2	Chr5/Chr8 translocation†	Parental	translocation	N/A	p.?	Not tiered	3	NR
	(NM_024867.4)	No 2nd variant found	unknown	-	-	-	-	-	-
26	*RSPH4A*	6:116628175C>T	Pat	c.1468C>T	N/A	p.(Arg490*)	Not tiered	5	rs118204043 [2, 3]
	(NM_001010892.3)	No 2nd variant found	-	-	-	-	-	-	-

SpliceAI delta score: AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss. **DNAH8* and *DNAH17* are not known to be expressed in cilia [4]. †Details in main text.

Full materials and methods

The 100,000 Genomes Project

The UK 100,000 Genomes Project, overseen by Genomics England Ltd (GEL; https://www.genomicsengland.co.uk), a company wholly owned by the UK Department for Health, was initiated in 2013 to sequence 100,000 genomes from NHS patients (and family members) in the UK affected by a rare disease, or cancer [5]. Participants were recruited from 2014 until September 2018 across 13 Genome Medicine Centres (GMCs), with blood or saliva taken from affected individuals and, where possible, both parents. In some cases patients were recruited with other family members, in other cases as singletons. DNA was extracted and stored centrally in the UK Biocentre.

142 people were recruited as 'non-CF bronchiectasis' (with 107 additional family members in the total cohort of 249 individuals). All participants provided written informed consent. Importantly, a separate group of 274 participants also recruited to the 100,000 Genomes Project were excluded from the present study, comprising 118 affected individuals with clinical suspicion of PCD (plus 156 family members) [6].

All whole genome library preparation and sequencing was carried out at the Wellcome Sanger Institute in Cambridge using Illumina short-read sequencing. All germline genomes were sequenced PCR-free to a depth allowing 95 percent of the autosomal genome (as defined by GRCh38) to be read by at least 15 or more independent observations, each having a quality of >Q30 and mappability of >mapQ20 (germline to 30x). Illumina, Inc. oversaw all sequencing and quality control of sequencing. GEL developed standardised data analysis pipelines, with raw sequence data initially aligned to GRCh37 early in the project, and later to GRCh38, using small nucleotide variant (SNV) calling and filtering pipelines to remove common SNVs (>0.001 in gnomAD exomes, gnomAD genomes [7]) and SNVs which did not segregate with disease or show the pattern of inheritance reported for the gene and disorder.

Filtered SNVs were then 'tiered' to prioritise the variants most likely to be clinically relevant for subsequent variant classification within the recruiting GMC, which was done according to the consensus American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP) guidelines [8]. Tiering was based on the use of virtual gene panels for different genetic conditions, constructed and curated by clinical and research experts in a platform

called PanelApp [9]. At the time of analysis, the non-CF bronchiectasis gene panel consisted of 5 genes (see **Table E1**) and the 'primary ciliary disorders' gene panel (relevant to respiratory ciliopathies and PCD) consisted of 32 genes (31 genes earlier in the project) (**Table E2**). Gene panels were automatically assigned to the analysis of patient genomes by a programme called 'Panel Assigner' which selected panels based on human phenotype ontology (HPO) terms entered by clinicians to describe that patient's phenotype on recruitment [10]. The non-CF bronchiectasis gene panel was applied to all patients recruited with non-CF bronchiectasis. The primary ciliary disorders gene panel was applied to 73 non-CF bronchiectasis patients' genomes.

Variants were tiered based on whether they were protein truncating (i.e. nonsense, frameshift, splice donor or splice acceptor) or *de novo* in a gene in a selected gene panel (Tier 1), protein altering (i.e. missense or splice region) in a gene in a selected gene panel (Tier 2), or were considered to be potentially disease-causing in a gene not on a selected panel (Tier 3). Tier 1 and tier 2 variants were returned to recruiting GMC laboratories for variant classification using ACMG guidelines, to allow the GMC to determine whether these variants were pathogenic or likely pathogenic and could provide a genetic diagnosis [8]. Later in the project, structural variants (SVs) >10kb identified by Canvas [11] and Manta [12] were also returned to GMCs for classification.

Concurrent with GEL variant filtering and tiering and GMC variant classification, approved researchers who are members of Genomics England Clinical Interpretation Partnerships (GeCIPs) have access to individual variant call format (vcf) files for SNVs/Indels or Structural variants (SVs) within a highly secure GEL 'research environment', alongside tables of key sample parameters and HPO terms. As members of the Respiratory GeCIP and Paediatrics GeCIP, we formed PCD and Bronchiectasis sub-domains, carried out expert curation of all tiered variants in non-CF bronchiectasis patients tiered by GEL, and developed complementary variant analysis pipelines. These pipelines interrogated an expanded panel of 91 genes linked to PCD, MMAF (multiple morphological abnormalities of the sperm flagella syndrome) and heterotaxy phenotypes associated with motile ciliopathies (**Table E3**) for exonic or splice donor/acceptor single nucleotide, short insertion/deletion, copy number and structural variants that were predicted to be protein altering. Recently developed variant annotation tools (SpliceAI [13], UTRannotator [14]), that were not part of the GEL internal pipeline, were also applied to screen for additional potentially protein altering variants. The GMC exit questionnaires (filled in by GMC clinicians for each closed case) and Tiered variants from GEL's internal pipeline were analysed alongside the variants in genes prioritised by our additional analyses. PubMed literature and genomics/disease database (ClinVar [15], gnomAD, BRAVO/TOPMed (https://bravo.sph.umich.edu/freeze5/hg38/), GeneCards [16])

searches were used to categorise known pathogenic variants, likely pathogenic variants in known disease genes for motile ciliopathies using the ACMG/AMP guidelines [8]. We performed a separate analysis of CNVs (copy number variants) and structural variants (SVs) using the variant calls generated by the GEL pipeline.

A summary of methods and findings are presented in the flow diagram in Figure E1.

100,000 Genomes Project inclusion and exclusion criteria for non-CF bronchiectasis domain

Inclusion criteria, any one of the following:

FEV1< 30% predicted and 4 lobes involved in bronchiectasis.

Aged 50 OR extensive multi-lobar disease.

≥2 immediate family members affected.

High chloride in sweat test but *CFTR* mutation analysis (including extended NHS funded analysis) negative.

Bronchiectasis with any suspected underlying immunodeficiency aspect to be cross referenced with immunodeficiency GeCiP, e.g. bronchiectasis and recurrent non pulmonary infections.

Bronchiectasis with any suspected underlying ciliopathy OR Young's Syndrome OR Mounier Kuhn syndrome (tracheobronchomegaly).

Exclusion criteria:

Late onset.

Single lobe disease.

Those where Asthma or COPD are felt much more clearly the primary driver/ aetiology of the bronchiectasis.

Variant filtering strategy for 100,000 Genomes Project data

The following pipeline was used to interrogate SNV/Indel variants with MAF<0.001 within a panel of known PCD genes. Steps 3-8 were run separately for the groups

- Probands with Non-CF Bronchiectasis,
- Relatives of Non-CF Bronchiectasis probands,
- Relatives of other rare disease probands (excluding families with either HPO-term of bronchiectasis or diagnosis of PCD).

Variant calls from the probands were then filtered to only include those with a vcf file 'PASS' filter, MAF<0.001 and not seen as homozygous in relatives of rare disease probands (excluding families with PCD or bronchiectasis). Variants annotated as exonic and predicted to be protein altering or splice site variants were considered. The VEP UTRannotator plugin was also used to interrogate potential 5'UTR disrupting variants.

1. Identify Genes -> 'bed' format file of chromosome, start, end position coordinates.
2. Identify Participant Cohorts -> text file lists of per participant 'vcf' file locations, per cohort (Labkey tables).
3. Extract all variation within genes of interest, per participant (BEDTools).
4. Split multi-allelic variant calls (Bcftools).
5. Catalogue variants per cohort, count genotypes, output variants in a single 'vcf' format file (custom perl script).
6. Split per cohort vcfs by chromosome (Tabix).
7. Annotate per cohort variants (VEP, including UTRannotator plugin).
8. Convert annotated vcf to text format (custom perl scripts).
9. Combine variants and genotype counts seen across all cohorts into one text file (Rscript).
10. Create list of 'PASS' quality variant genotypes for probands and relatives (custom perl script).
11. Combine proband, family, annotated and filtered variant information into text report (custom perl script).

Analysis scripts used to extract SNV/indel calls are available at: https://github.com/Helgriff/Gel-rare-pipe.

Copy number and structural variant analysis

A separate analysis was performed of CNVs and SVs using the variant calls generated by the Genomics England pipeline. For each participant, SVs were detected using Manta (version 0.28.0), which combines paired and split-read evidence for SV discovery and scoring, while CNVs were called with Canvas (version 1.3.1) that employs coverage and minor allele frequencies to assign copy number.

Low-quality variants were then filtered according to the following criteria:

- Manta-called SVs with a normal sample depth near one or both variant break-ends three times higher than the chromosomal mean
- Manta-called SVs with quality score < 30

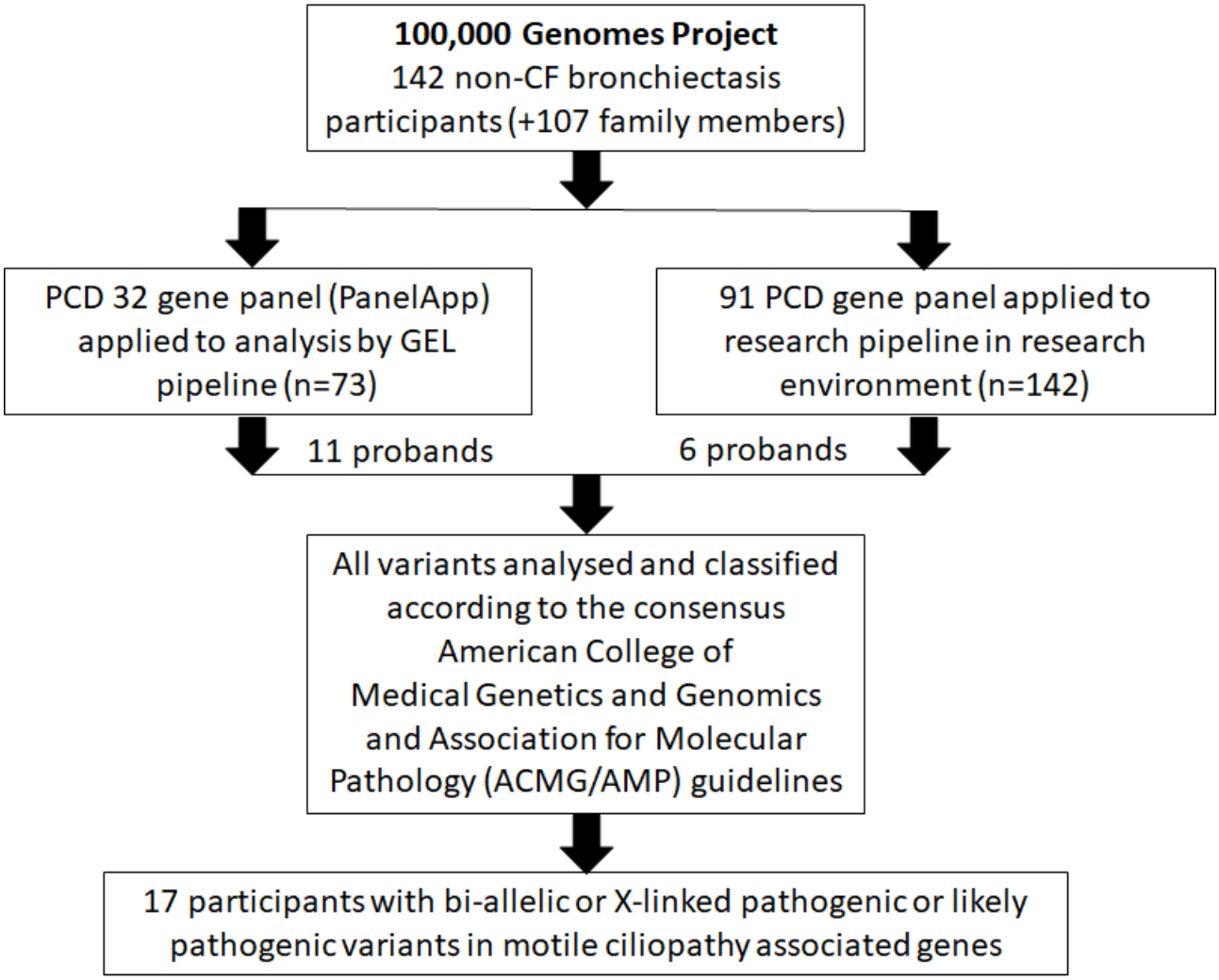
- Manta-called small variant (<1kb) where the fraction of reads with MAPQ0 around either break-end > 0.4
- Canvas-called CNVs with length < 10kb
- Canvas-called CNVs with quality score < 10

Potentially disruptive SV/CNV variants were then selected, based on the overlap with known gene regions according to the following rules:

- For deletion, duplication and insertion, overlap with an exon of a gene
- For inversion, one of the breakpoints is located within a gene and the gene is not fully contained in the inversion
- For generic breakpoints, site located within a gene

Supplementary Figure 1

Summary of 100,000 Genomes project methods and results



References

1. Hornef N, Olbrich H, Horvath J, Zariwala MA, Fliegauf M, Loges NT, Wildhaber J, Noone PG, Kennedy M, Antonarakis SE, Blouin JL, Bartoloni L, Nusslein T, Ahrens P, Griese M, Kuhl H, Sudbrak R, Knowles MR, Reinhardt R, Omran H. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am J Respir Crit Care Med* 2006: 174(2): 120-126.
2. Castleman VH, Romio L, Chodhari R, Hirst RA, de Castro SC, Parker KA, Ybot-Gonzalez P, Emes RD, Wilson SW, Wallis C, Johnson CA, Herrera RJ, Rutman A, Dixon M, Shoemark A, Bush A, Hogg C, Gardiner RM, Reish O, Greene ND, O'Callaghan C, Purton S, Chung EM, Mitchison HM. Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am J Hum Genet* 2009: 84(2): 197-209.
3. Zietkiewicz E, Bukowy-Bieryllo Z, Voelkel K, Klimek B, Dmenska H, Pogorzelski A, Sulikowska-Rowinska A, Rutkiewicz E, Witt M. Mutations in radial spoke head genes and ultrastructural cilia defects in East-European cohort of primary ciliary dyskinesia patients. *PLoS One* 2012: 7(3): e33667.
4. Whitfield M, Thomas L, Bequignon E, Schmitt A, Stouvenel L, Montantin G, Tissier S, Duquesnoy P, Copin B, Chantot S, Dastot F, Faucon C, Barbotin AL, Loyens A, Siffroi JP, Papon JF, Escudier E, Amselem S, Mitchell V, Toure A, Legendre M. Mutations in DNAH17, Encoding a Sperm-Specific Axonemal Outer Dynein Arm Heavy Chain, Cause Isolated Male Infertility Due to Asthenozoospermia. *Am J Hum Genet* 2019: 105(1): 198-212.
5. The National Genomics Research and Healthcare Knowledgebase v5, Genomics England. doi:10.6084/m9.figshare.4530893.v5. . 2019.
6. Wheway G, Mitchison HM. Opportunities and Challenges for Molecular Understanding of Ciliopathies-The 100,000 Genomes Project. *Frontiers in genetics* 2019: 10: 127.
7. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020: 581(7809): 434-443.
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* 2015: 17(5): 405-424.
9. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, Leong IUS, Smith KR, Gerasimenko O, Haraldsdottir E, Thomas E, Scott RH, Baple E, Tucci A, Brittain H, de Burca A, Ibañez K, Kasperaviciute D, Smedley D, Caulfield M, Rendon A, McDonagh EM. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature genetics* 2019: 51(11): 1560-1565.
10. Kohler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yuksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN. The Human Phenotype Ontology in 2021. *Nucleic Acids Res* 2021: 49(D1): D1207-D1217.
11. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* 2016: 32(15): 2375-2377.
12. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016: 32(8): 1220-1222.

13. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 2019: 176(3): 535-548 e524.
14. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics* 2021: 37(8): 1171-1173.
15. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018: 46(D1): D1062-D1067.
16. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 2016: 54: 1 30 31-31 30 33.