High prevalence of primary ciliary dyskinesia – results from a 'big data' genomics approach

Hannah M. Mitchison[1,2], Damian Smedley[3]

[1]Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH; [2]NIHR Biomedical Research Centre at Great Ormond Street Hospital, London, UK; h.mitchison@ucl.ac.uk.

[3]William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ; d.smedley@qmul.ac.uk;

How can the real-life frequency be calculated, of a complicated condition that occurs irregularly, with a clinically variable presentation that is often difficult to diagnose? This is a challenge for understanding societal and medical impacts of the estimated 7,000 rare disorders (those affecting > 1 in 2,000 people each) that can be poorly understood and often missed, especially if lacking clear-cut indicatory tests. The collective rare disease burden impacts over 1 in 20 people, involving a disparate range of healthcare services and creating unmet medical needs with potential for incorrect management. Better diagnostics and disease prevalence estimates are key to improving inequalities for millions of affected families globally.

Hannah *et al* address the difficulty of meaningful rare disease estimates using a genetics-led approach, as applied to an exemplar of these challenges, the rare ciliopathy disorder primary ciliary dyskinesia (PCD). Like many rare diseases, PCD lacks international data for prevalence and clinical course. The disease-causing genes govern multiciliogenesis and structural motility of cilia and sperm flagella, leading to a biologically complicated 'motile ciliopathy' disorder with variable multisystem involvement. PCD mutations notably affect airway mucociliary clearance, but also organ laterality, sperm/fallopian tube/efferent duct gamete movement, cardiac and brain development [1]. It is a monogenic and primarily autosomal recessive disorder, with X-linked recessive and autosomal dominant subtypes and there is extensive underlying genetic (locus) heterogeneity, caused by mutations in >50 different ciliary genes [1]. Bottleneck/founder effects occur, but allelic heterogeneity is overall high, with frequent family-specific variants. ClinVar lists >1,300 pathogenic/likely pathogenic variants (P/LP), mostly nonsense, frameshift, splice site and missense point mutations, plus thousands of variants of uncertain significance (VUS). For these reasons PCD diagnosis is complex, needing expertise in clinical evaluation, cilia functional testing and, increasingly, gene tests. There is no single benchmark diagnostic test and diagnostic algorithms differ in the weight given to different clinical and functional measures [2].

PCD is usually reported to affect 1 in 10,000–20,000 people, judged by methods often pre-dating gene-identification advances (situs inversus frequency, cilia structure and motility surveys, patient questionnaires) [3]. Hannah *et al* apply an a priori, i.e. phenotype-free, variant classification approach, using genomic variation datasets from a privately sequenced clinical cohort and the public gnomAD population sequencing resource. They estimate PCD prevalence and carrier frequency based on 29 genes in which recessive mutations were presumed to cause 65% of disease. First excluding any likely PCD patients, P/LP variants in these genes were catalogued in up to 41,225 individuals per gene, from diagnostic gene panel sequencing by Invitae. Sherloc, adapted from ACMG-AMP variant classification guidelines, was used to assign P/LP for single variants primarily based on predicted loss of function (LoF) and rarity, hence 'a priori' interpreted in individuals without a known PCD diagnosis. Additional LoF variants in these genes were extracted from presumed unaffected individuals within gnomAD (141,456 exome/genome sequenced controls), then all interpreted LoF variants were queried for their gnomAD allele frequency. Hardy-Weinberg equilibrium (HWE) for autosomal recessive inheritance was applied to calculate carrier frequency as well as genetic heterogeneity according to ethnicity, by estimating the number/prevalence of individuals carrying biallelic P/LP across different gnomAD-recorded populations.

Striking findings of the study are that the global PCD prevalence is substantially higher than previously reported: overall at least one in 7554 people are affected, around twice as common as typically quoted. This calculation based on numbers of recessive LoF alleles in only a subset of PCD genes, could furthermore be quite a conservative underestimate. Individuals of African ancestry had the highest PCD prevalence, greater than the next highest risk European (non-Finnish), Latino and East Asian ancestries. Different ethnic populations had markedly different disease genes and variants, with an intriguing list of the globally most mutated genes and recurrent disease mutations found per population. Twelve P/LP variants with a carrier frequency > 1 in 500, in the ethnicity in which they were most common, have mostly not been reported before or are not familiar as recurrent mutations in the PCD literature/ClinVar, being enriched in minority ethnicities.

These outcomes imply that people who are not of European/north American origin are more commonly affected by PCD than previously reported, contributing a significant impact in terms of global disease. A note of caution is that gnomAD allele frequencies might be less accurate for populations (e.g. African) where fewer genomes have been sequenced, meaning some of these variants could be false positives. However, it was already suspected that the previously understood leading causes of PCD may be heavily biased towards European/north American populations, as these are the so-far best studied. The emerging view, that specific ethnicities carry characteristic PCD gene variants, concurs with previous reports of

different human populations carrying their own specific mutational landscape and predominant PCD disease variants [4-7]. Overall, this study highlights a need for new understanding about the most at-risk communities. Understanding the distribution of PCD-causing variants in different populations is potentially field-changing, for better targeted diagnostics, better global-level disease recognition, and future disease management potentially tailored to a more/less severe predicted clinical course, along with more accurately targeted therapeutic programmes, including genetically targeted ones [8].

Like sequence-based revisits of other disease incidences, this study is reliant on robust variant pathogenicity algorithms, with potential for over-interpretation if predicted LoF alleles prove benign or VUS prove pathogenic[9]. The authors acknowledge other study limitations, including more rigorous variant classification of Invitae than gnomAD variants. Missing data bias could arise from variable sequencing coverage of gene panels versus omics, and from variant exclusions necessitated by merging clinical and control datasets, as some Invitae variants were absent from gnomAD (e.g., copy number variants) and gnomAD variants absent from the smaller Invitae dataset. The calculation of disease prevalence summed PCD incidence per gene within each ethnicity, based on a suggested 65% of cases having biallelic disease-causing variants in one of the 29 (of the known 50) genes analysed, which omits the disease contribution of many PCD genes and mutations including X-linked and dominant effects [1]. Other omissions were any population bottleneck effects or deviations from HWE e.g. from possible embryonic lethal alleles. Population-specific disease prevalence accuracies could also be limited, until our understanding of PCD mutations in all ethnicities is more complete, with potential bias arising from the limited gnomAD-recorded ethnicities and no record of ethnicity effects in the Invitae cases. VUS are a well-recognised complication in PCD research, and the comparison global disease prevalence estimates incorporating VUS in this study should be viewed with great caution.

Such analyses need further refinement, as our knowledge about VUS functions increase and the number of recognised motile ciliopathy genes and variants expands. With 100s of variants excluded as VUS or simply not identified in this study, PCD prevalence could still be greatly under-estimated. In future this approach will hopefully improve, linked into new clinical translational initiatives from the PCD community that continue to define and refine the most critical genetic contributions to disease [10,11].

References

1. Legendre M, Zaragosi LE, Mitchison HM. Motile cilia and airway disease. *Semin Cell Dev Biol* 2021; **110**: 19-33.
2. Shoemark A, Rubbo B, Haarman E, et al. The Controversies and Difficulties of Diagnosing Primary Ciliary Dyskinesia. *Am J Respir Crit Care Med* 2020; **201**(1): 120-2.
3. Kuehni CE, Frischer T, Strippoli MP, et al. Factors influencing age at diagnosis of primary ciliary dyskinesia in European children. *Eur Respir J* 2010; **36**(6): 1248-58.
4. 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): 322-30.
5. Zietkiewicz E, Nitka B, Voelkel K, et al. Population specificity of the DNAI1 gene mutation spectrum in primary ciliary dyskinesia (PCD). *Respir Res* 2010; **11**: 174.
6. Mani R, Belkacem S, Soua Z, et al. Primary ciliary dyskinesia gene contribution in Tunisia: Identification of a major Mediterranean allele. *Hum Mutat* 2020; **41**(1): 115-21.
7. Yiallouros PK, Kouis P, Kyriacou K, et al. Implementation of multigene panel NGS diagnosis in the national primary ciliary dyskinesia cohort of Cyprus: An island with a high disease prevalence. *Hum Mutat* 2021; **42**(6): e62-e77.
8. Paff T, Omran H, Nielsen KG, Haarman EG. Current and Future Treatments in Primary Ciliary Dyskinesia. *Int J Mol Sci* 2021; **22**(18).
9. Evans DG, Turnbull C, Woodward ER. Concern regarding classification of germline TP53 variants as likely pathogenic. *Hum Mutat* 2019; **40**(6): 828-31.
10. Raidt J, Pennekamp P, Riepenhausen S, et al. The international primary ciliary dyskinesia (PCD) registry reports genotype/phenotype-correlations. *European Respiratory Journal* 2021; **58**: PA2233.
11. Mani R, Gomes M, Rodríguez González A, et al. Development and first results of the BEAT-PCD international Primary Ciliary Dyskinesia gene variant database: CiliaVar. *European Respiratory Journal* 2021; **58** PA3458.